



Developing A Dialogic Digital Reading Learning Environment For Indonesian Literacy: Evidence From Annotation, Discussion, and Comprehension Traces

Nursaid*, Mohamad Hafrison, Muhammad Ismail Nasution, Bima Mhd Ghaluh

Department of Indonesian and Regional Language and Literature, Universitas Negeri Padang, Indonesia

ARTICLE INFO

Article History:

Received: 27 June 2026

Accepted: 06 July 2026

Published: 11 July 2026

Keywords:

Digital reading;
Dialogic engagement;
Learning environment;
Social annotation;
Indonesian literacy.

*Corresponding Author:

nursaid@fbs.unp.ac.id

ABSTRACT

Digital reading environments make students' reading processes more visible because they capture annotations, discussion posts, comprehension attempts, and activity logs. However, these traces are often reduced to participation totals rather than used to inform the design of reading pedagogy. This study developed and evaluated a dialogic digital reading learning environment for Indonesian literacy. Using a design-based mixed-methods approach, 100 undergraduate students engaged with genre-based Indonesian texts over six weeks through interactive annotation, threaded discussion, audio support, and comprehension practice. The dataset comprised 1,842 annotations, 1,126 discussion comments, pretest and posttest reading comprehension scores, platform activity logs, and 20 follow-up interviews. The learning environment was organized around five dialogic engagement dimensions: text-grounded noticing, inferential linking, evaluative stance, dialogic uptake, and metacognitive reflection. Coding showed strong interrater reliability (Cohen's kappa = .86). Reading comprehension increased from pretest ($M = 69.84$, $SD = 9.18$) to posttest ($M = 83.72$, $SD = 7.64$), $t(99) = 15.99$, $p < .001$, with a repeated-measures effect size of $d_z = 1.60$. Dialogic uptake and inferential linking were the strongest predictors of comprehension gain, and the regression model explained 65% of the variance. Interview findings indicated that annotations helped students locate textual difficulty, discussion threads supported peer-mediated interpretation, and comprehension tasks prompted students to return to the text rather than respond superficially. The study offers a pedagogical model for designing Indonesian digital reading environments as social, reflective, and evidence-informed learning spaces.

To cite this article:

Nursaid et al. (2026). Developing a dialogic digital reading learning environment for Indonesian literacy: Evidence from annotation, discussion, and comprehension traces. *Pedagogy Review*, 5(3), 301-317. doi: <https://doi.org/10.61436/pedrev/v5i3.pp301-317>

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

■ INTRODUCTION

Reading in higher education increasingly takes place in digital environments. Students encounter texts through screens that combine layout, hyperlinks, audio, comment spaces, search tools, and traces of activity. This matters because digital reading involves attention, navigation, monitoring, evaluation of evidence, and, in many learning settings, public traces of interpretation. Meta-analytic and experimental studies show that reading medium, navigation, and calibration can shape comprehension and reading behavior (Clinton, 2019; Delgado et al., 2018; Singer & Alexander, 2017). Recent work also links on-screen reading to attention under time pressure and to unequal experiences of digital academic reading (Delgado & Salmerón, 2021; Kuhn et al., 2024). From a design-based perspective, the educational value of technology depends on how tools are

embedded in tasks, feedback, and interaction (Anderson & Shattuck, 2012; Wang & Hannafin, 2005).

This issue becomes more specific in a genre-based Indonesian reading environment. Recent work on Indonesian language education has positioned digital and multimodal literacy as part of broader twenty-first-century engagement (Ghaluh et al., 2026). In the learning environment examined here, students read narrative, poetry, factual text, opinion, essay, drama, and scientific writing, each of which places different demands on attention, inference, evaluation, and textual justification. Digital environments can capture aspects of this work through annotations, discussion comments, attempts at comprehension, and activity logs. However, learning analytics research has shown that such traces need to be interpreted through educational constructs

rather than treated as transparent evidence of learning (Ifenthaler & Yau, 2020; Ye & Pennisi, 2022; Zamecnik et al., 2023). In this study, the central concern is therefore not the amount of platform activity, but the kind of reading that becomes visible through that activity.

Social annotation offers one route into this problem. Because comments and highlights are attached to specific textual segments, annotation can show where students notice difficulty, ask questions, connect ideas, or respond to peers. Reviews and empirical studies in higher education report that social annotation can support collaborative learning when students use it to work with textual evidence rather than merely mark participation (Gao, 2013; Novak et al., 2012). Studies in literature and graduate reading contexts also show its relevance for metacognitive reading and text-based participation (Clapp et al., 2021; Wilson et al., 2024). For the present study, this literature matters because annotation is treated not as a stand-alone feature, but as one source of evidence about students' dialogic engagement with Indonesian texts.

Discussion spaces extend annotation by allowing students to develop ideas beyond the margin of a single passage. In collaborative reading, students may test claims, request clarification, compare textual evidence, and regulate the direction of group understanding. Research on collaborative learning and socially shared regulation shows that interaction becomes academically productive when learners have structures that support planning, monitoring, explanation, and uptake (Kielstra et al., 2022; Vandenberg et al., 2021). Work using epistemic network analysis also shows that discourse quality cannot be reduced to the number of contributions, because learning depends on how ideas, evidence, questions, and feedback are connected across interaction (Csanadi et al., 2018; Viberg et al., 2024). In the present study, discussion is therefore examined as a text-linked practice rather than as a separate comment forum.

In this study, dialogic engagement refers to observable interpretive activity in which students question, respond, extend, or revise ideas with others through annotation and discussion, consistent with research on classroom dialogue, collaborative learning, and social annotation (Allred et al., 2025; Howe & Abedin, 2013; Mercer et al., 2019; Novak et al., 2012). Audio is therefore treated more narrowly as multimodal support for rereading rather than as a dialogic space. It is included because studies on multimedia learning and multimodal reading show that coordinated written-spoken input can support attention to

meaning when it is aligned with the task and does not impose unnecessary cognitive load (Moreno & Mayer, 2002; Pellicer-Sánchez et al., 2021).

Comprehension practice provides further evidence. Retrieval-practice research shows that answering questions can strengthen learning when it prompts learners to retrieve, monitor, and correct their understanding (Karpicke & Roediger, 2008; Littrell-Baez et al., 2015; Roediger & Karpicke, 2006). Reading assessment research also shows that item format shapes the construct being measured, as different task types elicit different cognitive processes (Alderson & Banerjee, 2002; Rupp et al., 2006). In this study, comprehension traces are therefore interpreted not as isolated scores, but as evidence to be read alongside annotations, discussion comments, and platform logs.

Taken together, these studies point to a design problem for digital reading pedagogy. Digital reading, social annotation, collaborative discourse, auditory input, retrieval practice, and learning analytics have each been examined in substantial bodies of work. Less clear is how these elements can be organized into a single reading environment in which platform traces are interpreted as evidence of reading quality. This is the gap addressed in the present study: the focus is not only on whether students participate or improve their scores, but on how their annotations, discussion comments, and comprehension traces represent attention, inference, evaluation, and dialogue during reading.

The present study responds to this problem by developing and evaluating a dialogic digital reading learning environment for Indonesian literacy. The environment uses genre-based Indonesian texts and integrates four pedagogical components: interactive annotation, threaded discussion, audio support, and comprehension practice. Rather than testing a single isolated feature, the study examines how these components provide evidence of students' reading engagement and development of comprehension. Dialogic engagement is organized into five dimensions: text-grounded noticing, inferential linking, evaluative stance, dialogic uptake, and metacognitive reflection. These dimensions make it possible to interpret annotations and discussion comments as reading evidence, while comprehension traces provide a complementary indicator of learning. This study is guided by three research questions (RQs):

RQ1. How can a dialogic digital reading learning environment be designed to

support Indonesian literacy through annotation, discussion, audio support, and comprehension practice?

