IDEAS

Journal on Language Teaching and Learning, Linguistics and Literature

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Volume 14, Number 2, December 2026
pp. 9457 - 9477

Copyright © 2026 The Author
IDEAS is licensed under CC-BY-SA 4.0 License



Issued by English study program of UIN Palopo

Revealing Ninth-Grade EFL Students' Reading Engagement Through the Read-STOP-Write Strategy

Dwi Arif Wibowo[1], Rudi Hartono[2], Rahayu Puji Haryanti[3], Intan Permata Hapsari[4]

[1,2,3,4]Universitas Negeri Semarang

Article Info

Received: 2026-10-31
Revised: 2026-11-02
Accepted: 2026-11-07

Keywords:
EFL;
junior high school;
mixed methods;
Read-STOP-Write; reading engagement

DOI:
10.24256/ideas. v14i2. 13056

Corresponding Author:
Dwi Arif Wibowo
dwiarifwibowo@gmail.com
Universitas Negeri Semarang

Abstract

Reading engagement is considered important for learning to read in a foreign language, yet little is known about how lower-secondary learners of English as a foreign language (EFL) engage across the stages of a demanding multistep reading and writing routine such as Read-STOP-Write. This explanatory sequential mixed-methods study examined ninth-grade students' reading engagement during six Read-STOP-Write sessions in an Indonesian junior high school. Questionnaire data came from 64 students in an experimental and a control class (n = 32 each); qualitative data came from narrative observations of the six sessions, interviews with six students, and an interview with their teacher. Self-reported engagement rose in both classes (experimental: 54.28 to 59.84; control: 55.34 to 58.88), but after controlling for pretest scores the classes did not differ at posttest (adjusted difference = 1.84, 95% CI [−3.95, 7.63], p = .528), nor on any of the four engagement dimensions. Individual change was far more variable in the experimental class (gain SD = 15.01 vs. 7.09). Observations and interviews showed volunteering, presenting, rereading, dictionary use, and peer help alongside passivity, fatigue, copying, and confusion, with engagement stronger once students reached writing and presentation. The findings do not show that Read-STOP-Write increased engagement. They indicate that engagement was conditional on task complexity, language demands, scaffolding, time, fatigue, and peer interaction.

1. Introduction

Reading engagement is commonly described as a multidimensional construct that includes behavioral, cognitive, affective, and social involvement in reading(Lee et al., 2021). It matters because it is linked to how well learners comprehend what they read. In EFL settings, engagement has been related to comprehension (Shehzad et al., 2023), and engagement has also been examined as a predictor of struggling readers' comprehension (Cho et al., 2019).

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Among Finnish third and fourth graders, cognitive engagement, particularly the use of response strategies, mediated the link between reading enjoyment and comprehension (Ronimus et al., 2022), and a network analysis of Hong Kong fourth graders placed intrinsic motivation at the center of connections among strategy use and reading habits (Zhang et al., 2026). Engagement is also not fixed. Teacher practices, text difficulty, and classroom interaction shape how learners engage with reading tasks(Abdelhalim, 2017; Protacio, 2017) , and affective engagement in academic reading has been studied among EFL learners (Yulia et al., 2020). Taken together, this work shows that engagement matters for comprehension and responds to instruction, texts, and interaction, but most of the evidence comes from first-language, elementary, or university settings, so it says little about lower-secondary EFL learners working through a complex routine.

These conditions are directly relevant to Read-STOP-Write, a six-stage routine in which students read an informational passage, summarize it, identify its text structure, organize ideas in a matching graphic organizer, plan a paragraph, and write it(Strong, 2020, 2023). The routine was developed and tested mainly in first-language (L1) elementary classrooms. In one development study, students' motivation to read nonfiction declined after the intervention (d = −0.26), possibly because the texts were too difficult, although teachers and students reported positive perceptions of the intervention (Strong et al., 2026). The finding suggests that a routine that supports understanding can also place heavy demands on the people using it. Whether the same is true for EFL learners in lower secondary school, who must manage both the routine and a second language, has not been documented. In principle, each stage gives students a concrete task and a visible product (a summary, an organizer, a paragraph) that can sustain behavioral and cognitive effort, which is why the routine is a plausible candidate for supporting engagement.

