

EXAMINING THE RELATIONSHIP BETWEEN PERCEIVED DIFFICULTY, READING SELF-EFFICACY, AND DISCUSSION TEXT COMPREHENSION OF GRADE XII EFL STUDENTS

Syafaruddin Marpaung¹, M. Manugeran², Jumino Suhadi³, Purwanto Siwi⁴

^{1,2,3,4}Doctoral Program in English Studies, Universitas Islam Sumatera Utara, Medan, Indonesia
email: syafaruddinmarpaung37@gmail.com

Abstract

This study investigates Grade XII EFL students' perceived reading difficulty, reading self-efficacy, and comprehension of discussion texts from the government textbook Life Today. Guided by Cognitive Load Theory (to interpret perceived difficulty), Bandura's self-efficacy framework (to examine confidence and persistence), and Barrett's taxonomy (to structure literal, inferential, and evaluative comprehension), the study addresses three research questions: (1) What features of the Life Today discussion texts do students perceive as most difficult? (2) How do students describe their reading self-efficacy when engaging with these texts? and (3) How is comprehension demonstrated across literal, inferential, and evaluative levels? A qualitative descriptive design was employed at SMA Negeri 2 Kota Tanjungbalai, Indonesia. Thirty-six Grade XII students completed Written Comprehension Responses (WCR) aligned with Barrett's levels, and fifteen students were purposively selected through maximum variation sampling for in-depth Think-Aloud Protocol (TAP) sessions and Semi-Structured Interviews (SSI). Data were collected sequentially (TAP–WCR–SSI) and analyzed using thematic analysis with directed coding supported by a transparent codebook and intercoder agreement procedures. Findings indicate that perceived difficulty clusters around tracking pro–contra arguments, processing academic vocabulary, interpreting contrastive cohesion markers, and generating inferences when conclusions are implicit. Students' self-efficacy was relatively strong in persistence but weaker in evaluating arguments and justifying judgments with textual evidence. Comprehension patterns revealed stable literal understanding, whereas inferential and evaluative performance frequently weakened when students struggled to integrate claims, reasons, and evidence across paragraphs. The study highlights the importance of psychologically informed scaffolding such as argument mapping, explicit cohesion instruction, and mastery-oriented feedback to strengthen higher-order comprehension and reading confidence in textbook-based argumentative tasks.

Keywords: *Perceived difficulty; Reading self-efficacy; Discussion text; Think-Aloud Protocol (TAP); Qualitative descriptive; Cognitive Load Theory*

INTRODUCTION

Reading in senior secondary EFL classrooms increasingly requires more than decoding vocabulary and retrieving stated information. Students are expected to interpret arguments, weigh competing

claims, and justify evaluations using textual evidence competencies closely connected to higher-order literacy demands in contemporary schooling (Adeimi & Baumann, 2023; Deane, 2020). In practice,

however, such demands often intensify the cognitive and motivational challenges of reading because learners must coordinate linguistic processing, attention, and strategy use while still meeting curriculum expectations.

In Indonesia's Kurikulum Merdeka (Phase F), English learning outcomes emphasize learners' ability to engage with spoken, written, and multimodal texts and to participate in open-ended discussions using appropriate strategies. Because government textbooks frequently serve as the primary instructional pathway in many schools, the demands embedded in textbook tasks can significantly shape students' learning opportunities and reading outcomes. This is particularly relevant when textbooks require students to construct interpretations and evaluations rather than simply answer literal questions (Francois & Hood, 2021).

One mandated resource in this context is the government EFL textbook *Life Today for Grade XII*, which includes a dedicated unit on discussion texts. Discussion texts are a high-stakes genre because students are expected to comprehend two opposing positions (pro-contra), track stance shifts across paragraphs, and evaluate reasons and conclusions. These genre features can push comprehension beyond surface understanding toward integrative processing, where readers must connect claims, reasons, and evidence and justify judgments with textual support (Adeimi & Baumann, 2023; Strobel, 2023). For EFL learners, difficulty may be amplified when academic vocabulary is dense, cohesion signals are subtle, and warrants or

conclusions are implied rather than explicitly stated.

To examine comprehension demands in a classroom-feasible and pedagogically meaningful way, this study adopts Barrett's taxonomy to structure evidence of understanding across three levels: literal, inferential, and evaluative comprehension. This framework is especially relevant to discussion texts because inferential and evaluative processing is central to the genre's purpose (Kanik Uysal, 2025; Koparan, 2025; Putri, 2025). Using Barrett's levels allows the study to identify not only whether students understand the text, but also where comprehension weakens as task demands increase.