RQ2. What patterns of dialogic engagement appear in students' annotations and discussion comments during the six-week reading cycle?

RQ3. To what extent are dialogic engagement dimensions associated with students' reading comprehension gains?

By addressing these questions, the study contributes to digital reading pedagogy in three ways. Conceptually, it treats platform traces as evidence of dialogic reading rather than as simple participation counts. Pedagogically, it integrates annotation, discussion, audio, and comprehension practice into a single learning environment. Methodologically, it combines coded discourse traces, comprehension data, platform logs, and interview evidence to examine how students read with others in a digital Indonesian literacy environment.

■ LITERATURE REVIEW

Digital reading environments and reading evidence

Digital reading research has moved beyond the simple question of whether screens are better or worse than print. Meta-analytic work shows that reading medium can affect comprehension, especially when readers work with informational texts or must regulate attention and understanding over extended tasks (Clinton, 2019; Delgado et al., 2018; Kong et al., 2018). Mode-effect studies indicate that comprehension and metacomprehension may shift when reading and assessment occur on-screen, suggesting that students' confidence in understanding does not always align with performance (Ben-Yehudah & Eshet-Alkalai, 2021; Halamish & Elbaz, 2020; Singer & Alexander, 2017). Other studies show that digital academic reading can create uneven demands because navigation, time pressure, and reader resources shape how students move through texts (Delgado & Salmerón, 2021; Kuhn et al., 2024; Støle et al., 2020).

This evidence does not, by default, make digital reading pedagogically weak. It shows that digital reading needs design. Reviews of computer-based reading environments indicate that comprehension support is strongest when digital tools are connected to explicit reading processes, feedback, and opportunities for monitoring (ter Beek et al., 2018). Design-based research offers a useful frame here because it treats technology-enhanced learning as a situated arrangement of tasks, tools, feedback, and interaction rather than as a

neutral delivery system (Anderson & Shattuck, 2012; Wang & Hannafin, 2005). From this perspective, the central question is not whether a platform is available, but what kind of reading behavior the platform makes possible and interpretable.

Trace data are important in this discussion because digital environments can record learners' reading behavior. Early work on trace data demonstrated its potential to examine self-regulated learning processes that are difficult to observe from final scores alone (Hadwin et al., 2007). More recent research has emphasized that trace-based measurement requires attention to validity, because logs only become meaningful when they are linked to defensible constructs of learning (Fan et al., 2022). This concern is echoed in learning analytics studies, which show that analytics can support feedback, study success, and self-regulation when interpreted in relation to educational purposes rather than as raw activity counts (Banihashem et al., 2022; Ifenthaler & Yau, 2020; Larrabee Sønderlund et al., 2019).

For this study, digital reading is therefore approached as a designed environment in which student activity becomes meaningful only when it is tied to reading quality. Platform traces are not treated as proof of learning on their own. They are interpreted as partial evidence that needs to be read alongside annotations, discussions, comprehension performance, and students' own accounts of how they read.

Social annotation, discussion, and dialogic engagement

Social annotation has become an important mechanism for making reading visible in digital environments. Reviews of social annotation in higher education describe its value in helping students externalize questions, connect interpretations to specific textual segments, and respond to peers within the space of reading itself (Novak et al., 2012). Empirical studies further show that annotation can support collaborative learning when comments are used to clarify meaning, build on peers' observations, and work with textual evidence (Gao, 2013). Recent studies extend this argument by showing that social annotation can strengthen active learning, engagement, and peer interaction when task design gives students a reason to return to the text (Cui & Wang, 2024; Li et al., 2024).

The pedagogical value of annotation is especially relevant for literature and language learning because interpretation often develops through visible negotiation. Studies in literature education position social annotation as a way of organizing shared attention to textual detail, while research with graduate readers shows that

annotations can reveal metacognitive reading practices often hidden in conventional assignments (Clapp et al., 2021; Wilson et al., 2024). Recent work in English language arts also connects social annotation with dialogic teaching and learning, indicating that annotation can support interpretive talk when students are asked to question, elaborate, and respond rather than merely mark a text (Allred et al., 2025). In Indonesian language education, digital reader-response work has likewise begun to examine how technology can support literary interpretation in the digital age (Nursaid et al., 2025).

Discussion spaces extend this dialogic work beyond the annotated line or paragraph. Research on classroom dialogue shows that academically productive interaction depends on how learners use talk to reason, explain, challenge, and build shared understanding (Howe & Abedin, 2013; Mercer et al., 2019). In computer-supported collaborative learning (CSCL), regulation studies similarly show that collaboration becomes more productive when learners have structures for planning, monitoring, explanation, and uptake (Järvelä & Hadwin, 2013; Järvelä et al., 2016; Kielstra et al., 2022). These perspectives are useful for digital reading because discussion can easily become detached from the text unless the task structure keeps it anchored to evidence and interpretation.

Analytic work on collaborative discourse reinforces this point. Epistemic network analysis studies show that the quality of interaction lies in the connections among ideas, evidence, questions, and feedback, not simply in the number of turns produced (Csanádi et al., 2018; Vandenberg et al., 2021; Viberg et al., 2024). For a reading environment, this distinction matters because a student may comment frequently without developing an interpretation, while another may contribute fewer but more textually consequential responses. The value of discussion is therefore not measured by volume alone, but by how it extends, qualifies, or redirects reading.

In this study, annotation and discussion are treated as connected practices. Annotation provides local evidence of what students notice in the text, while discussion shows how those observations are taken up, challenged, extended, or revised with others. Together, they form the main basis for examining dialogic engagement.

Audio support, comprehension practice, and trace-based interpretation

Within this framework, annotation and discussion provide the main evidence of dialogic engagement because they make

students' questions, responses, elaborations, and revisions visible within a shared reading space (Allred et al., 2025; Gao, 2013; Howe & Abedin, 2013; Mercer et al., 2019; Novak et al., 2012). Audio is positioned differently. It is not coded as a dialogic turn, but as multimodal support that may help students revisit the written text before they produce or revise an interpretation in annotation, discussion, or comprehension practice.

The rationale is not that adding more modes automatically improves reading. Multimedia and cognitive-load research suggests that written and spoken modes can support one another when learners are helped to coordinate information without unnecessary load (Moreno & Mayer, 1999, 2002; Tabbers et al., 2004), but may become less useful when additional modalities are poorly aligned with task demands and processing capacity (Anmarkrud et al., 2019). Research on multimodal input in language learning provides a more specific basis for this decision. Eye-tracking and comprehension studies show that auditory input can influence how learners allocate attention to written texts and process multimodal input during reading (de-la-Peña et al., 2024; Pellicer-Sánchez et al., 2020, 2021). These findings justify placing audio after initial silent reading, where it supports renewed attention to the text rather than replacing reading or peer interaction.

Comprehension practice adds another layer of evidence by asking students to retrieve and test their understanding. Retrieval-practice research has shown that answering questions can support learning when it encourages learners to retrieve, monitor, and revise their understanding (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006; Rowland, 2014). Meta-analytic work further shows that practice testing can support retention and transfer when used as a learning activity rather than only as evaluation (Adesope et al., 2017; Pan & Rickard, 2018). Reviews of learning techniques likewise identify practice testing as one of the more robust strategies for durable learning (Dunlosky et al., 2013). Classroom-oriented research has also emphasized the value of combining retrieval practice with metacognitive activity, particularly when students are expected to reflect on what they know and what still needs to be repaired (Littrell-Baez et al., 2015).