The study uses the four-dimension view of engagement (Fredricks et al., 2004; Lee et al., 2021; Li et al., 2026). Behavioral engagement refers to observable participation, effort, persistence, and time spent on reading tasks. Cognitive engagement refers to strategy use, monitoring of understanding, and connecting new information to prior knowledge. Affective engagement refers to emotional responses such as interest, enjoyment, boredom, or anxiety. Social engagement refers to interaction with others in constructing meaning. The EFL Reading Engagement Questionnaire, on which this study's questionnaire is based, was validated with university EFL learners in China and supports a four-factor structure; in that sample emotional engagement was the strongest positive predictor of examination scores, whereas social engagement showed a small negative effect (Li et al., 2026) . This finding is a reminder that the dimensions do not behave alike and that more social activity is not automatically more productive. Because the questionnaire was validated with university learners, it was adapted for ninth-grade students in the present study.

Read-STOP-Write Strategy sequences a complex task into ordered stages and externalizes relations among ideas in organizers. Sequencing and organizing can reduce the number of elements a learner must handle at once (Sweller, 2010), which is one reason the routine might support engagement. The same sequencing, however, adds steps, materials, and time. Strong et al. (2026) attributed the decline in nonfiction reading motivation they observed to text difficulty. For EFL learners, additional language demands may amplify this tension. The present study therefore treats engagement as

something that can rise, stay flat, or fall unevenly across students and activities, and it looks for the conditions under which each occurs. In other words, engagement is treated as conditional on the fit among task demands, learners' language competence, available support, and time, not as a stable trait or a direct product of the strategy.

Questionnaires alone are poorly suited to answering this question. They can show whether reported engagement changes, but they cannot show how engagement appears in the classroom or what enables and limits it. Self-report measures also carry known limitations, which has led researchers to call for observations and interviews alongside questionnaires (Zhai, 2025). Classroom observation has been used to identify how the behavioral, cognitive, affective, and social dimensions of engagement are realized and how they intertwine(Gustafsson et al., 2026). A study that combines questionnaire scores with observation of every lesson and with student and teacher accounts can therefore describe engagement more fully than either source alone. The distinction matters here: the questionnaire asks about English reading in general, whereas the observations and interviews concern engagement during Read-STOP-Write lessons, so the two strands answer related but different questions.

This article is part of a broader explanatory sequential mixed-methods study of the Read-STOP-Write Strategy in an EFL junior high school context. The present article focuses on reading engagement and does not evaluate the strategy's effect on reading comprehension. The gap it addresses is narrower than whether the strategy has been studied in EFL: little is known about how engagement unfolds across the stages of a complex reading and writing routine among lower-secondary EFL learners, and under what conditions it rises or falls. Each of the three research questions corresponds to one component of the analysis:

1. To what extent did students' reading engagement differ between the experimental and control classes after the intervention, controlling for pretest scores?
2. How did students display behavioral, cognitive, affective, and social engagement during the Read-STOP-Write sessions?
3. What conditions supported or constrained students' engagement during the sessions?

2. Method

The study used an explanatory sequential mixed-methods design (Creswell & Plano Clark, 2017). The quantitative strand was a quasi-experimental, non-equivalent control group pretest–posttest design with two intact ninth-grade classes. Engagement questionnaires were completed by both classes. The qualitative strand was conducted only in the experimental class and consisted of classroom observation, student interviews, and a teacher interview. Because no qualitative data were collected in the control class, the study cannot compare how engagement looked in the two classrooms; it can compare their questionnaire totals only. The quantitative results (no clear difference between classes and much greater variability in the experimental class) informed the qualitative strand by guiding the selection of interviewees with contrasting patterns of change, and the qualitative findings were then used to interpret the quantitative pattern (Table 6 and the Discussion).

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

The study took place in a public junior high school in Central Java, Indonesia, in the 2026/2027 academic year. Two intact ninth-grade classes took part (N = 64; 32 per class), and the students were approximately A1–A2 learners on the Common European Framework of Reference. Class IX A was the experimental class and Class IX B the control class; the classes were assigned to conditions on the basis of their results on a previous English test, and the study did not collect other information on pre-existing differences between them. Six students from the experimental class, coded SW01–SW06, were selected purposively on the basis of their pretest–posttest changes in the reading comprehension test and the engagement questionnaire, so as to represent different patterns of change (for example, increases in both, an increase in reading with a decrease in engagement, little change in either, and a large decline).

The experimental class's regular English teacher, who delivered the intervention, was also interviewed. The researcher did not teach and acted as a non-participant observer in the classroom. However, the researcher prepared the lesson plans and teaching materials, discussed the implementation closely with the teacher beforehand, and after each lesson gave the teacher feedback based on the observation sheet, including implementation fidelity.