At the same time, comprehension outcomes are influenced by both task demands and learner beliefs. From a cognitive perspective, Cognitive Load Theory (CLT) explains that comprehension becomes fragile when working memory is overloaded by task complexity and by the need to coordinate multiple cues (e.g., vocabulary, cohesion, stance, implied reasoning) during reading (Jiang, 2024). CLT therefore provides a principled lens for interpreting students' perceived difficulty as an indicator of where cognitive demands may exceed available processing resources during engagement with discussion texts.

From a motivational perspective, self-efficacy is a central belief shaping persistence, strategy application, and self-regulation during challenging tasks (Schunk & DiBenedetto, 2021, 2022). In reading, students who believe they can handle difficult texts are more likely to persist, monitor comprehension, and attempt higher-

level reasoning, whereas lower efficacy may limit effort and strategic engagement—especially when tasks require evaluation and justification.

The present study integrates these strands with a tight analytic rationale: CLT is used to interpret what students find difficult in Life Today discussion texts (sources of perceived difficulty), self-efficacy explains how students regulate effort and confidence while facing those demands, and Barrett's taxonomy provides a structured way to examine where comprehension is demonstrated or breaks down across literal, inferential, and evaluative levels. Importantly, these lenses are not treated as separate "theories in parallel," but as complementary tools to explain students' textbook-anchored reading experiences in a single instructional unit.

Despite extensive work on reading comprehension and self-efficacy, a textbook-anchored empirical gap remains in the Indonesian EFL context. Existing studies often examine perceived difficulty, self-efficacy, or comprehension as broad constructs without closely anchoring analysis to a mandated textbook unit that students routinely encounter. Moreover, while many studies use quantitative measures, fewer provide qualitative evidence that captures learners' real-time reasoning and self-regulatory responses while reading argumentative genres. For teachers working with government textbooks, this creates a practical problem: limited psychologically informed evidence on (a) which features of Life Today discussion texts students experience as most difficult, (b) how students describe their confidence and persistence

during reading, and (c) how comprehension manifests across literal, inferential, and evaluative levels in response to textbook-aligned prompts. Without such evidence, pedagogical responses risk becoming generic (e.g., vocabulary translation) rather than targeted to the specific sources of breakdown in discussion-text comprehension.

To address this gap, the present study adopts a qualitative descriptive design situated in the Life Today Grade XII discussion-text unit. Using semi-structured interviews, think-aloud reading sessions, and document analysis of students' written responses to comprehension prompts aligned with Barrett's levels, the study provides a triangulated account of (1) perceived difficulty as interpreted through CLT, (2) reading self-efficacy as articulated through students' descriptions of confidence and persistence, and (3) comprehension evidence across literal, inferential, and evaluative levels. The study aims to generate actionable insights for psychologically informed scaffolding such as argument mapping, explicit cohesion support, inference prompting, and mastery-oriented feedback to strengthen both higher-level comprehension and confidence in textbook-based argumentative reading.

Accordingly, this study addresses the following research questions: 1. What features of the Grade XII Life Today discussion texts do students perceive as most difficult, and how are these difficulties described in cognitive terms? 2. How do Grade XII students describe their reading self-efficacy (e.g., confidence, persistence, and strategy use) when engaging with the Life Today discussion texts? 3. How do

students demonstrate comprehension of the Life Today discussion texts at literal, inferential, and evaluative levels, as reflected in think-aloud accounts and written responses?

METHOD

This study employs a qualitative research method using a literature study approach, which involves collecting, reviewing and analyzing written texts and documents relevant to the topic. The primary objectives are to explore the patterns of action of the revenge and the psychological factors which are shaping the act of revenge in the selected literary works through close reading and interpretive analysis. The data used in this research consists of two types: primary and secondary. The primary data are derived directly from the novels *The Unexpected Guest* and *Five little pigs* by Agatha Christie. These primary texts serve as the main source for analyzing the narrative patterns and characters motivations regarding revenge. The secondary data consist of various supporting literatures such as academic journals, books and other publications that provide theoretical insights and critical perspectives relevant to the analysis.

This study employed a qualitative descriptive design to generate experience-near accounts of Grade XII EFL students' (a) perceived difficulty in reading discussion texts, (b) reading self-efficacy beliefs during engagement with the texts, and (c) comprehension evidence across literal, inferential, and evaluative levels. The design was selected because it enables systematic description of participants' experiences and observable indicators of comprehension

while remaining theoretically informed through directed coding.

The analytic framework integrates three complementary lenses: 1. Cognitive Load Theory (CLT) to interpret perceived difficulty (Jiang, 2024), 2. Self-efficacy theory to interpret confidence and persistence (Schunk & DiBenedetto, 2021, 2022), and 3. Barrett's taxonomy to structure comprehension evidence (Kanık Uysal, 2025; Koparan, 2025; Putri, 2025).