Research on reading comprehension and reading assessment helps clarify why these traces need careful interpretation. Models of comprehension emphasize that readers build meaning by integrating textual information with prior knowledge and inference processes

(Cromley & Azevedo, 2007; Kintsch, 1988). Assessment studies likewise show that task design influences what can be inferred from performance, because item format and cognitive diagnostic choices shape the processes elicited by a reading task (Alderson & Banerjee, 2002; Li et al., 2016; Rupp et al., 2006). In a digital environment, this concern becomes sharper because comprehension scores can be examined alongside process traces such as annotations, discussions, audio use, and activity logs.

The present study therefore places comprehension practice within a wider evidence chain. Audio-use traces indicate opportunities for rereading and revisiting; they are not coded as dialogic turns. Comprehension tasks show what students can retrieve and apply, while annotation and discussion traces show how students moved from noticing to interpretation, uptake, and reflection. This arrangement keeps the concept of dialogic engagement anchored in interaction while still recognizing the role of audio in the reading sequence.

METHOD

Participants and Setting

The participants were 100 undergraduate students enrolled in Indonesian language and literacy-related courses at a public university in West Sumatra, Indonesia. The study used three intact course sections ($n = 34, 33$, and 33) to preserve the normal instructional setting. Because intact sections were used, demographic and academic indicators were

recorded prior to implementation to document comparability in students' reading backgrounds. Eligibility required active enrollment in the relevant semester, completion of at least one university-level reading or text-analysis course, regular access to a smartphone or laptop, no prior formal use of Inspira in a course, and consent to the use of anonymized platform traces for research. All participants completed the pretest, participated in the six-week reading cycle, and completed the posttest. Table 1 reports the participant profile.

Baseline reading equivalence was checked before the six-week cycle. Mean pretest scores were similar across the three intact course sections: course section A ($M = 71.30$, $SD = 9.36$), course section B ($M = 69.19$, $SD = 10.18$), and course section C ($M = 68.99$, $SD = 7.95$), with no statistically significant section difference, $F(2, 97) = 0.65$, $p = .525$. The intact sections were therefore comparable at baseline for implementation purposes, although the study does not claim random assignment or causal equivalence between groups.

Research Design and Procedures

This study used a design-based mixed-methods design to develop and evaluate a dialogic digital reading learning environment for Indonesian literacy. The design was appropriate because the study examined the platform as an integrated pedagogical environment in which reading, annotation, discussion, audio-supported rereading, and comprehension practice formed one

Table 1. Participant profile

Characteristic	Category/statistic	Value
Total participants	Undergraduate students	100
Course sections	Three intact sections	34; 33; 33
Age	$M = 20.50$, $SD = 1.23$; range	18-23
Gender	Female	68
	Male	32
Active semester	Second semester	21
	Fourth semester	43
	Sixth semester	36
Primary study background	Indonesian language/literature education or related courses	100
Prior university reading coursework	Completed at least one reading or text-analysis course	100
Prior formal use of Inspira	No previous course-based use	100
Prior use of digital reading platforms	Low	24
	Moderate	51
	High	25
Weekly digital reading outside the course	$M (SD)$, hours per week	4.21 (1.60)
Device used for platform access	Smartphone only	37
	Laptop/tablet only	18
	Multiple devices	45

instructional sequence. The quantitative strand measured reading comprehension gains and examined how coded engagement traces were associated with those gains. The qualitative strand analyzed annotations, discussion comments, and interviews to explain how students used the platform during reading. The two strands were integrated during interpretation so that score gains, platform traces, and students' accounts could be read as complementary evidence of reading activity and learning development.

The implementation lasted six weeks. Before the reading cycle began, students received a short orientation to Inspira, including how to access texts, use the Hypothesis annotation layer, join discussion threads, listen to audio support, and complete UjiBaca comprehension practice. They then completed the reading comprehension pretest. During Weeks 1-6, students read assigned Indonesian texts across genres and followed a recurring sequence: silent reading, annotation, peer discussion, audio-supported revisiting, and comprehension practice. The instructor monitored the activity but did not grade the number of annotations or comments, because

the study focused on the quality of reading evidence rather than participation volume. At the end of Week 6, students completed the posttest, and selected students participated in interviews.

The course required students to read Indonesian texts across genres. Platform tasks were embedded into regular course activities, while the pretest, posttest, and interviews were administered as research activities outside the graded course assessment. This separation was intended to reduce pressure on students to produce a particular kind of platform behavior.

Digital reading environment

The learning environment was implemented through Inspira (<https://inspiraku.id>). This web-based Indonesian literacy platform provides a curated library of texts across genre families, including story, opinion, factual, drama, poetry, Indonesian for foreign speakers, essay, scholarly, and other materials. In this study, Inspira served as the primary reading space, where students progressed from text selection to reading, annotation, discussion, audio-supported rereading, and comprehension practice. The

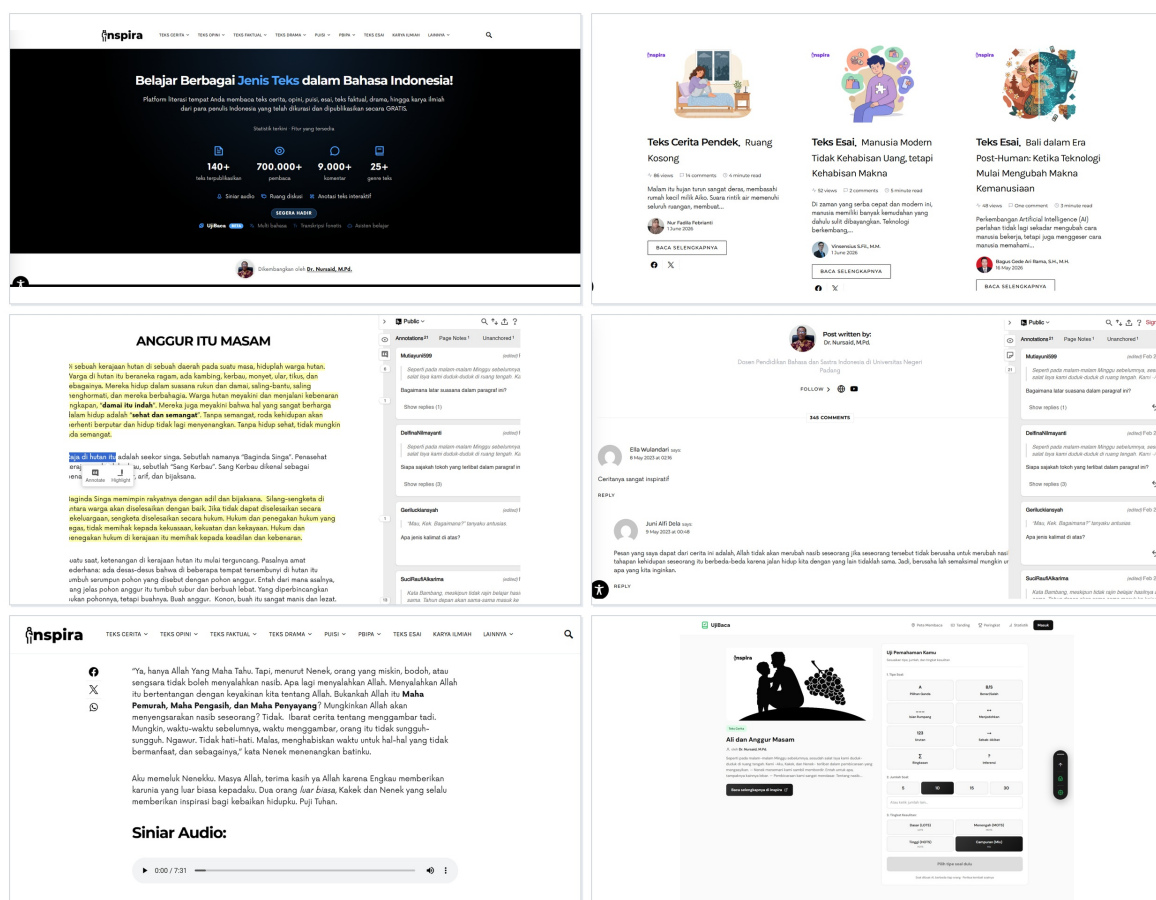


Figure 1. Digital reading environment used in the six-week reading cycle

main platform views used during the six-week reading cycle are shown in Figure 1.