Table 1. Data sources and the research questions they inform

Source	Content	Used for
Engagement questionnaire	Five-point Likert items adapted from Li et al. (2026); administered before and after the intervention to both classes (N = 64)	RQ1
Observation reports	Six narrative reports (one per session), written by the researcher using a protocol with stage-by-stage implementation check boxes, critical events, student responses, teacher practices, and a reflective summary	RQ2, RQ3
Student interviews	Semi-structured, with six students; the guide included one question for each engagement dimension, plus questions on experience and difficulties	RQ2, RQ3
Teacher interview	Semi-structured, with the English teacher who delivered the intervention	RQ2, RQ3

The questionnaire yielded a total score for each student. It consisted of 19 items adapted from the EFL Reading Engagement Questionnaire (Li et al., 2026), which was developed with university EFL learners and covers four dimensions of reading engagement: behavioral (items 1–4), emotional (items 5–9, called affective in this article), cognitive (items 10–14), and social (items 15–19). The items were translated into Indonesian and adjusted for ninth-grade students; for example, the original item "I read English exam-oriented materials after class, such as CET or TEM" became "I read English exam practice materials such as school exam or English test practice."

The Indonesian version was validated through expert judgment. Each item was rated on a five-point Likert scale, and the questionnaire was administered before and after the intervention to both classes. In the present sample, Cronbach's α for the total score was .92 at pretest and .95 at posttest, and for the four subscales it ranged from .67 to .76 at pretest and from .79 to .85 at posttest. The instrument was not separately piloted or factor-analyzed with lower-secondary learners, which is acknowledged as a limitation. Quantitative analysis focused on the total scores, with dimension scores analyzed as secondary analyses. The observation protocol was narrative and unscored; any proportions in the observation reports were the observer's rough estimates, so this article reports them only as approximate impressions ("most students") and does not treat them as measurements. Interviews were conducted in Indonesian, audio-recorded with consent, transcribed verbatim, and translated into English by the researcher for this article.

The school granted permission, institutional approval was obtained from Universitas Negeri Semarang, parents or guardians gave written consent, and students gave assent. Codes and pseudonyms were used in all records, and students and the teacher are not named in this article. Trustworthiness was supported by member checking, an audit trail with analytic memos, and triangulation across questionnaire, observation, student, and teacher data. The limits of the interview data are discussed under Study Limitations.

The experimental class received six sessions of Read-STOP-Write Strategy instruction between 22 August and 19 September 2026. Each session lasted two lesson periods, which are normally 40 minutes each. Two informational texts were used: *Global Warming* (cause–effect; sessions 1–3) and *Gorillas and Orangutans* (compare–contrast; sessions 4–6). Students used a summary sheet, a text-structure sheet with a cause–effect chart or Venn diagram, and a planning and writing sheet based on the TIDE frame (Topic, Ideas, Details, Ending). The control class studied the same objectives and texts through conventional reading instruction.

The observation reports show that the teacher also used activities that are not Read-STOP-Write Strategy stages: vocabulary ice-breakers (a letter-rearranging game, a "Vocabulary Tree," and a clapping routine), student presentations with audience questions in English using WH-question words, and pronunciation practice before presentations. These additions matter for interpretation. Engagement observed during them cannot be attributed to Read-STOP-Write Strategy itself. For this reason, the study cannot separate the contribution of the strategy from that of these additions, and the findings describe engagement during a lesson package that included Read-STOP-Write, not during the strategy alone.

Quantitative analysis used descriptive statistics and, as the main test for RQ1, an analysis of covariance (ANCOVA) comparing posttest scores between classes with pretest scores as the covariate, which suits intact classes whose pretest scores are not identical. The assumptions were checked: the pretest–posttest slopes did not differ between classes and the residuals were normally distributed, but the residual variances were unequal (Levene's test), so the group effect was also estimated with heteroscedasticity-robust (HC3) standard errors. The same ANCOVA was repeated for each of the four dimension scores as secondary analyses, interpreted descriptively because none approached significance.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

As supplementary analyses, paired-samples *t*-tests described change within each class (with Holm-adjusted *p* values for these two tests), an independent-samples *t*-test compared the classes at pretest, and gain scores (posttest minus pretest) were compared with Welch's *t*-test because the variances differed between classes. Normality was checked with the Shapiro–Wilk test; effect sizes are reported as Cohen's *d* and partial η^2. To describe individual change, each student's gain was classified as an increase, a decrease, or no change, and large changes (10 points or more) were counted. The significance level was set at .05 (two-tailed). Descriptive and *t*-test analyses were run in SPSS 27.

