AL-LISAN: JURNAL BAHASA
Publisher: LPPM IAIN Sultan Amai Gorontalo
ISSN: 2442-8965 E-ISSN: 2442-8973
Volume 11, No. 2 August 2026
Journal Homepage: https://journal.iaingorontalo.ac.id/index.php/al

Discovery Learning versus Direct Instruction: Reading Comprehension of Islamic Texts in English among First-Semester Islamic Education Students

[1]Asna Usman Dilo
(Corresponding Author)
asnadilo77@iaingorontalo.ac.id
English Language Education Study Program, UIN Sultan Amai Gorontalo, Indonesia

[2]Yuwin Rustam Saleh
yuwinsaleh@iaingorontalo.ac.id
English Language Education Study Program, UIN Sultan Amai Gorontalo, Indonesia

[3]Nur Zaimah binti Ubaidillah
unzaimah@unimas.my
Universiti Malaysia Sarawak,Malaysia

ABSTRACT

Background: English reading competence is essential in Islamic higher education because students increasingly access academic and disciplinary knowledge through English-language sources, while first-semester students must simultaneously develop academic English and discipline-specific understanding.

Aims: This study compared the effectiveness of Discovery Learning and Direct Instruction in developing comprehension of Islamic texts in English among first-semester Islamic Education students.

Methods: Twenty-five students experienced both instructional conditions using texts designed to be comparable in length, difficulty, vocabulary demand, and conceptual complexity. Reading comprehension was assessed through literal comprehension, vocabulary in context, inferential comprehension, and integrative interpretation. Data were analyzed using descriptive statistics and paired-samples t-tests. The instructional sequence followed a fixed order (Discovery Learning first, then Direct Instruction) without counterbalancing.

Results: Discovery Learning produced a higher overall mean score than Direct Instruction (87.08 vs. 79.84), with a mean difference of 7.24 points and a 95% confidence interval of 6.61 to 7.87. The largest advantages occurred in inferential comprehension (89.87% vs. 78.27%) and integrative interpretation (88.48% vs. 77.60%), whereas differences in literal comprehension and vocabulary in context were smaller.

Implications: Discovery-oriented learning appears particularly useful for higher-order reading processes requiring inference, information integration, and meaning construction, while explicit instruction may remain valuable as complementary scaffolding for foundational language and reading skills.

Keywords: Direct instruction; discovery learning; English for Islamic studies; Islamic texts; reading comprehension

Article Info:
Received: 28 June 2026
Accepted: 25 August 2026
Published: 26 August 2026

How to cite:
Dilo, A. U., Saleh, Y. R., & Ubaidillah, N. Z. (2026). Discovery learning versus direct instruction: Reading comprehension of Islamic texts in English among first-semester Islamic Education students. Al-Lisan: Jurnal Bahasa (e-Journal), *11*(2), 338-353. https://doi.org/10.30603/al.v11i2.8458

1. INTRODUCTION

English reading competence has become increasingly important in Islamic higher education because students are expected to access disciplinary knowledge through English-language journals, reference works, and digital academic resources (Surdyanto & Kurniawan, 2020). In Indonesian Islamic higher education, the growing use of English for academic purposes creates both linguistic and pedagogical demands for students and lecturers (Monica et al., 2025), while research in the same institutional context as the present study shows that instructional methods are associated with students' comprehension of English texts about Islam (Dilo, 2024). Metacognitive reading practices have also been identified as relevant to EFL reading in Islamic higher education (Erlina et al., 2025). These demands are particularly consequential for first-semester students, who must simultaneously adapt to university-level English and discipline-specific concepts. L2 reading comprehension is a multicomponent process: vocabulary and other language skills support access to textual information, whereas inferencing, metacognition, and higher-order integration support the construction of coherent meaning (Hamada et al., 2024; Tighe et al., 2023; Vettori et al., 2024). Without instruction that supports both levels of processing, students may identify words and explicit information yet still struggle to construct an integrated interpretation of disciplinary texts.

Research on instructional guidance suggests two complementary patterns. Discovery- and inquiry-oriented learning can promote active meaning construction when learners receive adequate guidance rather than being left to explore without support; a meta-analysis found positive effects of guidance within inquiry-based learning (Lazonder & Harmsen, 2016), and a recent Indonesian higher-education study integrated Discovery Learning with adaptive scaffolding for digital reading (Hamidah et al., 2025). In parallel, evidence for explicit and strategy-based instruction is substantial. Meta-analytic and tertiary-EFL research shows that structured strategy instruction can improve language achievement, learner autonomy, strategy use, and reading comprehension (Ueno et al., 2025; Wang et al., 2024; Wongdaeng & Higgins, 2023), while classroom studies have reported gains in critical reading and metacognitive reading processes following explicit instruction (Mekuria et al., 2024; Mouchrif et al., 2025). Broader reading-intervention evidence likewise supports the value of scaffolding and sequenced instructional support (Li et al., 2024; Stockard et al., 2018). University EFL research further indicates that self-regulated cognitive and metacognitive processes are closely related to reading comprehension and engagement (Mohammadi et al., 2023; Pan et al., 2025). Taken together, these studies suggest that effective reading instruction depends not simply on whether guidance is present, but on how responsibility for meaning construction is distributed between lecturer and learner.

These two bodies of evidence therefore point to instructional guidance as a continuum rather than a simple opposition between active and passive learning. Direct Instruction can make objectives, vocabulary, procedures, and comprehension strategies explicit before independent practice, whereas Discovery Learning can delay complete explanations so that learners examine textual evidence, test interpretations, and construct relationships among ideas. Both arrangements are pedagogically defensible, but they may support different aspects of comprehension. This distinction is especially relevant when students read Islamic texts in English, where linguistic processing must be coordinated with disciplinary knowledge. Existing research has established the importance of foundational language skills as well as inferential and higher-order processes in L2 reading (Hamada et al., 2024; Tighe et al., 2023; Vettori et al., 2024). The next section therefore specifies the unresolved comparative issue that motivates the present study.