Students entered a text through the library, read it in the platform interface, and used annotation and discussion tools to mark difficulty, ask questions, respond to peers, and elaborate interpretations. Annotation was supported through an integration with Hypothesis (<https://web.hypothes.is>), which allows highlights and comments to be anchored to specific spans of text. This integration was important because it kept student responses attached to textual evidence rather than moving them into a detached discussion forum. Discussion comments were also kept within the reading space so that peer responses could be related to the text being interpreted.

Audio support was placed after initial silent reading. Draft narration scripts and audio files were prepared with an AI-assisted text-to-speech drafting tool (<https://notebooklm.google.com>) and checked by the research team for content fidelity, pacing, and pronunciation before being used in the reading cycle. Students could listen to the audio after silent reading to revisit phrasing, rhythm, emphasis, and continuity of meaning. Comprehension practice was delivered through UjiBaca (<https://ujibaca.inspiraku.id>). This linked practice module draws on the Inspira text library. It allows students to select item type, number of questions, and difficulty level before completing practice tasks based on the selected text. The available formats include multiple-choice, true/false, cloze, matching, sequencing, cause-and-effect, summary, and inference tasks, with difficulty levels organized into lower-order thinking skills (LOTS), middle-order thinking skills (MOTS), higher-order thinking skills (HOTS), and mixed-mode.

In the present study, UjiBaca was analyzed as the comprehension-practice component of the wider dialogic reading environment.

Instruments and Data Sources

The study drew on reading comprehension tests, platform-generated traces, coded annotation and discussion data, and interviews. These sources were selected to keep the analysis aligned with the study's central concern: how students' reading processes became visible through dialogic platform activity and how those processes related to gains in comprehension.

Reading comprehension tests

Reading comprehension was measured using parallel pretest and posttest forms developed for the study. Each form contained 30 items, was scored on a 0-100 scale, and addressed literal comprehension, inference, evaluation, and text-based justification. The forms used texts that were comparable in length and difficulty but did not repeat the platform reading materials. Two Indonesian language education lecturers first reviewed the forms for relevance, clarity, and level appropriacy. The forms were then piloted with 32 students from a comparable course who were not included in the main sample. Item analysis showed acceptable empirical validity: corrected item-total correlations ranged from .36 to .72 for the pretest form and from .35 to .74 for the posttest form, exceeding the critical r value for the pilot sample ($r = .349$, $n = 32$, $p < .05$). Internal consistency was also acceptable, with Cronbach's alpha values of .86 for the pretest form and .88 for the posttest form. Constructed-response items were scored by two raters, and interrater agreement was strong, with an

Table 2. Sample reading comprehension test items

Target process	Item format	Sample item	Scoring focus
Literal comprehension	Multiple choice	<i>Menurut teks, apa alasan utama tokoh menolak ajakan temannya?</i> (According to the text, what is the main reason the character refuses the friend's invitation?)	Identifying explicitly stated information
Inferential comprehension	Short answer	<i>Apa yang dapat disimpulkan dari perubahan sikap tokoh pada paragraf terakhir?</i> (What can be inferred from the character's change of attitude in the final paragraph?)	Drawing meaning from textual cues
Evaluative comprehension	Constructed response	<i>Apakah argumen penulis cukup kuat? Jelaskan dengan bukti dari teks.</i> (Is the author's argument sufficiently strong? Explain using evidence from the text.)	Evaluating claims and using textual evidence
Text-based justification	Constructed response	<i>Pilih satu kalimat yang paling mendukung jawabanmu, lalu jelaskan alasannya.</i> (Choose one sentence that best supports your answer and explain why.)	Linking an interpretation to textual evidence

intraclass correlation coefficient (ICC) = .87. Items with unclear wording or weak discrimination were revised before administration. The tests were administered under the same conditions at the beginning and end of the six-week cycle. Table 2 provides sample item types used in the reading comprehension tests.

Annotation and discussion corpus

The annotation and discussion corpus was exported from Inspira after the implementation. The final dataset contained 1,842 annotations and 1,126 discussion comments. Annotations were treated as local evidence of what students noticed at specific points in a text. In contrast, discussion comments were treated as evidence of how students responded to, extended, or revised interpretations with peers. Both datasets were coded using the dialogic engagement scheme described in the Data analysis subsection.

Platform activity logs

Platform logs recorded text access in Inspira, annotation and discussion activity through the Hypothesis layer, audio use, UjiBaca practice completion, and revisit activity. These logs were not interpreted as direct evidence of learning. They were used to describe patterns of platform use and to normalize activity measures across students with differing total reading volumes.

Semi-structured interviews

Follow-up interviews were conducted with 20 students selected to represent high, moderate, and low levels of platform engagement. The interviews focused on students' reading routines, use of annotation and discussion, experience with audio support, and perceptions of comprehension practice. Interviews lasted approximately 20-30 minutes, were conducted in Indonesian, audio-recorded with permission, and transcribed for thematic

analysis. Excerpts reported in the Results section were translated into English after coding.

Data Analysis

Quantitative analysis proceeded in four stages. First, descriptive statistics were calculated for pretest and posttest reading comprehension scores. Second, coded engagement units were summarized across the five analytic dimensions, and interrater reliability for the double-coded subset was estimated using Cohen's kappa. Third, a paired-samples *t* test was used to examine reading comprehension gain from pretest to posttest, with *dz* reported as the repeated-measures standardized mean difference. The value was calculated by dividing the mean paired difference by the standard deviation of the paired differences (Lakens, 2013). Fourth, regression analysis examined whether the coded dimensions of dialogic engagement predicted comprehension gain. Annotation and discussion counts were normalized by the number of assigned reading tasks to avoid treating activity volume as equivalent to reading quality.

Qualitative analysis focused on annotations, discussion comments, and interviews. Two trained coders segmented annotations and discussion comments into meaning units and coded them using five dimensions: text-grounded noticing, inferential linking, evaluative stance, dialogic uptake, and metacognitive reflection. A 20% subset of the corpus was double-coded to estimate interrater reliability. Disagreements were resolved through discussion, and the coding guide was revised before the remaining data were coded. Interview analysis followed three operational steps adapted from Miles et al. (2014): data reduction, data display, and conclusion drawing/verification. First, the interview transcripts were read repeatedly and reduced into meaning units related to students' use of annotation, discussion, audio support, comprehension practice, and rereading behavior. Second, the reduced data were organized in matrices by participant engagement level and platform component so that patterns could be compared across high-, moderate-, and low-engagement students. Third, preliminary themes were checked against excerpts from annotations, discussion comments, platform logs, and comprehension-practice records. This triangulation was used to confirm that interview claims were supported by other forms of evidence rather than by retrospective accounts alone.