The study was conducted at SMA Negeri 2 Kota Tanjungbalai, Indonesia, where the government-issued textbook Life Today is used as the core instructional resource for Grade XII English. Data collection focused specifically on the discussion-text unit. Participants were 15 Grade XII students selected from a cohort of 36 students through purposive maximum variation sampling. The aim was not statistical representativeness but the inclusion of information-rich cases reflecting diverse reading profiles.

Selection criteria included variation in: 1. Prior English reading performance (high, medium, low based on recent assessment records), and 2. Teacher-observed reading confidence (students who frequently express confidence versus those who often report difficulty). This strategy enhances transferability by ensuring analytic comparison across varied learner profiles and supporting thick description (Thorne, 2025). All participants volunteered and were assigned anonymized codes (P01–P15).

One discussion text (or a short excerpt set) from the Life Today discussion-text unit was selected based on the following criteria: a) explicit pro–contra structure, b)

presence of contrastive cohesion markers (e.g., however, on the other hand), c) inclusion of Grade XII-level academic vocabulary, and d) sufficient length to elicit literal, inferential, and evaluative processing.

All participants worked with the same selected text to ensure procedural consistency. Data were generated through three complementary techniques to enable triangulation: 1. Think-Aloud Protocol (TAP) Sessions. 2. Written Comprehension Responses (WCR). 3. Semi-Structured Interviews (SSI). All instruments were researcher-developed and informally piloted with two non-participating students to ensure clarity. Language and Translation Decisions

To minimize language-related response bias and focus on cognition rather than productive English ability: a) Interviews and WCR prompts were conducted in Bahasa Indonesia. b) TAP verbalizations were allowed in Bahasa Indonesia and/or English. Excerpts originally produced in Bahasa Indonesia were translated into English for reporting and reviewed by a bilingual colleague to ensure semantic equivalence.

Each participant completed one TAP session while reading the selected discussion text. A brief 5–7 minute practice session using a non-study text familiarized participants with verbalizing their thoughts. During the main session, only neutral prompts were used (e.g., “Please continue telling me what you are thinking while reading.”) to maintain verbalization without directing content. Each session lasted approximately 25–40 minutes, was audio-recorded, and transcribed verbatim.

Immediately following the TAP session, participants completed three written

prompts aligned with Barrett’s levels: 1. Literal: Identify the main idea and explicitly stated reasons. 2. Inferential: Draw a plausible conclusion not directly stated. 3. Evaluative: Judge which argument (pro or contra) is stronger and justify using textual evidence. Prompts were presented in Bahasa Indonesia to ensure the task measured comprehension rather than English writing proficiency.

SSI were conducted after TAP and WCR to allow participants to reflect on the specific text they had read. Interviews lasted 20–30 minutes, were audio-recorded, and transcribed verbatim. Interview questions explored: a) perceived sources of difficulty, b) confidence and persistence, c) strategy use, and d) experiences with inferential and evaluative demands.

To ensure procedural reproducibility, the sequence was identical for all participants: 1. Orientation and consent (5–10 minutes). 2. TAP session (25–40 minutes) 3. WCR completion (15–25 minutes) 4. SSI (20–30 minutes). Total time per participant ranged from 65–105 minutes and was conducted in a quiet room at school. Data were analyzed using thematic analysis combined with directed coding (Thorne, 2025). Step 1: Familiarization, All TAP transcripts, SSI transcripts, and WCR documents were read repeatedly. Step 2: Codebook Development. An initial codebook was developed deductively from: CLT-informed difficulty indicators (e.g., vocabulary load, cohesion processing, inference burden), self-efficacy indicators (confidence statements, persistence, help-seeking), and Barrett-aligned comprehension indicators (literal, inferential, evaluative

evidence). Inductive codes were added when new recurring patterns emerged. Step 3: Iterative Coding. Two rounds of coding were conducted: a) Round 1: broad application of the preliminary codebook. b) Round 2: refinement of code boundaries, merging overlaps, and clarification of definitions. All revisions were documented in an audit trail. Step 4: Intercoder Agreement, A second qualitative researcher independently coded 20% of the dataset using the revised codebook. Discrepancies were discussed and resolved through consensus, resulting in a refined final codebook applied consistently across all data. Step 5: Theme Development and Triangulation

Codes were clustered into themes aligned with each research question. Themes were cross-checked across TAP, WCR, and SSI data to confirm convergence or explain divergence. Credibility and transferability were strengthened through: a) method triangulation (TAP, WCR, SSI), b) audit trail documentation, c) peer debriefing and intercoder agreement, and d) thick description of context and participant variation (Thorne, 2025). Participation was voluntary. Students were informed of the study purpose, procedures, and their right to withdraw without penalty. All data were anonymized and stored securely for research purposes only.