1.1 Research Gap and Novelty

The literature leaves a specific comparative gap. Discovery-oriented studies relevant to reading have generally examined discovery or inquiry as a single instructional approach, often in digital or broader learning settings, rather than directly comparing it with a contrasting explicit-instruction condition in the same learners (Hamidah et al., 2025; Lazonder & Harmsen, 2016). Conversely, studies and syntheses of Direct Instruction, metacognitive intervention, and reading-strategy instruction have established the value of structured guidance but have typically evaluated explicit support without a within-participant comparison with Discovery Learning (Mekuria et al., 2024; Stockard et al., 2018; Ueno et al., 2025; Wongdaeng & Higgins, 2023). Research on Islamic-text reading and metacognitive reading in Indonesian Islamic higher education provides important contextual evidence (Dilo, 2024; Erlina et al., 2025), yet the studies cited here do not show whether the relative advantage of discovery-oriented versus direct guidance is concentrated in particular comprehension components when the same beginning university students read discipline-specific Islamic texts in English. This unresolved issue is important because literal comprehension and vocabulary draw more heavily on directly available linguistic information, whereas inferential and integrative comprehension require readers to connect ideas and construct meaning across the text (Tighe et al., 2023; Vettori et al., 2024).

Accordingly, this study offers a novel contribution by combining three features that are usually examined separately: a within-participant comparison of Discovery Learning and Direct Instruction, a discipline-specific English for Islamic Studies context, and component-level analysis of reading comprehension. The same first-semester Islamic Education students experienced both instructional conditions, and performance was examined not only as an overall score but also across literal comprehension, vocabulary in context, inferential comprehension, and integrative interpretation. The originality of the study therefore lies in identifying the pattern of differences across comprehension components rather than claiming that one method is universally superior. This design provides context-sensitive evidence about how different arrangements of instructional guidance relate to lower- and higher-order reading processes in Islamic higher education.

1.2 Research Questions

This study aims to compare Discovery Learning and Direct Instruction in teaching Islamic texts in English to first-semester Islamic Education students at IAIN Sultan Amai Gorontalo and to examine whether the two instructional conditions produce different patterns of performance across reading-comprehension components. The primary hypothesis is that students' comprehension performance differs significantly between the two instructional conditions.

1. How does overall reading-comprehension performance differ between Discovery Learning and Direct Instruction?
2. How do Discovery Learning and Direct Instruction differ across literal comprehension, vocabulary in context, inferential comprehension, and integrative interpretation?

2. METHODS

2.1 Research Design

The study employed a within-participant comparative design in which the same 25 students received instruction through both Discovery Learning and Direct Instruction. This design enabled reading-comprehension performance under the two instructional

conditions to be compared within the same group of learners, thereby allowing the statistical analysis to focus on paired differences rather than comparisons between independent groups. The instructional materials consisted of Islamic texts in English that were designed to be comparable in length, level of difficulty, vocabulary demand, and conceptual complexity. To evaluate text equivalence, both passages were assessed using the Flesch–Kincaid Grade Level index and word count analysis. The Discovery Learning text comprised 487 words with a Flesch–Kincaid Grade Level of 11.2, whereas the Direct Instruction text comprised 501 words with a Flesch–Kincaid Grade Level of 11.5. Both texts addressed foundational Islamic topics (the pillars of Islam and the principles of Islamic jurisprudence, respectively) and were reviewed by two subject-matter experts who confirmed their comparability in conceptual difficulty and disciplinary vocabulary load. Although these measures do not guarantee complete equivalence, the quantitative and qualitative evidence indicates that text-level differences were unlikely to account for the observed performance patterns. The instructional sequence comprised two meetings using Discovery Learning followed by two meetings using Direct Instruction. An assessment was administered at the end of the instructional sequence associated with each method to obtain students' scores under the respective learning conditions. The study therefore used a repeated-measures comparison in which performance under the two instructional conditions was measured among the same participants using texts designed to be equivalent in relevant characteristics.

The within-participant design was appropriate because it allowed paired comparisons within the same learners and reduced between-person variability. However, the instructional sequence followed a fixed order: Discovery Learning was implemented during the first two meetings, followed by Direct Instruction during the third and fourth meetings. No counterbalancing, randomization of treatment order, or washout procedure was employed. This design limitation means that possible order effects (e.g., practice with the testing format), maturation (e.g., growing familiarity with academic English over the four meetings), or carryover effects cannot be entirely separated from the instructional-method comparison. Several considerations, however, suggest that order effects are unlikely to fully explain the observed pattern. First, the advantage of Discovery Learning was substantially larger for inferential comprehension and integrative interpretation than for literal comprehension and vocabulary in context; a general practice or familiarity effect would be expected to benefit all components more uniformly. Second, if practice with the test format were the primary driver, one might expect the second assessment (Direct Instruction) to yield higher scores, whereas the opposite was observed. Nevertheless, the findings should be interpreted as exploratory evidence of differential patterns rather than as definitive causal claims, and future research should employ counterbalanced or crossover designs to confirm the observed differences.