Table 3. Coded dialogic engagement in platform data

Dimension	Coded units	Percentage
Text-grounded noticing	734	24.7
Inferential linking	642	21.6
Evaluative stance	401	13.5
Dialogic uptake	693	23.3
Metacognitive reflection	498	16.8
Total	2,968	100.0

Ethical Considerations

Participation was voluntary, and students were informed that their course standing would not be affected by participation in the study. Consent was obtained before data collection. Platform data were anonymized before analysis, and interview excerpts were coded using participant labels rather than real names. Only aggregated results are reported in the manuscript.

RESULTS AND DISCUSSIONS

Platform traces and dialogic engagement profile (RQ1 and RQ2)

The six-week implementation generated 1,842 annotations and 1,126 discussion comments. After segmentation and coding, the corpus yielded 2,968 coded engagement units. Interrater reliability for the coded subset was strong (Cohen's kappa = .86). Table 3 presents the distribution of coded engagement units across the five analytic dimensions.

The coded profile shows that students used the platform primarily to annotate textual information, build links between ideas, and respond to peers' interpretations. Evaluative stance appeared less frequently than the other dimensions, indicating that explicit judgment of claims and evidence was present but less common than noticing, linking, and uptake.

Reading comprehension gains (RQ3)

Reading comprehension increased from pretest to posttest. Figure 2 displays each student's pretest and posttest scores against a diagonal no-gain line, allowing the distribution of individual score changes to be examined alongside the aggregate gain.

The scatter pattern shows that most participants were located above the diagonal line, indicating higher posttest than pretest scores. The mean score increased from 69.84 ($SD = 9.18$) to 83.72 ($SD = 7.64$), with a mean gain of 13.88 points. The gain was statistically significant, $t(99) = 15.99$, $p < .001$, with a large repeated-measures effect size, $d_z = 1.60$. Given the one-group design, this result is interpreted as an improvement associated with participation in the dialogic digital reading environment.

Figure 3 shows the pretest-posttest shifts for the four target processes represented in the 30-item reading forms: literal comprehension, inferential comprehension, evaluative comprehension, and text-based justification. The process-level profile indicates improvement across all four target processes. The largest gains appeared in evaluative comprehension and inferential comprehension, followed by text-based justification and literal comprehension. This pattern suggests that the

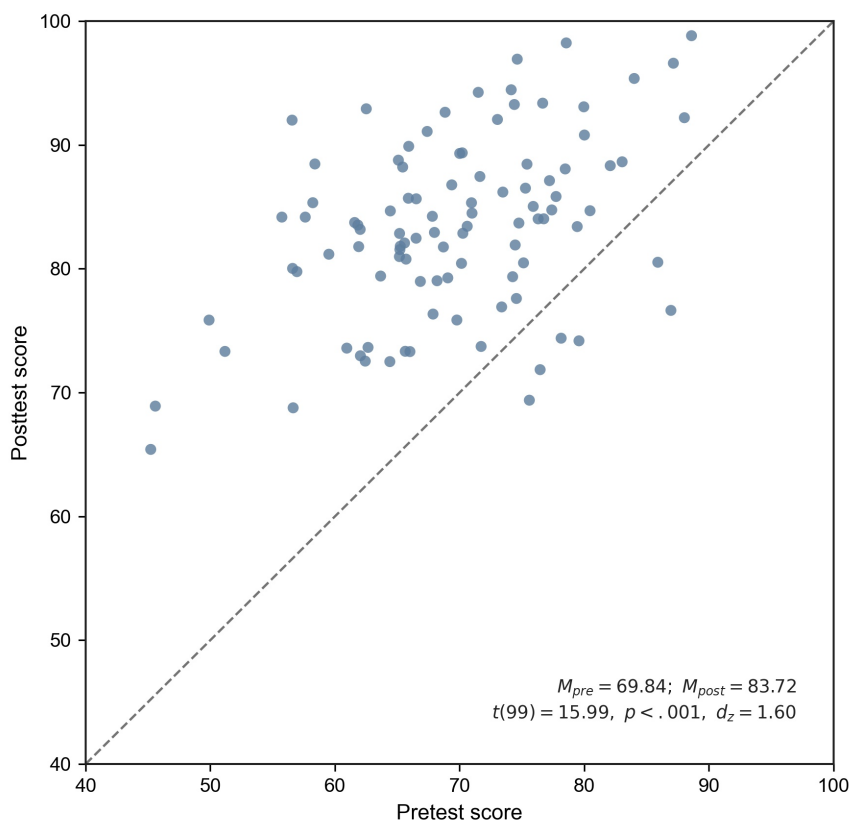


Figure 2. Individual pretest-posttest score distribution

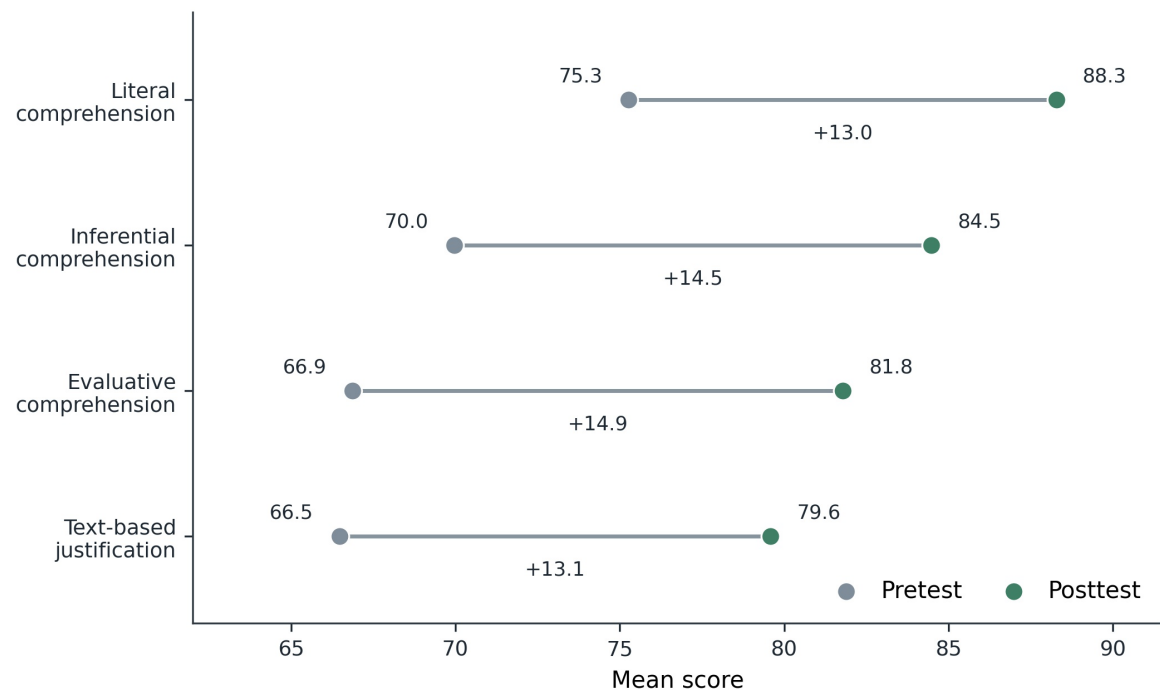


Figure 3. Reading comprehension profile by target process

reading cycle supported not only the retrieval of explicit information but also the interpretive and evidence-based processes emphasized in annotation, discussion, and comprehension practice.