FINDING AND DISCUSSION

Findings are reported by research question and are triangulated across three data sources: Semi-Structured Interviews (SSI), Think-Aloud Protocol (TAP) sessions, and Written Comprehension Responses (WCR). Across participants, four recurring difficulty patterns emerged (RQ1), followed

by three self-efficacy patterns (RQ2) and level-based comprehension patterns (RQ3). Because this study was conducted in a single school context, the findings are presented as context-specific patterns that may be transferable to similar EFL textbook settings rather than universal claims.

RQ1. What features of Life Today discussion texts do students perceive as most difficult? Theme 1. Difficulty tracking pro–contra argument structure (argument tracking). Across SSI and TAP, students frequently reported losing track of which side (pro or contra) was being advanced, particularly at contrastive shifts. This pattern also appeared in WCR, where summaries blended or omitted positions. SSI (representative quote): P01 stated, “When it moves into the contra part, I often forget what the pro side said, so it gets mixed.” TAP (processing evidence): P01 verbalized uncertainty at a contrastive marker: “This was pro... then there’s however... so now it’s contra? But I’m confused which reason is the main one.” WCR (aggregated pattern): Several students produced summaries that either merged both sides into one conclusion or represented only one side, indicating breakdown in maintaining stance-based structure.

Triangulated synthesis: Taken together, SSI and TAP suggest that contrast-driven stance shifts increased cognitive coordination demands, while WCR shows the downstream product of that breakdown (mixed or one-sided summaries). Theme 2. Academic and topic-specific vocabulary load disrupts meaning construction. Students described unfamiliar academic vocabulary as interrupting reading flow, pushing them

toward word-by-word translation and repeated rereading. This theme was consistently visible across the three datasets. SSI: P02 explained, “There are many academic words, so I read slowly and sometimes I have to translate first.” TAP: P02 paused at an unfamiliar term: “I don’t know this word... what does it mean... I’ll reread the sentence.” WCR (representative pattern): When key terms were not understood, responses became overly general or conceptually inaccurate (e.g., interpreting an academic term literally), suggesting that vocabulary gaps weakened access to the text’s central claims.

Triangulated synthesis: SSI and TAP show vocabulary as a moment-to-moment processing barrier, while WCR indicates how that barrier can distort or dilute the conceptual content students ultimately report. Theme 3. Interpreting contrastive cohesion markers and references hinders integration across sentences and paragraphs. Students often understood sentences locally but struggled to integrate them into coherent relations across the paragraph (e.g., treating contrast markers as continuation). In TAP, this appeared as rereading earlier paragraphs to confirm what a connector was contrasting. SSI: P04 stated, “I can understand sentence by sentence, but when it becomes a paragraph, I get confused about how the sentences relate.” TAP: P03 questioned the function of a connector: “On the other hand... does this add the same reason or a different one? I’m not sure, so I’ll read the previous paragraph again.” WCR (pattern): In responses, students sometimes presented reasons without showing how contrasts reorganize the argument, suggesting that

cohesion cues were not consistently used to structure comprehension.

Triangulated synthesis: SSI/TAP indicate that difficulty lies not only in knowing connector meanings, but in mapping cohesion to stance structure, which supports the earlier argument-tracking theme. Theme 4. Inferential demands become a major barrier when reasons or conclusions are implicit. A repeated pattern across SSI and TAP was students’ difficulty responding when answers were not explicitly stated. In WCR, inferential prompts frequently elicited literal restatements rather than warranted inferences. SSI: P03 said, “If the answer isn’t written, I’m confused about where to infer it from.” TAP: P03 verbalized uncertainty: “The question asks for a conclusion... but the exact answer isn’t in the text... I’m confused.” WCR (representative pattern): Some inferential answers repeated explicit facts (literal recall) rather than connecting multiple ideas to derive an implied conclusion, indicating a breakdown in inference construction. Triangulated synthesis: The convergence suggests that students often equate comprehension with locating explicit sentences, making inferencing cognitively and motivationally risky (“I’m afraid of being wrong”), especially in discussion texts where implied warrants are common.