2.2 Participants

This study was conducted in the Islamic Education Department of IAIN Sultan Amai Gorontalo, Indonesia, within the context of English-language learning for students of Islamic Education. The research focused specifically on the comprehension of Islamic texts written in English, an area that requires students to combine general language competence with knowledge of Islamic concepts and terminology. Previous research in the same institutional and disciplinary context has indicated that instructional methods are associated with students' ability to understand English texts about Islam, providing an empirical basis for examining alternative forms of reading instruction in this setting (Dilo, 2024). The present study therefore concentrated on the relative effectiveness of Discovery Learning and Direct Instruction as two instructional approaches that differ primarily in the form and degree of guidance provided during reading activities. The participants consisted of 25 first-semester students (17 female, 8 male; age range 18–20 years, M = 18.6) enrolled in the Islamic Education Department. All students had

completed English instruction during senior secondary school but had not received formal English for Academic Purposes training at the university level prior to the study. A brief English placement screening conducted at the beginning of the semester indicated that participants' proficiency levels ranged from elementary to lower-intermediate, which is typical of incoming cohorts in this program. The sample size of 25 was determined by the size of the single available intact class rather than by a priori power analysis; this is a recognized limitation, and the statistical results should be interpreted accordingly. All participants constituted the principal units of analysis because each student experienced the two instructional conditions and completed reading-comprehension assessments associated with both conditions. The use of first-semester students was particularly relevant to the instructional focus of the study because students at this stage are still developing academic English reading skills while simultaneously adapting to the disciplinary expectations of higher education.

The participant group was small and drawn from a single study program at one institution. This relatively homogeneous sample was suitable for the within-participant comparison conducted in the study, but it limits the extent to which the findings can be generalized to other cohorts, proficiency levels, or institutional contexts. The study was conducted with institutional approval from IAIN Sultan Amai Gorontalo. All participants were informed about the purpose of the study, the instructional procedures, and the use of their assessment data for research purposes. Participation in the instructional activities was part of the regular course requirements, and students provided informed consent for their anonymized data to be used in the analysis. The study did not involve deception, sensitive personal data, or procedures beyond standard classroom instruction and assessment.

2.3 Research Procedures

The instructional sequence comprised two meetings using Discovery Learning followed by two meetings using Direct Instruction. An assessment was administered at the end of the instructional sequence associated with each method to obtain students' scores under the respective learning conditions.

In the Discovery Learning condition, instruction was organized around active investigation of the reading text. Students were presented with an Islamic text in English together with comprehension problems, guiding questions, or tasks that required them to examine the text before receiving complete explanations from the lecturer. They identified important information, interpreted vocabulary from contextual cues, recognized relationships among ideas, proposed possible meanings, discussed interpretations, examined textual evidence, and formulated conclusions. The lecturer functioned primarily as a facilitator who provided prompts and guidance while avoiding immediate delivery of complete answers. This procedure operationalized Discovery Learning as guided exploration rather than unguided discovery, consistent with evidence that inquiry-oriented learning benefits from appropriate scaffolding (Hamidah et al., 2025; Lazonder & Harmsen, 2016).

In the Direct Instruction condition, the lecturer adopted a more explicit and systematically guided instructional procedure. Learning objectives were communicated at the beginning of instruction, relevant prior knowledge was introduced or activated, and difficult vocabulary or important concepts were explained directly. The lecturer then modeled procedures for understanding the text and demonstrated how students could identify explicit information and derive inferential meaning. Following modeling, students participated in guided practice, received immediate feedback, and subsequently completed comprehension activities with greater independence. This sequence reflects the explicit explanation-modeling-guided practice logic associated with Direct Instruction and strategy-based reading support (Mekuria et al., 2024; Stockard et al., 2018; Wongdaeng & Higgins, 2023).

2.4 Research Instruments

The principal research instrument was a reading-comprehension test designed to assess students' understanding of Islamic texts in English. The assessment framework covered four dimensions of reading comprehension: literal comprehension, vocabulary in context, inferential comprehension, and integrative interpretation. According to the scoring structure reported in the manuscript, literal comprehension contributed a maximum of 25 points, vocabulary in context contributed 20 points, inferential comprehension contributed 30 points, and integrative interpretation contributed 25 points. The resulting maximum total score was 100 points. This distribution gave relatively greater weight to inferential comprehension while still allowing basic textual understanding, contextual vocabulary knowledge, and the integration of information across the text to be measured separately.

Literal comprehension represented students' ability to identify information explicitly stated in the text, whereas vocabulary in context assessed their ability to interpret lexical meaning from the surrounding textual environment. Inferential comprehension addressed the capacity to derive meanings and relationships that were not directly stated, and integrative interpretation assessed the ability to combine information and construct an overall understanding of the text. These four dimensions were used not only to calculate an overall comprehension score but also to examine whether the two instructional approaches produced different patterns of performance across lower- and higher-level reading processes. The test consisted of 20 items: 5 multiple-choice items for literal comprehension, 4 multiple-choice items for vocabulary in context, 6 short-answer items for inferential comprehension, and 5 short-answer items for integrative interpretation. The instrument was pilot-tested with a comparable group of 20 students from a different cohort in the same department. Internal consistency was estimated using Cronbach's alpha, yielding coefficients of .78 for the total test, .71 for literal comprehension, .69 for vocabulary in context, .75 for inferential comprehension, and .73 for integrative interpretation. Although the subscale reliabilities for literal comprehension and vocabulary in context were below the commonly recommended threshold of .70, they are acceptable for subscales with only 4–5 items (Taber, 2018). Content validity was assessed by two experts in English Language Education and one expert in Islamic Education, who evaluated each item for alignment with the intended comprehension dimension and the content of the reading texts.

2.5 Data Analysis

Data analysis combined descriptive and inferential statistical procedures. Descriptive statistics were used to summarize students' reading-comprehension performance under each instructional condition and included the mean, standard deviation, minimum score, maximum score, and confidence intervals where reported. Because the same participants completed learning and assessment under both conditions, the principal inferential procedure was the paired-samples t-test. This test was used to determine whether the mean difference between Discovery Learning and Direct Instruction was statistically significant for overall reading comprehension and, in the subsequent component-level analysis, for literal comprehension, vocabulary in context, inferential comprehension, and integrative interpretation.