Predictors of comprehension gain (RQ3)

Regression analysis examined whether the five coded engagement dimensions predicted gains in comprehension. The model was significant, $F(5, 94) = 34.41$, $p < .001$, and explained 65% of the variance in gain scores. Table 4 shows that dialogic uptake and inferential linking were the strongest predictors.

Table 4. Regression model predicting reading comprehension gain

Predictor	Standardized beta	<i>t</i>	<i>p</i>
Text-grounded noticing	.18	2.68	.009
Inferential linking	.31	4.67	< .001
Evaluative stance	.23	3.22	.002
Dialogic uptake	.40	5.95	< .001
Metacognitive reflection	.13	1.74	.085
Model summary	$R^2 = .65$; adjusted $R^2 = .63$	$F(5, 94) = 34.41$	< .001

The model suggests that students gained more when their platform activity moved beyond merely marking information toward linking ideas and engaging with peers' interpretations. Evaluative stance and text-grounded noticing also contributed to the model, while metacognitive reflection showed a positive but non-significant association after the other engagement dimensions were entered.

Interview patterns

The interview data explained how students experienced the reading sequence behind the trace patterns reported above. Across the 20 interviews, students described Inspira, annotation, discussion, audio support, and UjiBaca as interconnected parts of a single reading cycle. Their accounts commonly moved from reading to marking, to checking ideas with peers, to revisiting the text, and finally to answering comprehension tasks. The main themes and their relationship to the sequence of platform activity are summarized in Figure 4.

The first theme was annotation as a way of locating textual difficulty. Students described highlighting and short notes as useful because these actions helped them identify the exact part of a text that needed attention. Difficulty became more manageable when it could be attached to a word, sentence, or paragraph rather than described as a general failure to understand the whole text. S07, a moderate-

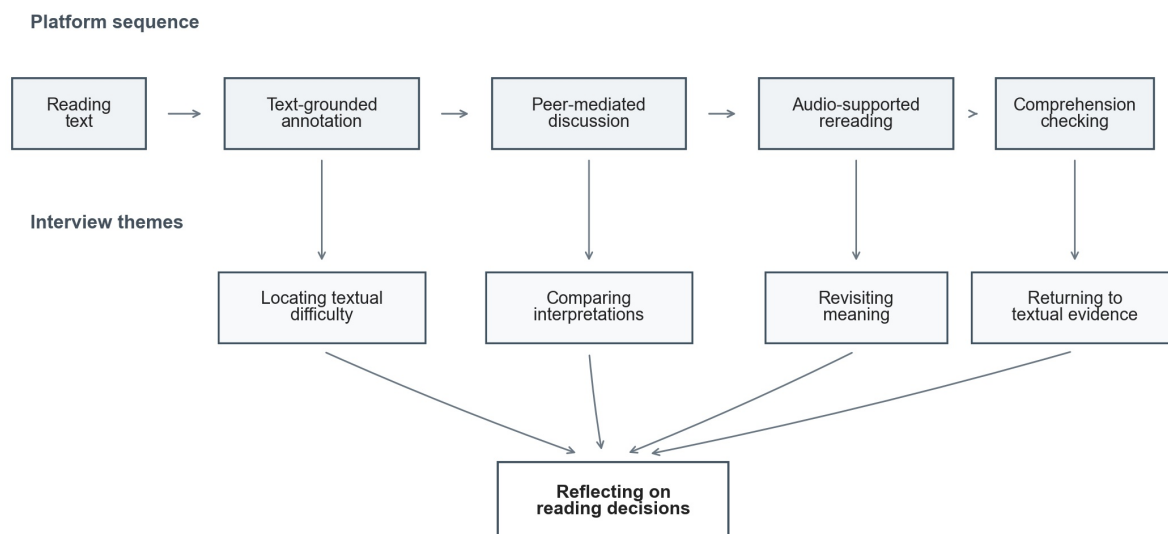


Figure 4. Thematic map of interview findings

engagement participant, explained:

"At first I only said, this text is hard. After I marked the sentence, I knew which part I wanted to ask about." (S07, moderate engagement)

Annotation, therefore, served as a textual anchor for later questioning and discussion.

The second theme concerned discussion as a form of peer-mediated interpretation. Students described discussion threads as useful because they exposed different readings of the same passage and encouraged them to return to the text before agreeing or disagreeing. S12, a high-engagement participant, noted:

"My friend's comment made me check the paragraph again. I did not always agree, but I had to find the sentence that supported my answer." (S12, high engagement)

In this sense, discussion was most useful when it did not replace reading with opinion, but pushed students back to textual evidence.

The third theme involved UjiBaca as a prompt for rereading. Students reported that comprehension practice encouraged them to check the text again before submitting an answer. This was particularly clear among students with lower engagement, who described the tasks as reminders that understanding could not be based solely on memory. S03, a low-engagement participant, said:

"When the questions appeared, I opened the text again. If I answered only from

memory, I was not sure." (S03, low engagement)

These accounts help explain why comprehension practice functioned as part of the reading cycle rather than only as an end-of-task score.

The fourth theme related to audio as support for revisiting meaning. Students described audio as helpful after silent reading, especially when a sentence felt long, dense, or difficult to follow. They did not generally present listening as a shortcut. Instead, they used it as a second pass through the text. S18, a moderate-engagement participant, explained:

"The audio helped after I had read it once. Some sentences became clearer when I heard the pause and then read them again." (S18, moderate engagement)

This pattern suggests that audio worked best when it supported rereading rather than replacing the written text.

Taken together, the interviews show that students experienced the environment as a connected reading sequence. Annotation made difficulty visible, discussion opened interpretation to peer response, audio supported a second encounter with meaning, and UjiBaca prompted a return to the text. One high-engagement participant described this sequence directly:

"I usually read first, mark the part that confuses me, and then look at what my friends wrote. After that, the questions make me go back to the text again." (S15, high engagement)

These accounts give experiential support to the quantitative pattern: students benefited most when visible platform activity became a form of interpretive work.

Designing a sequenced reading environment

The findings show that a dialogic digital reading environment can make students' reading processes visible in ways that go beyond attendance, clicks, or task completion. The trace data did not simply show that students were active on the platform; they showed how students marked difficulty, connected ideas, responded to peers, revisited meaning, and checked comprehension. In relation to the first research question, the learning environment worked as a sequence of reading supports: Inspira provided the genre-based text space, Hypothesis kept students' responses attached to specific textual segments, audio support invited rereading, and UjiBaca returned students to comprehension decisions. Such sequencing matters because research on digital reading has shown that screen-based access alone does not guarantee stronger comprehension; task design has to guide attention, monitoring, and the use of textual evidence (Clinton, 2019; Delgado et al., 2018; Kong et al., 2018). In this study, the platform served as a pedagogical arrangement to make reading activities more visible and accountable.

Interpreting dialogic traces

The second research question concerned the patterns of engagement that appeared in students' annotations and discussion comments. The trace data showed that engagement was not evenly distributed across all forms of reading work. Text-grounded noticing appeared most frequently, followed by dialogic uptake and inferential linking. This pattern is consistent with social annotation research showing that annotation can help learners externalize questions, mark uncertainty, and work with textual evidence (Gao, 2013; Novak et al., 2012; Wilson et al., 2024). However, the present findings add an important qualification. The annotation was valuable, but it was not enough by itself. The stronger pedagogical evidence appeared when students used a marked passage to connect ideas, respond to peers, or revise an interpretation. In other words, annotation became dialogic only when it moved beyond marking into interpretive exchange.