RQ2. How do Grade XII students describe their reading self-efficacy when engaging with these texts? Theme 1. Task-specific self-efficacy: higher confidence for literal tasks, lower confidence for inferential–evaluative tasks Across SSI and TAP, students expressed relatively higher confidence when tasks required locating

explicit information, but lower confidence when they had to infer or evaluate. SSI: P02 noted, “When I have to draw a conclusion, I’m afraid of being wrong because there is no clear sentence saying it.” SSI: P05 stated, “Evaluating arguments is hard... both seem correct, and I’m confused which text evidence counts.” Triangulated synthesis: These statements align with RQ1 (inferential/evaluative difficulty) and help explain why higher-level comprehension tasks trigger uncertainty even when literal comprehension is adequate. Theme 2. Persistence is present, but often supported by external help (translation and help-seeking) Many students described persistence through rereading, yet their ability to continue often depended on translation tools or asking peers/teachers. SSI: P04 explained, “I want to continue, but when I’m stuck, I look up word meanings or ask a friend.” TAP: Students’ rereading moves frequently occurred after confusion at stance shifts or cohesion markers, indicating persistence paired with repair strategies.

Triangulated synthesis: Persistence appears as a strength, but the reliance on external support suggests that mastery experiences for autonomous comprehension—especially at higher levels—may still be limited in this context. Theme 3. Strategy-based confidence: students feel more capable when they have a usable repair strategy. Self-efficacy became more observable when students could name or enact strategies (rereading, slowing down, translating key terms, revisiting earlier paragraphs). SSI: Students commonly linked confidence to strategy availability (e.g., translating key words to “make it connect”).

TAP: Strategy use was visible as deliberate rereading and checking earlier paragraphs when coherence was lost. Triangulated synthesis: Confidence is not only a stable trait but is situated in strategy access, which suggests pedagogical leverage through explicit strategy coaching and guided practice (Schunk & DiBenedetto, 2021, 2022).

RQ3. How is comprehension demonstrated across Barrett’s levels? Literal comprehension: relatively stable across students. In WCR, many students could identify the topic, main idea, or explicitly stated reasons with fewer breakdown signals compared to higher-level prompts. TAP verbalizations also showed fewer “confusion” markers when students were retrieving directly stated information. Integrated synthesis: Literal tasks aligned with students’ stronger task-specific self-efficacy and required less cross-paragraph integration, making performance more stable. Inferential comprehension: vulnerable when integration across paragraphs and implied reasoning is required. Inferential comprehension weakened when students needed to integrate claims and reasons across paragraphs, interpret implied warrants, or connect contrastive relations. TAP evidence: Students explicitly noted that answers were “not written exactly,” then hesitated or guessed. WCR pattern: Responses often remained literal (restating explicit facts) rather than constructing a justified inference.

Integrated synthesis: The convergence across TAP and WCR indicates that inferencing requires coordination of cohesion cues, stance tracking, and evidence integration—precisely where perceived

difficulty was highest (RQ1). Evaluative comprehension: the key difficulty is justifying judgments with textual evidence. At the evaluative level, students could sometimes state a preference (pro or contra) but struggled to justify it with relevant textual evidence. SSI: P05 said, “Evaluating arguments is hard... I’m confused which evidence from the text to use.” WCR pattern: Some responses provided opinions without citing reasons/evidence from the text, or treated both sides as equally valid without weighing argument quality.

Integrated synthesis: Evaluative breakdown appears to be less about having an opinion and more about evidence-based justification, suggesting that students need explicit modeling of how to select and use textual evidence in argumentative reading. Across data sources, students’ experiences suggest a coherent pattern: as task demands shift from literal retrieval to cross-paragraph integration, inference construction, and evidence-based evaluation, both perceived difficulty and self-efficacy risk increase, and comprehension becomes more fragile. These findings are specific to the Life Today discussion-text unit in this school context but may be transferable to similar textbook-driven Grade XII EFL settings.

DISCUSSION

The findings suggest that students’ difficulty with Life Today discussion texts goes beyond unfamiliar vocabulary and reflects cognitive overload created by genre-specific demands. Across SSI, TAP, and WCR, students struggled to (a) track pro–contra stance shifts, (b) interpret contrastive cohesion markers, (c) process academic vocabulary, and (d) generate inferences when

reasons or conclusions were implicit. From a Cognitive Load Theory (CLT) perspective, these tasks require students to coordinate multiple information cues simultaneously, which can strain working memory and weaken coherence-building during reading (Jiang, 2024). The observed breakdowns such as losing track of the argument flow at contrastive transitions or rereading earlier paragraphs to re-establish stance are consistent with the idea that comprehension becomes fragile when processing demands exceed available cognitive resources.