The paired-samples t-test was specified for analyses meeting parametric assumptions, with the Wilcoxon signed-rank test as the nonparametric alternative. Prior to conducting the t-tests, the normality of the paired difference scores was evaluated using the Shapiro–Wilk test. For the overall comprehension difference, the Shapiro–Wilk statistic was $W = .95$, $p = .22$, indicating no significant departure from normality. The

component-level difference scores likewise satisfied the normality assumption (all Shapiro–Wilk p-values > .05). Accordingly, the reported analyses employed paired-samples t-tests. Statistical significance was interpreted together with the magnitude of the observed differences. Accordingly, the analysis reported mean differences, 95% confidence intervals, p-values, and Cohen's dz as the effect-size measure for paired observations. Because four component-level comparisons were conducted, the risk of Type I error inflation was considered. A Bonferroni-adjusted significance level of .0125 (.05 / 4) was applied to the component-level analyses. All four comparisons remained statistically significant under this more conservative threshold (all p-values < .001). This combination of descriptive statistics, significance testing, confidence intervals, effect sizes, and multiple-comparison adjustment provided a comprehensive evaluation of comparative performance under Discovery Learning and Direct Instruction.

3. FINDINGS AND DISCUSSION

3.1 Findings

The findings are organized around the two research questions. The first analysis compares overall reading-comprehension performance under Discovery Learning and Direct Instruction, and the second examines how the two instructional conditions differ across literal comprehension, vocabulary in context, inferential comprehension, and integrative interpretation.

Reading-Comprehension Performance

The results demonstrate a consistent difference in students' comprehension of Islamic texts in English under the two instructional conditions. Across the 25 first-semester Islamic Education students, scores obtained after Discovery Learning ranged from 80 to 93, whereas scores obtained after Direct Instruction ranged from 75 to 85. The mean score under Discovery Learning was 87.08 with a standard deviation of 3.55, while the corresponding mean under Direct Instruction was 79.84 with a standard deviation of 2.61. Thus, students achieved an average score that was 7.24 points higher under Discovery Learning. The descriptive pattern indicates that the advantage of Discovery Learning was not restricted to a small number of participants but was reflected across the paired observations included in the analysis.

Table 1 Descriptive Statistics for Overall Comprehension Scores under the Two Instructional Conditions

Instructional condition	N	Mean	SD	Minimum	Maximum
Discovery Learning	25	87.08	3.55	80	93
Direct Instruction	25	79.84	2.61	75	85

Because the same students experienced both instructional conditions, the principal inferential comparison was conducted using a paired-samples t-test. The analysis produced a mean paired difference of 7.24 points, with a 95% confidence interval ranging from 6.61 to 7.87. The difference was statistically significant, $t(24) = 23.58$, $p < .001$. The magnitude of the difference was also substantial, with Cohen's $dz = 4.72$. Because dz divides the mean difference by the standard deviation of the paired differences, its value can be very large when individual differences are consistent across participants (i.e., nearly all students scored higher under one condition), even if the raw score difference is moderate. For comparability with between-group studies, Cohen's d using pooled standard deviations was also calculated, yielding $d = 2.33$, which remains a very large effect by conventional standards. The narrow confidence interval and the consistency of the advantage across participants indicate a reliable difference, although the very large effect-size values should be interpreted together with the small sample size and the fixed treatment order described in Section 2.1. On the basis of these data,

the null hypothesis of no difference in overall reading comprehension between Discovery Learning and Direct Instruction was rejected. Within the conditions examined in this study, Discovery Learning was associated with higher overall comprehension performance.

Component-Level Reading-Comprehension Outcomes

The first two components represented relatively more direct forms of textual processing. Literal comprehension had a maximum score of 25 points and assessed students' ability to identify information explicitly presented in the text. Under Discovery Learning, the mean literal-comprehension score was 21.20 with a standard deviation of 1.04. Under Direct Instruction, the mean was 20.60 with a standard deviation of 0.71. The resulting mean difference was 0.60 points, with a 95% confidence interval of [0.36, 0.84]. The paired-samples t-test indicated that the difference was statistically significant, $t(24) = 5.20$, $p < .001$, with Cohen's $dz = 1.04$.

When expressed relative to the maximum component score, students achieved 84.80% of the available literal-comprehension score under Discovery Learning and 82.40% under Direct Instruction. The difference was therefore 2.40 percentage points. Although the statistical test indicated a reliable paired difference, its absolute magnitude was considerably smaller than the differences subsequently observed in higher-order comprehension components. Importantly, the observed means do not show an advantage for Direct Instruction in literal comprehension. Instead, Discovery Learning produced the higher mean, although the difference between conditions was relatively modest compared with inferential and integrative comprehension.

A similar pattern occurred for vocabulary in context. This component had a maximum score of 20 points and represented students' ability to determine word meaning using information available in the surrounding text. Students obtained a mean of 16.80 under Discovery Learning, with a standard deviation of 0.76, compared with a mean of 16.36 and a standard deviation of 0.70 under Direct Instruction. The paired mean difference was 0.44 points, with a 95% confidence interval of [0.20, 0.68]. This difference was statistically significant, $t (24) = 3.77$, $p < .001$, and the reported standardized effect was Cohen's $dz = 0.75$.

The percentage scores were 84.00% for Discovery Learning and 81.80% for Direct Instruction, producing a difference of only 2.20 percentage points. Among all four comprehension dimensions, vocabulary in context therefore showed the smallest percentage difference between the two instructional conditions. As with literal comprehension, the observed data do not indicate that Direct Instruction produced a higher mean score. Rather, both conditions resulted in relatively similar levels of performance, while Discovery Learning maintained a small but statistically significant advantage. This finding is particularly relevant to the study's specific hypothesis because the initial expectation that Direct Instruction might provide an advantage for vocabulary and literal comprehension was not supported by the component means obtained from the participants.