Qualitative material (observation reports and interview transcripts) was coded deductively by the four engagement dimensions and inductively for what supported or constrained engagement, using thematic analysis as a general guide (Braun & Clarke, 2022). Evidence for each dimension was then compared across sources and sessions in a joint display, and contradictory evidence was retained rather than removed. Findings are reported separately from interpretations. Because one researcher coded all material, there was no second coder; analytic memos, an audit trail, and member checking were used instead. Where sources disagreed, the disagreement was kept and shown in the joint display (Table 6) rather than resolved. The two strands were integrated by comparing them in the joint display and in the Discussion, which used the qualitative evidence to interpret the quantitative pattern at the class level; interview codes were not matched to individual questionnaire records.

3. Result

Questionnaire Scores

Table 2 presents total engagement scores. Both classes started at similar levels and both increased.

Table 2. Total engagement questionnaire scores before and after instruction

Class	*n*	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Gain *M* (*SD*)
Experimental	32	54.28 (14.51)	59.84 (17.67)	5.56 (15.01)
Control	32	55.34 (12.41)	58.88 (14.00)	3.53 (7.09)

Note. Totals of the 19-item questionnaire (possible range 19–95). Gain = posttest minus pretest. Adjusted posttest means from the ANCOVA (covariate evaluated at the pretest mean of 54.81): experimental 60.28, control 58.44.

The main comparison used ANCOVA on posttest scores with pretest scores as the covariate. Pretest scores strongly predicted posttest scores ($B = 0.82$, $p < .001$), and the slopes did not differ between classes, $F(1, 60) = 1.47$, $p = .231$. After adjustment, the experimental class's posttest mean (adjusted $M = 60.28$) did not differ significantly from the control class's (adjusted $M = 58.44$): adjusted difference = 1.84, 95% CI [−3.95, 7.63], $F(1, 61) = 0.40$, $p = .528$, partial $\eta^2 = .007$. Residual variances were unequal (Levene's $F = 15.61$, $p < .001$), so the model was re-estimated with robust standard errors, which led to the same conclusion (95% CI [−4.02, 7.70], $p = .533$). The supplementary tests agreed. The classes did not differ at pretest, $t(62) = -0.315$, $p = .754$, or at posttest, $t(62) = 0.243$, $p = .809$ ($d = 0.061$, 95% CI [−0.430, 0.551]), and gains were similar, Welch's $t(44.19) = 0.692$, $p = .492$ ($d = 0.173$, 95% CI [−0.319, 0.663]).

The within-class increases were statistically significant in the experimental class, $t(31) = -2.096$, Holm-adjusted $p = .044$, and in the control class, $t(31) = -2.816$, Holm-adjusted $p = .017$. Both classes therefore improved, but there is no evidence that the experimental class improved more. With 32 students per class the comparison has limited precision (the confidence interval for the adjusted difference runs from about −4 to +8 points), so a non-significant difference should be read as absence of evidence for a difference, not as evidence that none exists.

The same pattern held in each engagement dimension (Table 3). Both classes' means rose in every dimension, but the adjusted posttest differences between classes were small and not significant (behavioral $p = .475$, emotional $p = .570$, cognitive $p = .769$, social $p = .747$; robust estimates agreed). Gains were numerically largest in the social dimension in both classes (experimental 1.81, control 1.47 points), but the classes did not differ more in one dimension than in another, so the questionnaire gives no evidence that the strategy changed any one dimension more than the others. Gain variability was larger in the experimental class in all four dimensions and was statistically unequal only for the social dimension (Levene's test, $p = .007$). The subscales are short, so these results are secondary and should be read with caution.