In addition, the results indicate that cohesion cues and connectives function as “high-impact” triggers for comprehension disruption. Students often reported understanding sentences locally but struggling to integrate them across the paragraph structure, especially when markers such as *however* and *on the other hand* signaled stance changes. Such integration difficulties matter because discussion texts are structurally organized around competing claims and require readers to maintain a stable mental representation of both sides. When stance tracking collapses, WCR summaries tend to become mixed, one-sided, or overly generalized outcomes that align with work emphasizing the importance of building and justifying interpretations through coherent integration of textual evidence and relations (Deane, 2020). In this sense, the Life Today discussion-text unit is cognitively demanding not only because of lexical difficulty, but because it requires readers to manage contrastive relations and argument structure while constructing a coherent representation.

The findings also show that reading self-efficacy was task-specific rather than global. Students expressed higher confidence when tasks were perceived as “retrievable” (i.e., answers could be located explicitly) and lower confidence when comprehension required inferential or evaluative reasoning. This pattern is consistent with self-efficacy theory, which emphasizes that efficacy beliefs are shaped by perceived control over strategies, actions, and outcomes (Schunk & DiBenedetto, 2021, 2022). In this study, students’ statements and behaviors suggest that inferential and evaluative tasks reduced perceived control because students could not see a clear path from the text to an answer, especially when conclusions had to be constructed rather than located.

At the same time, the dataset indicates meaningful persistence (e.g., rereading and continuing despite confusion), but persistence was often supported by external help such as translation tools or assistance from peers/teachers. This suggests resilience, yet it also implies that students may not consistently experience mastery-oriented control over comprehension, an important source of efficacy strengthening (Schunk & DiBenedetto, 2021, 2022). Therefore, strengthening self-efficacy in discussion-text reading is not only a matter of encouragement but requires explicit, internalizable strategies that students can apply independently when stance tracking and integration become difficult.

Across data sources, comprehension patterns showed relatively stable performance at the literal level but more fragile performance at inferential and evaluative levels. This aligns with Barrett-

based distinctions, where literal comprehension emphasizes retrieval of explicit information, while inferential and evaluative comprehension require integration, implied reasoning, and evidence-based judgment (Kanık Uysal, 2025; Koparan, 2025; Putri, 2025). In the present findings, inferential difficulties commonly occurred when students needed to integrate ideas across paragraphs and construct an implied conclusion. Evaluative difficulties emerged when students expressed an opinion but struggled to justify it with textual evidence indicating that the challenge was not only “having an opinion” but performing evidence-based justification.

These results echo the broader argument that reading, particularly in academically demanding contexts, requires students to build interpretations and justify them using textual support rather than relying on surface-level recall (Deane, 2020). For discussion texts specifically, evaluation depends on identifying claims and reasons, comparing argument strength, and selecting relevant evidence. Without explicit routines, students may default to translation, guessing, or opinion statements unsupported by the text.

Methodologically, the triangulation across SSI, TAP, and WCR strengthens interpretive validity by showing convergence between (a) what students say is difficult (SSI), (b) where processing breaks down in real time (TAP), and (c) how understanding is ultimately represented in products (WCR). This integration helps distinguish vocabulary difficulty as a general claim from the specific moments when vocabulary or cohesion cues disrupt stance tracking, and it clarifies how

those disruptions translate into mixed summaries or weak evidence-based evaluations.

Practical implications: actionable instructional routines grounded in the data To respond directly to the patterns identified across the three data sources, the following instructional routines are recommended for Grade XII textbook-based discussion-text reading: 1. Two-column argument mapping (Pro–Contra Tracker). Routine: While reading, students maintain a two-column chart (Pro vs. Contra) and label each paragraph’s stance (Pro/Contra/Shift). Data link: This targets the most frequent difficulty pattern-losing track of stance shifts-observed in SSI, TAP confusion points, and mixed WCR summaries. 2. Cohesion-and-stance “signal training” (Connectives → Function → Stance). Routine: Teach a small set of contrastive markers (however, on the other hand, nevertheless) using a consistent three-step prompt: Which marker? What function (contrast/addition)? Which stance does it introduce or challenge? Data link: This addresses the recurring difficulty with contrastive cohesion and paragraph integration that triggered rereading and uncertainty in TAP. 3. Inference scaffolds using structured stems (Claim–Reason–Implied Conclusion). Routine: Students complete inference frames such as: “Based on Reason A and Reason B, the implied conclusion is.” Data link: This responds to repeated student difficulty when “the answer is not written,” which led to guessing or literal restatement in WCR. Evidence-based evaluation routine (CER for reading: Claim–Evidence–Reasoning). Routine: For evaluative questions, require