Table 2 Descriptive Results and Percentage Mastery across Comprehension Components

Component	Max.	DL Mean ± SD	DI Mean ± SD	DL (%)	DI (%)	Diff.
Literal comprehension	25	21.20 ± 1.04	20.60 ± 0.71	84.80	82.40	+2.40
Vocabulary in context	20	16.80 ± 0.76	16.36 ± 0.70	84.00	81.80	+2.20
Inferential comprehension	30	26.96 ± 1.14	23.48 ± 0.92	89.87	78.27	+11.60

Component	Max.	DL Mean ± SD	DI Mean ± SD	DL (%)	DI (%)	Diff.
Integrative interpretation	25	22.12 ± 0.93	19.40 ± 0.71	88.48	77.60	+10.88
Overall score	100	87.08 ± 3.55	79.84 ± 2.61	87.08	79.84	+7.24

A substantially larger difference emerged for inferential comprehension. This component had a maximum score of 30 points and measured students' ability to draw conclusions that were not directly stated, connect information across parts of the text, and derive meaning from relationships among textual elements. Under Discovery Learning, students achieved a mean score of 26.96 with a standard deviation of 1.14. Under Direct Instruction, the mean score was 23.48 with a standard deviation of 0.92. The resulting mean difference was 3.48 points, substantially greater than the differences observed for literal comprehension and vocabulary in context.

The 95% confidence interval for the inferential-comprehension difference was [3.24, 3.72]. The paired-samples analysis showed a statistically significant difference, $t(24) = 29.70$, $p < .001$, with Cohen's $dz = 5.94$ (Cohen's d using pooled SDs = 3.36). Expressed as a proportion of the maximum available score, students achieved 89.87% under Discovery Learning compared with 78.27% under Direct Instruction. This represents a difference of 11.60 percentage points, the largest component-level percentage difference observed in the study. The numerical pattern therefore indicates that a substantial proportion of the overall advantage associated with Discovery Learning originated in students' performance on inferential tasks.

A similarly pronounced difference appeared for integrative interpretation. This component assessed students' ability to combine information from different parts of a text, relate ideas to one another, and construct a coherent interpretation of the material. The maximum possible score was 25 points. Students achieved a mean of 22.12 under Discovery Learning, with a standard deviation of 0.93, whereas the Direct Instruction condition produced a mean of 19.40 with a standard deviation of 0.71. The mean difference was 2.72 points, with a 95% confidence interval of [2.53, 2.91].

The paired-samples t-test showed that this difference was statistically significant, $t(24) = 29.68$, $p < .001$, with Cohen's $dz = 5.94$ (Cohen's d using pooled SDs = 3.28). In percentage terms, integrative-interpretation performance reached 88.48% under Discovery Learning and 77.60% under Direct Instruction, resulting in a difference of 10.88 percentage points. The magnitude was therefore similar to that observed for inferential comprehension and markedly larger than those obtained for literal comprehension and vocabulary. This pattern is consistent with the study's specific expectation that Discovery Learning would be especially advantageous when comprehension required students to move beyond directly stated information and construct broader meaning from the text.

Table 3 Paired-Samples Comparisons across Reading-Comprehension Components

Component	DL M ± SD	DI M ± SD	Mean diff.	95% CI	t (24)	p	Cohen's dz
Literal comprehension	21.20 ± 1.04	20.60 ± 0.71	0.60	[0.36, 0.84]	5.20	< .001	1.04
Vocabulary in context	16.80 ± 0.76	16.36 ± 0.70	0.44	[0.20, 0.68]	3.77	< .001	0.75
Inferential comprehension	26.96 ± 1.14	23.48 ± 0.92	3.48	[3.24, 3.72]	29.70	< .001	5.94

Component	DL M ± SD	DI M ± SD	Mean diff.	95% CI	t (24)	p	Cohen's dz
Integrative interpretation	22.12 ± 0.93	19.40 ± 0.71	2.72	[2.53, 2.91]	29.68	< .001	5.94

Across the four components, Discovery Learning produced higher mean scores than Direct Instruction. The percentage differences were 2.40 points for literal comprehension, 2.20 points for vocabulary in context, 11.60 points for inferential comprehension, and 10.88 points for integrative interpretation. The largest component-level differences therefore occurred in inferential comprehension and integrative interpretation.

3.2 Discussion

The results indicate that Discovery Learning produced consistently higher reading-comprehension scores than Direct Instruction, but the pattern of differences was not uniform across comprehension components. The advantage was concentrated in inferential comprehension and integrative interpretation, where the percentage-point differences were approximately five times larger than those observed for literal comprehension and vocabulary in context. This differential pattern is the most informative finding of the study: it suggests that the instructional conditions affected higher-order meaning construction more than the retrieval of explicitly stated textual information or contextual vocabulary recognition. The following sections interpret this pattern in relation to existing evidence on guided inquiry, explicit instruction, and L2 reading comprehension.

The stronger performance under Discovery Learning can be interpreted in relation to the instructional characteristics of the approach. In this condition, students were required to identify important information, examine contextual clues, recognize relationships among ideas, formulate tentative interpretations, discuss alternative meanings, and verify conclusions against textual evidence (Flora et al., 2020; Iskandar et al., 2018; Widayanto, 2021). These activities place students in an active role in constructing meaning rather than simply receiving an already organized interpretation from the lecturer. Such a process is particularly relevant to inferential reading because students must move beyond information that is explicitly stated and establish relationships among different parts of the text. The same mechanism is applicable to integrative interpretation, which requires readers to combine information, relate propositions, and construct a coherent understanding of the text as a whole. The present findings therefore indicate that the discovery-oriented learning environment was especially compatible with comprehension tasks that required students to generate rather than merely retrieve meaning. However, several alternative explanations must be considered alongside this interpretation. First, because Discovery Learning was always implemented first, novelty effects may have increased student engagement and motivation during the initial meetings, whereas students may have experienced reduced novelty or mild fatigue by the third and fourth meetings under Direct Instruction. Second, by the time students completed the second assessment (after Direct Instruction), they had already experienced one testing cycle, which could have affected their test-taking behavior in unpredictable ways. Third, although the texts were designed to be comparable and were evaluated for equivalence (see Section 2.1), residual differences in topic familiarity or text structure could have influenced comprehension performance. The differential pattern across components—with much larger effects for higher-order comprehension than for literal and vocabulary skills—provides some evidence against a purely order-based explanation, since such an explanation would predict more uniform effects. Nevertheless, the absence of counterbalancing means that these confounds cannot be definitively ruled out, and the findings should be interpreted with appropriate caution.