Table 3. Engagement dimension scores before and after instruction

Dimension	Class	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Adjusted difference [95% CI]
Behavioral (items 1–4; α = .67/.82)	Experimental	11.34 (3.70)	12.69 (3.81)	0.55 [−0.99, 2.10]; *p* = .475
	Control	11.53 (2.96)	12.25 (3.59)	
Emotional (items 5–9; α = .73/.85)	Experimental	14.41 (3.92)	15.88 (5.06)	0.56 [−1.40, 2.53]; *p* = .570
	Control	14.69 (3.19)	15.53 (4.49)	
Cognitive (items 10–14; α = .76/.83)	Experimental	14.34 (4.11)	15.28 (4.78)	0.27 [−1.57, 2.12]; *p* = .769
	Control	14.78 (3.72)	15.28 (3.96)	
Social (items 15–19; α = .76/.79)	Experimental	14.19 (3.89)	16.00 (4.96)	0.28 [−1.46, 2.02]; *p* = .747
	Control	14.34 (4.04)	15.81 (3.29)	

Note. Dimension scores are sums of 4 (behavioral) or 5 items; α = Cronbach's alpha at pretest/posttest. Emotional engagement corresponds to affective engagement in the text. Adjusted difference = experimental minus control, from ANCOVA with the pretest score as covariate.

The spread of gains differed sharply between classes. The standard deviation of experimental gains (15.01) was more than twice that of control gains (7.09), and Levene's test for gain scores was significant, $F = 17.768$, $p < .001$. Table 4 shows what this means for individual students.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Table 4. Direction and size of individual change in total engagement score

Class	Increased	Decreased	No change	Rose ≥ 10 points	Fell ≥ 10 points	Range of gains	Median gain
Experimental (n = 32)	19	13	0	11	8	−18 to +35	4.5
Control (n = 32)	19	11	2	8	1	−10 to +20	2.5

Note. Counts were calculated from students' pretest and posttest total scores.

In both classes most students' totals rose, but the experimental class contained many more large falls: eight students dropped by 10 points or more, compared with one in the control class. The average gain in the experimental class therefore combines substantial increases for some students (11 rose by 10 points or more) with substantial decreases for others. This is a pattern in the data. It does not explain why students changed, and the questionnaire cannot say which activities were responsible. It also asks about English reading in general rather than engagement during Read-STOP-Write Strategy lessons, so it describes change in reported reading engagement, not engagement in the activities themselves. The main quantitative message is therefore that the classes did not differ in average change, whereas the experimental class showed much greater dispersion.

An exploratory look at the data suggests that part of this dispersion may reflect starting points: in the experimental class, students below the median pretest score gained more on average (M = 9.5, SD = 17.0) than those above it (M = 1.6, SD = 12.0), although the correlation between pretest score and gain was weak and not significant (r = −.28, p = .115), and regression toward the mean cannot be excluded. The qualitative evidence below offers possible explanations for the divergence, which the Discussion takes up.

Engagement Across Sessions

The observation reports show that engagement varied with what the class was doing (Table 5). Three patterns stand out: early confusion that required heavy scaffolding (sessions 1 and 2); higher participation once students reached writing and presenting (sessions 3, 5, and 6); and reduced engagement under time pressure and fatigue (sessions 2 and 4).

Table 5. Stages enacted and observed engagement, by session

Session (date, students present)	Text	Stages enacted	Observed engagement
1 (22 Aug, 31)	*Global Warming*	Read, Summarize; Text Structure only orally; Organize, Plan, Write not reached	Active vocabulary talk and read-aloud; almost the whole class struggled to find main ideas; one drowsy student, one off-topic

Session (date, students present)	Text	Stages enacted	Observed engagement
2 (24 Aug, 24)	*Global Warming*	Text Structure, Organize, Plan; Write postponed	About five students volunteered in the ice-breaker; only one or two answered main-idea questions; widespread confusion at the TIDE sheet; sleepiness in the last period
3 (31 Aug, 26)	*Global Warming*	Write completed; presentations	Eight volunteers at the Vocabulary Tree; most focused on writing (observer's estimate); six presenters, audience asked WH-questions; two or three noisy students
4 (5 Sep, 30)	*Gorillas and Orangutans*	Read through Plan; Write limited to two sentences; periods cut to 36 minutes	Most read carefully (observer's estimate); four students dominated oral answers; three sleepy; objection when class ended abruptly
5 (12 Sep, 28)	*Gorillas and Orangutans*	Full cycle: summary review, Venn diagram, plan, paragraph	Students took turns at the board, two before being asked; three started the task early; writer's block for some; one drowsy
6 (19 Sep, 28)	*Gorillas and Orangutans*	Presentations; most stages completed earlier	Eight presenters, some competing to go first; spontaneous peer correction; classmates whispering help; nervousness and reading from drafts

These patterns can be described in more detail. In sessions 1 and 2, engagement was effortful and uneven: students took part in vocabulary talk and followed instructions, but most could not find main ideas without heavy help (session 1) and many said they did not understand the TIDE sheet (session 2; "No. No. No, no."). In session 3 and again in sessions 5 and 6, participation was high once the class had worked through to writing and presenting. In session 4, a shortened lesson and morning fatigue limited both the writing and the closing, and oral participation was concentrated in four students.