students to write: (a) a claim (which side is stronger), (b) two direct evidence points from the text, and (c) reasoning explaining how the evidence supports the claim. Data link: This directly targets the most persistent evaluative breakdown: inability to identify what counts as evidence and justify judgments textually. Several limitations should be considered when interpreting these findings. First, the study was conducted in one school using a purposive sample, which limits broad generalization; results are best interpreted as context-specific patterns that may be transferable to similar textbook-driven Grade XII EFL settings through thick description rather than universal claims. Second, Think-Aloud Protocol (TAP) procedures may influence reading behavior by slowing reading or prompting additional monitoring; this potential reactivity may affect the frequency of rereading or strategy reporting. Third, because interviews and written prompts were conducted in Bahasa Indonesia, translation of excerpts into English for reporting may introduce subtle shifts in meaning despite careful checking. These limitations do not invalidate the findings but define the boundaries of inference and highlight the need for cautious transferability claims.

Future studies could extend the present work through multi-site qualitative replication, classroom-based intervention studies testing the proposed instructional routines (argument mapping, cohesion signal training, CER-based evaluation), and complementary quantitative profiling to examine how widely the identified patterns occur across different schools and learner profiles.

CONCLUSION

This qualitative descriptive study examined how Grade XII students at SMA Negeri 2 Kota Tanjungbalai experience discussion-text reading in the Life Today textbook through the integrated lenses of Cognitive Load Theory, self-efficacy theory, and Barrett's taxonomy. The findings demonstrate that discussion-text comprehension is cognitively demanding not merely because of vocabulary difficulty, but because students must coordinate stance tracking (pro–contra shifts), interpret contrastive cohesion markers, process academic vocabulary, and construct inferences when reasoning is implicit. When these demands accumulate, coherence building becomes fragile, particularly at inferential and evaluative levels.

It extends Cognitive Load Theory (CLT) to a textbook-anchored EFL genre context by showing how genre-specific structural features especially contrastive stance shifts function as cognitive load amplifiers during real-time reading. This highlights that cognitive load in EFL reading is not only lexical but also structural and relational. Second, the study refines understanding of reading self-efficacy as task-specific and level-sensitive. Students' confidence was relatively stable for literal retrieval but decreased when inferential and evaluative reasoning required evidence-based justification. This supports the view that self-efficacy is shaped by perceived strategic control and task transparency, particularly in argumentative genres. Third, by triangulating SSI, TAP, and WCR data, the study operationalizes Barrett's taxonomy not merely as a testing framework but as a

diagnostic tool for identifying where comprehension breaks down in authentic classroom texts. The decline from literal to inferential–evaluative performance illustrates how higher-order comprehension is especially vulnerable when cognitive coordination and evidence integration are required simultaneously.

The findings point to several actionable instructional implications for textbook-based EFL reading: 1. Structured argument mapping (two-column pro–contra tracking) to stabilize stance representation during reading. 2. Explicit cohesion instruction linking contrastive markers to stance shifts and paragraph integration. 3. Inference scaffolds that guide students to connect multiple textual reasons before drawing conclusions. 4. Evidence-based evaluation routines (e.g., Claim–Evidence–Reasoning structures) to strengthen justified argumentation rather than unsupported opinion.

These routines directly respond to the identified breakdown points in stance tracking, cohesion integration, and evidence-based evaluation, and may simultaneously strengthen higher-level comprehension and mastery-oriented self-efficacy. The conclusions should be interpreted within several boundaries. The study was conducted in a single school context using purposive sampling, which limits broad generalization. Additionally, Think-Aloud Protocol sessions may have influenced reading behavior, and translation from Bahasa Indonesia to English for reporting may have introduced subtle shifts in meaning despite verification procedures. Finally, the qualitative design

prioritizes depth of insight over statistical generalizability.

Future research should therefore: (1) replicate the study across multiple schools and regions to examine transferability of the identified patterns; (2) implement and evaluate classroom interventions based on the proposed routines (e.g., argument mapping and evidence-based evaluation structures) to test their impact longitudinally; and (3) integrate complementary quantitative profiling to examine relationships between perceived cognitive load, self-efficacy, and comprehension performance at scale. Overall, the study underscores that improving discussion-text comprehension in EFL classrooms requires coordinated attention to cognitive demand, strategic scaffolding, and the development of evidence-based evaluative reasoning. By aligning psychological theory with textbook-based classroom practice, the findings offer a grounded framework for strengthening higher-order reading in argumentative genres.