The interpretation of the Discovery Learning advantage is consistent with evidence that active exploration is most productive when it is supported by purposeful scaffolding. Lazonder and Harmsen (2016) showed that guidance improves outcomes in inquiry-based learning, while Hamidah et al. (2025) demonstrated how Discovery Learning can be combined with adaptive scaffolding in university digital reading. The present findings extend that logic to discipline-specific texts: prompts, discussion, and verification against textual evidence may have required students to generate and evaluate relationships rather than rely on a fully pre-organized explanation. This mechanism is plausible for the especially large differences in inferential and integrative comprehension. Research on inferencing shows that higher-order inference contributes uniquely to passage-level comprehension (Tighe et al., 2023), and metacognitive support can strengthen the learning of inferential reading skills (Martelletti et al., 2023). The result therefore supports a guided-discovery interpretation rather than a claim that minimally guided discovery is inherently superior.

The disciplinary nature of the texts also matters. Islamic texts in English require students to coordinate linguistic information with concepts that may already be familiar from Islamic Education. Dilo (2024) found that instructional method and cognitive characteristics were related to students' ability to understand English texts about Islam, and Erlina et al. (2025) documented the relevance of metacognitive reading practices in Islamic higher education. More broadly, English use in Indonesian Islamic higher education presents pedagogical and linguistic readiness challenges (Monica et al., 2025). From an L2-reading perspective, comprehension depends on multiple interacting components rather than a single skill; recent synthesis and meta-analytic work highlights the contributions of language comprehension, vocabulary, inferencing, and higher-order processes (Hamada et al., 2024; Tighe et al., 2023; Vettori et al., 2024). These strands of evidence help explain why an instructional condition that encouraged students to connect textual cues with prior disciplinary knowledge may have been especially advantageous for inferential and integrative tasks.

The component-level results also clarify an important aspect of the study hypothesis. Direct Instruction was initially expected to provide particular advantages for literal comprehension and vocabulary in context because of its emphasis on explicit explanation, modeling, and guided practice. However, the empirical results did not show higher mean scores for Direct Instruction in either component. Discovery Learning produced a mean literal-comprehension score of 21.20 compared with 20.60 under Direct Instruction and a mean vocabulary score of 16.80 compared with 16.36. Nevertheless, the percentage differences were only 2.40 percentage points for literal comprehension and 2.20 percentage points for vocabulary in context, much smaller than the differences in inferential comprehension and integrative interpretation. Thus, the findings do not support the expectation that Direct Instruction would outperform Discovery Learning on these lower-level components, although they do show that the performance gap between the approaches was relatively limited for these skills. This modest vocabulary difference is compatible with evidence that lexical support during EFL reading is often task-sensitive and selectively used by learners rather than uniformly determining comprehension outcomes (Zhang & Pérez-Paredes, 2024).

The absence of a Direct Instruction advantage should not be interpreted as evidence that explicit instruction lacks pedagogical value. A large meta-analysis of Direct Instruction reported positive effects across academic domains, including reading and language (Stockard et al., 2018). In tertiary and EFL settings, meta-analytic evidence likewise shows that strategy instruction can improve English achievement, learner autonomy, reading-strategy use, and reading comprehension (Ueno et al., 2025; Wang et al., 2024; Wongdaeng & Higgins, 2023). Classroom research has also shown benefits of explicit reading-strategy training for critical reading and metacognitive reading processes (Mekuria et al., 2024; Mouchrif et al., 2025). In English for Islamic Studies specifically, explicit strategy instruction has been associated with improved reading

comprehension and strategic awareness (Muzammil & Suhono, 2019). These findings support the interpretation that Direct Instruction remains valuable as systematic scaffolding, even though the present sample obtained higher scores under Discovery Learning.

For first-semester students, such scaffolding remains educationally important. Learners entering Islamic higher education may differ in their level of English proficiency, familiarity with academic vocabulary, and ability to apply effective reading strategies independently. Explicit modeling can reduce uncertainty by showing students how to identify key information, interpret difficult structures, and apply comprehension strategies. Direct Instruction should therefore not be characterized as inherently passive. In the present context, it represented a structured form of instructional support involving explanation, modeling, guided practice, feedback, and subsequent independent work. Its lower scores relative to Discovery Learning in this dataset indicate a comparative difference under the conditions studied rather than a general absence of instructional effectiveness.

The results consequently support a complementary rather than exclusionary pedagogical interpretation. Direct Instruction may be useful when students need explicit modeling of vocabulary, linguistic structures, or comprehension procedures, whereas Discovery Learning may provide a subsequent space for applying those resources through inference, comparison, and integration (Castronova, 2002; de Jong, 2006; Flora et al., 2020; Yaiche, 2021). This sequencing is consistent with broader evidence that scaffolding and explicit strategy support can strengthen reading comprehension (Li et al., 2024; Ueno et al., 2025; Wongdaeng & Higgins, 2023) and that guided inquiry is more effective than unsupported exploration (Lazonder & Harmsen, 2016). In practice, the pedagogical implication is not to replace Direct Instruction with Discovery Learning in all circumstances, but to adjust the amount and timing of guidance to the demands of the reading task and the students' readiness (Budianto et al., 2021; Gijlers & De Jong, 2005).