The distinction between frequency and explanatory value is important. Text-grounded noticing was the most frequent coded category and was a significant predictor of gain, but its standardized coefficient was smaller than those

of dialogic uptake and inferential linking. This means that noticing mattered, but it mattered most when it became part of a wider interpretive process. The result addresses the concern that frequent activity is not necessarily the most pedagogically powerful. In this study, the strongest evidence emerged when students used platform traces to link ideas and engage with others' interpretations, not merely when they produced many visible marks.

Comprehension gains and process-level evidence

The third research question shifted from description to association, examining whether the dimensions of dialogic engagement predicted gains in comprehension. The regression results suggest that the answer lies in the quality of interaction rather than the volume of platform use. Dialogic uptake and inferential linking were the strongest predictors of gain, while evaluative stance and text-grounded noticing also contributed significantly. Metacognitive reflection showed a positive but non-significant association after the other engagement dimensions were entered. This result does not necessarily mean that reflection was unimportant. It suggests that, within a six-week implementation, students' explicit reflective comments may have been less immediate as predictors of score gain than the more text-facing actions of linking ideas, responding to peers, and justifying interpretations. It is also possible that the coding scheme captured only visible metacognitive statements in annotations and discussion comments, while quieter forms of monitoring, hesitation, or self-correction remained less observable in the platform traces. Future iterations should therefore include more direct reflective prompts after audio-supported rereading and comprehension practice so that students' monitoring of difficulty, strategy choice, and revision can be captured more sensitively. This pattern fits work on classroom dialogue and computer-supported collaborative learning, where productive learning depends on explanation, uptake, regulation, and the connection of ideas across turns (Howe & Abedin, 2013; Järvelä et al., 2016; Mercer et al., 2019). The contribution of the present study is to show how the dialogic principle operates in digital reading: students gained more when they used visible traces of reading as a basis for interpretation, rather than merely as evidence that they had been active.

The process-level score profile also strengthens the interpretation of the comprehension gain. Improvement appeared not only in literal comprehension, where

students identified explicit information, but also in inferential and evaluative comprehension, where they had to connect ideas and judge textual evidence. This pattern aligns with the platform design because annotation and discussion repeatedly asked students to return to specific parts of the text and justify their interpretations. The gain should still be interpreted with caution because the study did not include a comparison group. However, the individual scatterplot and process-level profile show that the improvement was not limited to a single aggregate score.

Qualitative explanation and pedagogical implications

The interview findings sharpen this interpretation by showing how students understood each component of the sequence. Audio support was useful mainly after silent reading, when students used it to revisit phrasing and the flow of meaning. This is consistent with multimedia learning research, which suggests that combining written and spoken input can support learning when the modes are coordinated rather than redundant or distracting (Moreno & Mayer, 2002; Pellicer-Sánchez et al., 2021). UjiBaca played a different role. It prompted students to return to the text before answering, which connects with retrieval-practice research showing that questions are most useful when they encourage monitoring, checking, and correction (Adesope et al., 2017; Rowland, 2014). The point is not that audio or comprehension questions automatically improve reading. Their value in this study stemmed from their placement within a reading sequence that asked students to revisit, justify, and revise their meaning.

The study also contributes to the interpretation of trace data in digital literacy research. Learning analytics can make learner activity visible, but visibility is not the same as explanation. Prior research has cautioned that trace data require construct validity because clicks, time records, and participation counts do not speak for themselves (Fan et al., 2022; Ifenthaler & Yau, 2020; Ye & Pennisi, 2022). The present study addresses that concern by coding platform traces across five dimensions of dialogic engagement: text-grounded noticing, inferential linking, evaluative stance, dialogic uptake, and metacognitive reflection. This allowed the study to treat annotations, comments, and comprehension attempts as process evidence rather than simply as activity totals.

For Indonesian literacy pedagogy, the findings extend recent work that positions

digital and multimodal literacy as part of broader language-education development in Indonesia (Ghaluh et al., 2026; Nursaid et al., 2025). The specific contribution here is classroom-level evidence about how students work with Indonesian texts across genres when the platform asks them to annotate, discuss, listen, reread, and answer. The study therefore shifts attention from whether students use a digital platform to how the platform organizes reading as social, reflective, and evidence-based work.

Several limitations should be acknowledged. The study used a one-group design, so the comprehension gain cannot be interpreted as causal evidence against a comparison condition. The results show improvement associated with participation in the dialogic digital reading environment, but future research should test the model through a matched comparison or quasi-experimental design. The study also examined one institutional context and one six-week implementation cycle, which limits generalizability. Finally, UjiBaca was examined as one component within the reading environment. Future studies could investigate how annotation, discussion, audio support, and UjiBaca practice contribute individually and in combination across different learner groups, text genres, and instructional settings.

CONCLUSION

This study developed and examined a dialogic digital reading learning environment for Indonesian literacy by integrating genre-based texts, social annotation, threaded discussion, audio support, and comprehension practice into a single reading sequence. The findings show that students' platform activity became pedagogically meaningful when it was interpreted in terms of the quality of their reading work rather than participation counts alone. Annotation helped students identify textual difficulties, discussion supported peer-mediated interpretation, audio encouraged a second encounter with the meaning, and UjiBaca prompted students to return to the text before answering. Reading comprehension improved across the six-week implementation, and dialogic uptake and inferential linking were the strongest predictors of gain, suggesting that students benefited most when they used visible reading traces to connect ideas and respond to others' interpretations. Although the one-group design does not allow causal comparison with another instructional condition, the study offers a practical model for designing Indonesian digital reading environments as social, reflective, and evidence-informed spaces for

literacy learning.

■ REFERENCES

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>
- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35(2), 79–113. <https://doi.org/10.1017/S0261444802001751>
- Allred, J. B., Connors, S. P., & Goering, C. Z. (2025). Social annotation and dialogic teaching and learning in English language arts. *Journal of Adolescent & Adult Literacy*, 68(5), 515–525. <https://doi.org/10.1002/jaal.1382>
- Anderson, T., & Shattuck, J. (2012). Design-based research. *Educational Researcher*, 41(1), 16–25. <https://doi.org/10.3102/0013189X11428813>
- Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist*, 54(2), 61–83. <https://doi.org/10.1080/00461520.2018.1554484>
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. <https://doi.org/10.1016/j.edurev.2022.100489>
- Ben-Yehudah, G., & Eshet-Alkalai, Y. (2021). Print versus digital reading comprehension tests: Does the congruency of study and test medium matter? *British Journal of Educational Technology*, 52(1), 426–440. <https://doi.org/10.1111/bjet.13014>
- Clapp, J., DeCoursey, M., Lee, S. W. S., & Li, K. (2021). "Something fruitful for all of us": Social annotation as a signature pedagogy for literature education. *Arts and Humanities in Higher Education*, 20(3), 295–319. <https://doi.org/10.1177/1474022220915128>
- Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, 42(2), 288–325. <https://doi.org/10.1111/1467-9817.12269>
- Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, 99(2), 311–325. <https://doi.org/10.1037/0022-0663.99.2.311>
- Csanadi, A., Eagan, B., Kollar, I., Shaffer, D. W., & Fischer, F. (2018). When coding-and-counting is not enough: Using epistemic network analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 13(4), 419–438. <https://doi.org/10.1007/s11412-018-9292-z>
- Cui, T., & Wang, J. (2024). Empowering active learning: A social annotation tool for improving student engagement. *British Journal of Educational Technology*, 55(2), 712–730. <https://doi.org/10.1111/bjet.13403>
- Delgado, P., & Salmerón, L. (2021). The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure. *Learning and Instruction*, 71, 101396. <https://doi.org/10.1016/j.learninstruc.2020.101396>
- Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*, 25, 23–38. <https://doi.org/10.1016/j.edurev.2018.09.003>
- de la Peña, C., Chaves-Yuste, B., & Luque-Rojas, M.-J. (2024). Digital reading comprehension: Multimodal and monomodal inputs under debate. *Reading Psychology*, 45(5), 500–518. <https://doi.org/10.1080/02702711.2024.2339809>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*, 17(3), 949–987. <https://doi.org/10.1007/s11409-022-09291-1>
- Gao, F. (2013). A case study of using a social annotation tool to support collaboratively learning. *The Internet and Higher Education*, 17, 76–83. <https://doi.org/10.1016/j.iheduc.2012.11.002>
- Ghaluh, B. M., Wulandari, E., Swatika Sari, E.,