REFERENCE

- Adeimi, S. A., & Baumann, J. (2023). Roles of engagement: Analyzing adolescent students' talk during controversial discussions. *Journal of Adolescent & Adult Literacy*, 67(1), 42–52. <https://doi.org/10.1002/jaal.1289>
- Barber, A. T., & Klauda, S. L. (2020). How reading motivation and engagement enable reading achievement: Policy implications. *Policy Insights from the Behavioral and Brain Sciences*, 7(1), 27–34. <https://doi.org/10.1177/2372732219893385>
- Bellanca, J. A. (2020). *Personalized deeper learning: Blueprints for teaching complex cognitive, social-emotional, and digital skills*. Solution Tree Press.
- Bloms, M. (2023). *Exploring the think-aloud strategy to support inferential comprehension* (Master's thesis). Minot State University.
- Brisk, M. E. (2022). *Engaging students in academic literacies: SFL genre pedagogy for K–8 classrooms*. Routledge.
- Chong, S. W. (2021). University students' perceptions towards using exemplars dialogically to develop evaluative judgement: The case of a high-stakes language test. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), 1–17.
- Claire Acevedo. (2023). Reading to learn: Powerful pedagogy for disciplinary teaching in a high-stakes examination curriculum. *Didacticae*, 14, 30–53. <https://doi.org/10.1344/did.2023.14.30-53>
- Deane, P. (2020). Building and justifying interpretations of texts: A key practice in the English language arts. *ETS Research Report Series*, 2020(1), 1–53. <https://doi.org/10.1002/ets2.12304>
- Francois, C., & Hood, M. (2021). Rethinking the role of reading assessments for real learning. *Journal for Multicultural Education*, 15(3), 299–312. <https://doi.org/10.1108/JME-03-2021-0027>
- Jiang, D. (2024). *Cognitive load theory and foreign language listening comprehension*. Springer.

- Kanık Uysal, P. (2025). The use of taxonomies in constructing text-based reading comprehension questions: Bloom and Barrett taxonomies. *E-Kafkas Journal of Educational Research*, 12(3), 795–816. <https://doi.org/10.30900/kafkasegt.1716522>
- Koparan, B. (2025). An investigation of the effects of different text formats on middle school students' reading comprehension performance. *PLOS ONE*, 20(1), 1–20. <https://doi.org/10.1371/journal.pone.0331786>
- Krismadayanti, A., & Zainil, Y. (2022). The level of students' reading comprehension. *CulinguA*, 3(1), 39–48.
- Phillips Galloway, E., Qin, W., Uccelli, P., & Barr, C. D. (2020). The role of cross-disciplinary academic language skills in disciplinary, source-based writing. *Reading and Writing*, 33(1), 13–44. <https://doi.org/10.1007/s11145-019-09942-x>
- Putri, T. A. W. (2025). The study of reading comprehension questions in Bright an English 2 based on Barrett's taxonomy. *Retain*, 13(1), 1–10.
- Reinhold, F., Hoch, S., Schiepe-Tiska, A., Strohmaier, A. R., & Reiss, K. (2021). Motivational and emotional orientation, engagement, and achievement in mathematics. *Frontiers in Education*, 6, 1–18. <https://doi.org/10.3389/educ.2021.58472>
- Schunk, D. H., & DiBenedetto, M. K. (2021). Self-efficacy and human motivation. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 8, pp. 153–179). Elsevier. <https://doi.org/10.1016/bs.adms.2020.10.001>
- Schunk, D. H., & DiBenedetto, M. K. (2022). Self-efficacy and engaged learners. In A. L. Reschly & S. L. Christenson (Eds.), *Handbook of research on student engagement* (pp. 155–170). Springer. https://doi.org/10.1007/978-3-031-07853-8_8
- Strobel, M. (2023). Collaborative dialogue and deliberative communication: Reading circles with young adult novels and adolescent learners of English as a second language (Doctoral dissertation). University of Glasgow.
- Wei, L. (2023). Leveraging read-aloud instructional strategy (RAIS) for enhancing English reading fluency, anxiety, and confidence of EFL Chinese undergraduates. *European Journal of English Language Teaching*, 8(5), 61–91. <https://doi.org/10.46827/ejel.v8i5.5086>
- Yang, C., & Zhang, L. J. (2023). Think-aloud protocols in second language writing. Springer.
- Zhang, L., Zhang, L. J., & Xu, T. S. (2023). Effects of teacher explicit instruction in stance-taking on students' perceptions of stance and academic writing beliefs. *Frontiers in Psychology*, 14, 1–15. <https://doi.org/10.3389/fpsyg.2023.1140050>
- Zhao, J., Hwang, G.-J., Chang, S.-C., Yang, Q., & Nokkaew, A. (2021). Effects of gamified interactive e-books on

students' flipped learning performance, motivation, and metacognition tendency. Educational Technology Research and Development, 69(6), 3255–3280.

<https://doi.org/10.1007/s11423-021-10053-0>