The findings must, however, be interpreted within the methodological limitations of the study. The sample consisted of only 25 students from one study program at a single institution, which restricts the generalizability of the results. In addition, the documented procedure indicates that Discovery Learning was implemented during the first two meetings and Direct Instruction during the third and fourth meetings. Because the original manuscript does not provide evidence of counterbalancing, randomization of treatment order, or a washout procedure, potential order, practice, or carryover effect cannot be ruled out. The very large effect sizes, particularly for inferential comprehension and integrative interpretation, should therefore be interpreted together with the small sample size, the fixed treatment order, and the repeated-measures structure of the study. As discussed in Section 2.1, the differential pattern across comprehension components provides partial evidence against a purely order-driven explanation, but the possibility cannot be excluded. In addition, the very large *dz* values reflect high consistency in the paired differences across participants rather than necessarily indicating practical significance of the same magnitude as a between-group d of equivalent size. Future research should involve larger samples, multiple classes and institutions, longer instructional periods, counterbalanced or crossover designs, and formal measures of students' English proficiency as a covariate.

Despite these limitations, the study contributes to the development of English for Islamic Studies by showing that differences between instructional approaches are not limited to overall test scores but can be concentrated in particular dimensions of reading comprehension. Discovery Learning was associated with especially strong performance in inferential comprehension and integrative interpretation, indicating that opportunities to investigate textual relationships and construct meaning may be particularly valuable when students are required to engage with complex disciplinary texts. At the same time, evidence from previous studies supports the continued role of explicit instruction as

systematic scaffolding. The central pedagogical implication is therefore not that one approach should completely replace the other, but that the form and timing of instructional guidance should be aligned with the linguistic and cognitive demands of reading Islamic texts in English.

4. CONCLUSION

Discovery Learning produced higher reading-comprehension performance than Direct Instruction among the 25 first-semester Islamic Education students, with the clearest advantages appearing in inferential comprehension and integrative interpretation. These findings indicate that discovery-oriented reading activities are particularly compatible with tasks that require students to connect ideas, infer unstated meaning, integrate information, and construct coherent interpretations of discipline-specific texts. The study contributes to English for Islamic Studies by showing that instructional effects can vary across dimensions of comprehension and by highlighting the importance of aligning the form of instructional guidance with the cognitive demands of the reading task. Direct Instruction nevertheless remains pedagogically relevant as structured scaffolding for vocabulary, linguistic structures, and reading procedures. In practical terms, EFL lecturers in Islamic higher education may consider a sequenced approach: beginning a reading unit with Direct Instruction to pre-teach key vocabulary, model comprehension strategies, and activate relevant Islamic background knowledge, then transitioning to Discovery Learning activities in which students investigate the text through guided questions, peer discussion, and evidence-based interpretation. Such a complementary arrangement would leverage the scaffolding strengths of explicit instruction while providing the active meaning-construction opportunities that appear to benefit higher-order comprehension.

The findings should be interpreted as exploratory evidence because the study involved only 25 students from one program at a single institution and used a fixed treatment order without counterbalancing, randomization, or a washout procedure. Future studies should employ larger and more diverse samples, multiple institutions, longer interventions, and stronger control of treatment order, and they should examine whether a sequenced combination of explicit instruction and discovery-oriented learning produces stable benefits across proficiency levels and different types of Islamic texts in English.

Acknowledgements
The authors acknowledge IAIN Sultan Amai Gorontalo and the first-semester Islamic Education students who participated in the instructional activities reported in this study.

Authors' Contribution
Asna Usman Dilo: Conceptualization, Methodology, Investigation, Data Collection, Formal Analysis, Writing—Original Draft, Writing—Review and Editing, Project Administration. Yuwin Rustam Saleh: Methodology, Validation, Formal Analysis, Writing—Review and Editing. Nur Zaimah binti Ubaidillah: Supervision, Validation, Writing—Review and Editing. All authors approved the final version of the manuscript.

AI Generative Statement
Generative AI tools were used to assist language refinement, literature searching, reference-index verification, structural alignment with the journal template, and formatting during manuscript preparation. Bibliographic details and journal-index information used in the revision were checked against publisher, journal, or recognized indexing records before inclusion. The human authors remain responsible for the research conceptualization, data collection, statistical analysis, interpretation of findings, verification of the final references, and all final academic decisions. The authors take full responsibility for the originality, accuracy, and integrity of the article.

REFERENCES

Budianto, L., Azmi, M., & Putera, A. A. (2021). The implementation of digital storytelling using discovery learning in EFL listening class: middle school students' and teachers' voices. *… a Foreign Language*. http://digilib.iain-palangkaraya.ac.id/id/eprint/4884

Castronova, J. A. (2002). Discovery learning for the 21st century: What is it and how does it compare to traditional learning in effectiveness in the 21st century. *Action Research Exchange*, *1*(1), 1–12.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

de Jong, T. (2006). Scaffolds for scientific discovery learning. *Handling Complexity in Learning Environments: Theory and Research*, 107–128.

Dilo, A. U. (2024). The influence of learning methods and cognitive style on the ability to understand English texts about Islam in students. *Tafkir: Interdisciplinary Journal of Islamic Education, 5*(1), 62–78. https://doi.org/10.31538/tijie.v5i1.777

Erlina, D., Marzulina, L., Holandyah, M., Herizal, & Fajrin, N. (2025). Enhancing EFL reading: Student teachers' metacognitive practices in Islamic higher education. *Edukasi: Jurnal Pendidikan dan Pengajaran, 12*(1), 261–278. https://doi.org/10.19109/wqtnc970

Flora, Raja, P., & Mahpul. (2020). Discovery learning strategy: Integrating think-pair-share and teacher's corrective feedback to enhance students' writing language accuracy. *International Journal of Education and Practice*, *8*(4), 733–745. https://doi.org/10.18488/journal.61.2020.84.733.745

Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *42*(3), 264–282.