Two cautions apply. First, the highest observed participation often came in activities the teacher added (the Vocabulary Tree, audience questions, presentations), so it is not evidence about Read-STOP-Write Strategy alone. Second, these are one observer's impressions of a single class, with no comparable observation of the control class.

All six interviewed students said they enjoyed the lessons, and the teacher described the class as more active than in her usual reading lessons, with students competing to present their work. The observation of session 6, in which some students competed to present first, is consistent with her account. Other parts of her account are only partly borne out. She said students stayed focused on the text, but the observer recorded

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

passivity and drowsiness in sessions 2 and 4. The student accounts also contained qualifications that the one-word endorsement of enjoyment hides, which the RQ2 results below describe.

Engagement Across the Four Dimensions

Evidence for each dimension is presented by source, followed by a joint display (Table 6). Students' quotations were translated from Indonesian.

Behavioral Engagement

Observation. Behavioral engagement appeared as read-aloud, underlining difficult words, and dictionary use in session 1; volunteering at the Vocabulary Tree (eight students, session 3); coming to the board without being asked (session 5); and presenting (six students in session 3, eight in session 6, some competing to go first). Task focus was recorded as high in some lessons (most students reading carefully in session 4 and writing in session 3, by the observer's approximate judgment). Off-task behavior was also recorded in nearly every session: drowsiness (sessions 1, 2, 4, 5), off-topic talk, playing with a pencil or a desk, and shouting (session 3). Attendance ranged from 24 to 31 of 32 students, so not every student experienced every stage. In session 2 only one or two students answered the teacher's questions, and in session 4 four students dominated oral answers.

Students. All six students said they tried to complete the activities, but their reasons were mostly about finishing. SW01 said, "*Agar cepat selesai*" ("So that it finishes quickly"), SW03 wanted to finish on time, and SW06 said, "*Karena harus itu*" ("Because I have to"). Only SW02 gave a curiosity-related reason, saying they wanted to know all of the questions. Students' focus was not constant. SW02 said "*kadang iya, kadang enggak*" ("sometimes yes, sometimes no") and described waiting for a noisy class to quiet down or forcing themselves to read. SW04 said they were sometimes disturbed by a seatmate and responded with annoyance but no action.

Teacher. The teacher said students were more active and focused than usual and "*saling berebut*" (competed) to present their work. Participation should not be equated with engagement here: volunteering, finishing tasks, and presenting can reflect a wish to complete the task or to be seen as well as curiosity or cognitive investment, and the present data cannot always tell these apart.

Cognitive Engagement

Observation. Cognitive effort was visible in both success and struggle. In session 1 almost the whole class had trouble finding main ideas and relied on the teacher's scaffolding. In session 2, students had difficulty mapping information into the TIDE sheet. In sessions 3 and 5, some students faced writer's block when asked to turn points into a paragraph in their own words. Yet students also worked out text structure themselves: in session 4 they identified *compare and contrast* from the title and introduction, and in session 5 they sorted similarities and differences into the correct parts of the Venn diagram. In session 6, presenters stressed similarities and differences when speaking.

Students. Students described a range of strategies and difficulties. SW02 reread texts "*tiga sampai lima kali*" (three to five times) and tackled writing by taking words from the text and combining them with known English words. SW05 reread repeatedly. SW03 used a dictionary but, when finding the main idea was hard, "*ngasal*" (just guessed). SW06

managed writing by "*memahami dulu, baru menuliskan*" (understanding first, then writing). SW01 consulted the teacher's notes. SW04 reported general mental strain: "*Selalu diberikan tugas, selalu pusing mikirnya*" ("Whenever I am given a task, I always get dizzy from thinking").

Teacher. The teacher said students found main ideas and text structures easier and reported a slight change in critical thinking, a claim she made cautiously. Observation partly supports the first point (sessions 4–6) but also shows how heavily it depended on scaffolding in sessions 1 and 2.

Affective Engagement

Observation. Positive affect appeared as enthusiasm in ice-breakers (sessions 2 and 3), eagerness at the board (session 5), and relaxed, enthusiastic presentations (session 6). Negative or mixed affect also appeared: collective complaints at the TIDE sheet (session 2), an objection ("Oh my God") when the lesson ended abruptly (session 4), and nervousness when presenting (session 6).