- & Suryaman, M. (2026). Enhancing digital and multimodal literacy in Indonesia's language education: A strategic policy framework for twenty-first-century engagement. *Current Issues in Language Planning*, 1–38. <https://doi.org/10.1080/14664208.2026.2633476>
- Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2(2–3), 107–124. <https://doi.org/10.1007/s11409-007-9016-7>
- Halamish, V., & Elbaz, E. (2020). Children's reading comprehension and metacomprehension on screen versus on paper. *Computers & Education*, 145, 103737. <https://doi.org/10.1016/j.compedu.2019.103737>
- Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, 43(3), 325–356. <https://doi.org/10.1080/0305764X.2013.786024>
- Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4), 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>
- Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, 48(1), 25–39. <https://doi.org/10.1080/00461520.2012.748006>
- Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual- and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning*, 11(3), 263–280. <https://doi.org/10.1007/s11412-016-9238-2>
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Kielstra, J., Molenaar, I., van Steensel, R., & Verhoeven, L. (2022). Supporting socially shared regulation during collaborative task-oriented reading. *International Journal of Computer-Supported Collaborative Learning*, 17(1), 65–105. <https://doi.org/10.1007/s11412-022-09365-x>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>
- Kong, Y., Seo, Y. S., & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A meta-analysis. *Computers & Education*, 123, 138–149. <https://doi.org/10.1016/j.compedu.2018.05.005>
- Kuhn, A., Schwabe, A., Boomgarden, H., Brandl, L., Stocker, G., Lauer, G., Brendel-Kepser, I., & Krause-Wolters, M. (2024). Who gets lost? How digital academic reading impacts equal opportunity in higher education. *New Media & Society*, 26(2), 1034–1055. <https://doi.org/10.1177/14614448211072306>
- Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594–2618. <https://doi.org/10.1111/bjet.12720>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t* tests and ANOVAs. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, 33(3), 391–409. <https://doi.org/10.1177/0265532215590848>
- Littrell-Baez, M. K., Friend, A., Caccamise, D., & Okochi, C. (2015). Using retrieval practice and metacognitive skills to improve content learning. *Journal of Adolescent & Adult Literacy*, 58(8), 682–689. <https://doi.org/10.1002/jaal.420>
- Li, S., Huang, X., Lin, L., & Chen, F. (2024). Exploring influential factors in peer upvoting within social annotation. *British Journal of Educational Technology*, 55(5), 2212–2225. <https://doi.org/10.1111/bjet.13455>
- Mercer, N., Hennessy, S., & Warwick, P. (2019). Dialogue, thinking together and digital technology in the classroom: Some educational implications of a continuing line of inquiry. *International Journal of Educational Research*, 97, 187–199. <https://doi.org/10.1016/j.ijer.2017.08.007>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.
- Moreno, R., & Mayer, R. E. (1999). Cognitive

- principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2), 358–368. <https://doi.org/10.1037/0022-0663.91.2.358>
- Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, 94(1), 156–163. <https://doi.org/10.1037/0022-0663.94.1.156>
- Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *The Internet and Higher Education*, 15(1), 39–49. <https://doi.org/10.1016/j.iheduc.2011.09.002>
- Nursaid, N., Ghaluh, B. M., & Wulandari, E. (2025). An artificial intelligence assistant to reader response theory: Pioneering novel analysis in the digital age. *Language Teaching Research*. <https://doi.org/10.1177/13621688251368636>
- Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, 144(7), 710–756. <https://doi.org/10.1037/bul0000151>
- Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension. *Studies in Second Language Acquisition*, 42(3), 577–598. <https://doi.org/10.1017/S0272263120000091>
- Pellicer-Sánchez, A., Conklin, K., Rodgers, M. P. H., & Parente, F. (2021). The effect of auditory input on multimodal reading comprehension: An examination of adult readers' eye movements. *The Modern Language Journal*, 105(4), 936–956. <https://doi.org/10.1111/modl.12743>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. <https://doi.org/10.1037/a0037559>
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441–474. <https://doi.org/10.1191/0265532206lt337oa>
- Singer, L. M., & Alexander, P. A. (2017). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education*, 85(1), 155–172. <https://doi.org/10.1080/00220973.2016.1143794>
- Støle, H., Mangen, A., & Schwippert, K. (2020). Assessing children's reading comprehension on paper and screen: A mode-effect study. *Computers & Education*, 151, 103861. <https://doi.org/10.1016/j.compedu.2020.103861>
- Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, 74(1), 71–81. <https://doi.org/10.1348/000709904322848824>
- ter Beek, M., Brummer, L., Donker, A. S., & Opdenakker, M. J. (2018). Supporting secondary school students' reading comprehension in computer environments: A systematic review. *Journal of Computer Assisted Learning*, 34(5), 557–566. <https://doi.org/10.1111/jcal.12260>
- Vandenberg, J., Zakaria, Z., Tsan, J., Iwanski, A., Lynch, C., Boyer, K. E., & Wiebe, E. (2021). Prompting collaborative and exploratory discourse: An epistemic network analysis study. *International Journal of Computer-Supported Collaborative Learning*, 16(3), 339–366. <https://doi.org/10.1007/s11412-021-09349-3>
- Viberg, O., Baars, M., Mello, R. F., Weerheim, N., Spikol, D., Bogdan, C., Gasevic, D., & Paas, F. (2024). Exploring the nature of peer feedback: An epistemic network analysis approach. *Journal of Computer Assisted Learning*, 40(6), 2809–2821. <https://doi.org/10.1111/jcal.13035>
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5–23. <https://doi.org/10.1007/bf02504682>
- Wilson, N. S., Dussling, T., Adams, B., Stevens, E., Baumann, J., Yang, S., Smetana, L., Bean-Folkes, J., & Van Wig, A. (2024). What a multi-institutional collective case study of social annotation data reveals about graduate students' metacognitive reading practices. *Literacy*, 58(2), 190–203. <https://doi.org/10.1111/lit.12364>

- Ye, D., & Pennisi, S. (2022). Using trace data to enhance students' self-regulation: A learning analytics perspective. *The Internet and Higher Education*, 54, 100855. <https://doi.org/10.1016/j.iheduc.2022.100855>
- Zamecnik, A., Kovanović, V., Joksimović, S., Grossmann, G., Ladjal, D., Marshall, R., & Pardo, A. (2023). Using online learner trace data to understand the cohesion of teams in higher education. *Journal of Computer Assisted Learning*, 39(6), 1733–1750. <https://doi.org/10.1111/jcal.12829>