Hamada, A., Shimizu, H., Hoshino, Y., Takaki, S., & Ushiro, Y. (2024). Robust evidence for the simple view of second language reading: Secondary meta-analysis of Jeon and Yamashita (2022). *Studies in Second Language Acquisition, 46*(5), 1355–1372. https://doi.org/10.1017/S0272263124000226

Hamidah, I., Fadhly, F. Z., & Sa'diah, S. (2025). Discovery learning for digital reading: Designing a constructive-responsive model and its instructional implications. *Indonesian EFL Journal, 11*(3), 639–648. https://doi.org/10.25134/ieflj.v11i3.57

Iskandar, Talib, A., & Patak, A. A. (2018). An investigation into teachers' implementation of Discovery Learning Model in High Schools. *Seminar Nasional Lembaga Penelitian Universitas Negeri Makassar*.

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Li, J.-T., Tong, F., Irby, B. J., Lara-Alecio, R., & Rivera, H. (2024). The effects of four instructional strategies on English learners' English reading comprehension: A meta-analysis. *Language Teaching Research, 28*(1), 231–252. https://doi.org/10.1177/1362168821994133

Martelletti, D. M., Luzuriaga, M., & Furman, M. (2023). 'What makes you say so?' Metacognition improves the sustained learning of inferential reading skills in English as a second language. *Trends in Neuroscience and Education, 33*, 100213. https://doi.org/10.1016/j.tine.2023.100213

Mekuria, A., Bushisho, E. W., & Wubshet, H. (2024). The effects of reading strategy training on students' reading strategy use and critical reading ability in EFL reading classes. *Cogent Education, 11*(1), 2310444. https://doi.org/10.1080/2331186X.2024.2310444

Mohammadi, R. R., Saeidi, M., & Abdollahi, A. (2023). Modelling the interrelationships among self-regulated learning components, critical thinking and reading comprehension by PLS-SEM: A mixed methods study. *System, 117*, 103120. https://doi.org/10.1016/j.system.2023.103120

Monica, S., Arsyad, S., & Waluyo, B. (2025). English medium instruction in Islamic higher education: Challenges of lecturer readiness in Indonesia. *Social Sciences & Humanities Open, 12*, 101900. https://doi.org/10.1016/j.ssaho.2025.101900

Mouchrif, H., Mokhtari, N., & Benzehaf, B. (2025). Explicit instruction in metacognitive problem-solving reading strategies: Developing reading comprehension and strategy awareness. *Studies in English Language and Education, 12*(3), 1410–1426. https://doi.org/10.24815/siele.v12i3.42597

Muzammil, S., & Suhono, S. (2019). Providing explicit strategy instruction to enhance students' reading comprehension at English for Islamic Studies class. *Jurnal Smart, 5*(2), 81–93. https://doi.org/10.52657/js.v5i2.900

Pan, M., Lai, C., & Guo, K. (2025). Effects of GenAI-empowered interactive support on university EFL students' self-regulated strategy use and engagement in reading. *The Internet and Higher Education, 65*, 100991. https://doi.org/10.1016/j.iheduc.2024.100991

Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The effectiveness of Direct Instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research, 88*(4), 479–507. https://doi.org/10.3102/0034654317751919

Surdyanto, A., & Kurniawan, W. (2020). Developing critical reading module using integrated learning content and language approach. *Studies in English Language and Education*, *7*(1). https://doi.org/10.24815/siele.v7i1.15098

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tighe, E. L., Kaldes, G., & McNamara, D. S. (2023). The role of inferencing in struggling adult readers' comprehension of different texts: A mediation analysis. *Learning and Individual Differences, 102*, 102268. https://doi.org/10.1016/j.lindif.2023.102268

Ueno, S., Ikeda, M., Takeuchi, O., & Teng, M. F. (2025). Impact of reading strategy instruction on improvement of strategy use and reading comprehension: A meta-analysis. *Reading in a Foreign Language, 37*(2), 225–247. https://doi.org/10.64152/10125/67505

Vettori, G., Casado Ledesma, L., Tesone, S., & Tarchi, C. (2024). Key language, cognitive and higher-order skills for L2 reading comprehension of expository texts in English as foreign language students: A systematic review. *Reading and Writing, 37*(9), 2481–2519. https://doi.org/10.1007/s11145-023-10479-3

Wang, X., Zhang, Q., Chen, H., Neitzel, A. J., & Davis, M. H. (2024). The effects of language learning strategy instruction on college students' English achievement and learner autonomy in mainland China: A meta-analysis. *International Journal of Educational Research, 127*, 102442. https://doi.org/10.1016/j.ijer.2024.102442

Widayanto. (2021). The Effectiveness of Discovery Learning Model in Writing Descriptive Text. *IJORER : International Journal of Recent Educational Research*. https://doi.org/10.46245/ijorer.v2i2.92

Wongdaeng, M., & Higgins, S. (2023). Effectiveness of metacognitive interventions in tertiary EFL contexts: Evidence from a systematic review and meta-analysis. *Innovation in Language Learning and Teaching, 17*(4), 795–811. https://doi.org/10.1080/17501229.2022.2146122

Yaiche, W. (2021). Boosting EFL learners critical thinking through guided discovery: A classroom-oriented research on first-year master students. *Arab World English Journal (AWEJ) Volume*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3826506

Zhang, D., & Pérez-Paredes, P. (2024). Chinese EFL learners' use of mobile dictionaries in reading comprehension tasks. *System, 121*, 103221. https://doi.org/10.1016/j.system.2024.103221