Students. All six said they enjoyed the lessons, and SW06 added that they felt calm. Enjoyment coexisted with boredom and confusion. SW01 was bored when asked to explain the text, SW02 when the lesson ran long ("*jamnya terlalu lama*"), and SW03 when searching for paragraphs. SW05 said there were too many stages, which caused confusion. SW06 said the instructions were sometimes confusing and sometimes clear. These reports must be read with the limits of the interviews described under Study Limitations, because several enjoyment statements were agreements with yes/no questions.

Teacher. The teacher saw increased interest and confidence when students were asked to present.

Social Engagement

Observation. Peer interaction took several forms. In session 1 a student with a larger vocabulary became a spontaneous source of word meanings for classmates. In session 2, confused students asked classmates who understood, and peers corrected a spelling error in the ice-breaker. In sessions 3 and 6, presentations were followed by audience questions, and in session 6 classmates corrected a mispronunciation and whispered a forgotten word. In sessions 4 and 5 work was mostly individual, and the only peer interaction recorded in session 5 was borrowing writing materials.

Students. All six reported some discussion with classmates, mostly about word meanings (SW01, SW02), main ideas or paragraphs (SW03), and supporting details (SW05, SW06). Peer help varied in quality. SW03 said that when reading was hard, "*Mencontek teman*" (copying from a friend), and SW04 said a classmate sometimes gave the answer.

Teacher. She reported that students exchanged ideas to understand the text, but also that some students were reluctant to ask classmates for help. High interaction was therefore not the same as productive engagement: help with word meanings and peer correction support meaning-making, whereas copying or being handed answers does not, even though all of them appear as peer interaction.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Table 6. Joint display of evidence on the four dimensions

Dimension	Where sources converge	Where they diverge or call for caution
Behavioral	Volunteering, presenting, and finishing tasks (observation, teacher, students)	Students' reasons were mainly about finishing; teacher said "focused," but the observer recorded drowsiness and passivity in several sessions
Cognitive	Rereading, dictionary use, and notes (students); structure and Venn work (observation)	Main-idea, TIDE, and own-words tasks needed heavy scaffolding or produced guessing; SW04 reported general mental strain
Affective	Enjoyment and calm (students), enthusiasm in activities (observation), confidence in presenting (teacher)	Boredom, confusion, complaints, and nervousness appeared in all three sources
Social	Peer help with word meanings and supporting details; audience questions (observation, students)	Copying and being given answers (students); the teacher's "reluctant to ask" does not match the peer help seen in sessions 1–3 and 6, though sessions 4–5 were individual work

Across the four dimensions, the evidence points to interaction rather than independent categories. Confusion at the Summarize and TIDE stages limited cognitive work and, at the same time, dampened participation and affect in sessions 1 and 2, whereas success at the Venn diagram or in presentations coincided with eager participation, peer help, and enthusiasm in sessions 5 and 6. Participation (the most visible behavioral and social evidence) did not always coincide with cognitive investment or positive feeling, which the interviews made clear.

Conditions Supporting and Constraining Engagement

Table 7 lists conditions that students, the teacher, or the observation reports linked to engagement. Each condition is labeled by where it originates (the strategy itself, teacher practice, classroom conditions, student characteristics, or external scheduling), because only some of them belong to Read-STOP-Write itself.

Table 7. Conditions that supported or constrained engagement

Condition	Evidence	Source
Supports		
Teacher scaffolding (*teacher practice*)	Moving between desks, guided questions, splitting tasks by paragraph; SW01 asked the teacher when stuck; the teacher gave extra attention to struggling students	Observation, student, teacher
Visible structure and organizers (*strategy*)	The Venn diagram and cause–effect chart made structure concrete; students sorted points correctly in sessions 4–5	Observation

Condition	Evidence	Source
Peers (*classroom condition*)	Word meanings, supporting details, peer correction, audience support	Observation, students
Active formats (*teacher-added activities*)	Ice-breakers, board work, presentations with audience questions; bilingual presentation allowed when students seemed burdened	Observation
Repeated reading (*teacher practice*)	Three readings before lessons are routine for this teacher	Teacher, SW02, SW05
Constraints		
Time and lesson length (*external scheduling*)	Summarize overran in session 1; ice-breaker and repeated explanations in session 2; periods cut to 36 minutes in session 4; SW01 wanted more time ("*Banyak yang belum selesai*"); SW02 found the lesson too long	Observation, students
Main-idea, summary, and TIDE demands (*strategy*)	Almost the whole class struggled (session 1) and many did not understand the TIDE sheet (session 2); SW03 guessed; SW04 was confused by the worksheet columns	Observation, students
Number of stages (*strategy*)	SW05: "*Tahapannya terlalu banyak*" ("There are too many stages")	Student
Vocabulary and writing in English (*student characteristic*)	Own-words writing was hard for several students; the teacher linked this to limited vocabulary	Observation, students, teacher
Fatigue and timetable (*external scheduling*)	Last-period tiredness (session 2); early-morning sleepiness (sessions 4, 5)	Observation
Noise and distraction (*classroom condition*)	Off-topic talk and shouting (session 3); SW02 and SW04 reported being distracted	Observation, students
Attendance (*external scheduling*)	Between 24 and 31 students present	Observation

Students and the teacher focused on the usefulness of structure and scaffolding, whereas the observation reports added conditions external to the strategy, such as time cuts, the timetable, and absence. Only the demands of the Summarize, Organize, and TIDE stages and the number of stages belong to the strategy itself; the other conditions arise from teaching practice, the classroom, students' language proficiency, or scheduling, so the evidence does not support attributing the engagement patterns to Read-STOP-Write alone. The observer's own recommendations were a short mini-lesson on finding main ideas before the Summarize stage, a fully worked example of each organizer, shorter texts so the whole cycle can be completed in one lesson, and small-group work on Organize and Plan. These are the observer's suggestions, not findings.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

4. Discussion

This study set out to describe how reading engagement appeared when ninth-grade EFL students and their teacher used Read-STOP-Write. Its central finding is that the strategy did not produce a statistically superior increase in overall self-reported engagement, but its implementation showed that engagement was highly conditional on task demands, language demands, scaffolding, time, fatigue, and interaction. Total questionnaire scores rose in both classes without a difference between them once pretest scores were considered, and individual change was much more varied in the experimental class. The observer recorded engagement in all four dimensions, high in some lessons and low in others, and students and the teacher described both enjoyment and strain. The remainder of this section interprets these findings and marks where the evidence does not reach.

Engagement Was Conditional and Uneven

The most consistent pattern across sources is that engagement depended on the conditions of the lesson. Observed participation was highest in lessons where students could complete tasks they could manage and then show the product (sessions 3, 5, and 6) and lowest when the task was unclear, time was cut, or students were tired (sessions 1, 2, and 4). One reading of this is that students engaged when the demands of a stage matched what they could do with the support available, and disengaged when they did not. This is an interpretation. The questionnaire cannot test it, and the observation covers a single class. Large-scale evidence points in a similar direction: among 15-year-olds in 24 OECD countries, perceiving reading as difficult was negatively associated with digital reading performance, whereas reading enjoyment and perceived reading competence were positively associated with it (Wang & Hu, 2025). A clear purpose for reading has also been shown to prompt more rereading and more careful processing, particularly for second-language readers with lower working memory capacity (Özata & Rızaoğlu, 2026) . The routine of reading a text three times before lessons (Table 7) may serve a similar function, although the present data cannot show this.

These patterns suggest treating engagement as conditional rather than as a stable characteristic of students: it appeared to fluctuate with the interaction among learner competence, task complexity, instructional support, time, the classroom environment, and opportunities for meaningful participation. This is a conceptual interpretation offered from one classroom, not a tested model, but it organizes the findings and fits the introduction's view of engagement as something that can rise, stay flat, or fall.

Why Individual Change Varied More in the Experimental Class

The experimental class's gain scores were far more dispersed than the control class's, with eight students falling by 10 points or more and eleven rising by that amount. The qualitative evidence suggests a plausible reading: a structured routine may help some students, for whom organizers made structure concrete and gave them tasks they could manage (sessions 4 and 5; SW02's rereading and word-combining), while burdening others, who met many stages, vocabulary demands, and a writing requirement at once (SW05's "too many stages," SW04's mental strain, SW03's guessing).

The larger spread appeared in all four dimensions and was statistically clearest in the social dimension (Levene's test, $p = .007$), although short subscales limit what can be concluded. Starting points may also matter, since students with lower pretest scores tended to gain more, and regression toward the mean may contribute. Because interview codes were not linked to individual questionnaire records, and the questionnaire measures general reading engagement, these explanations remain hypotheses about the class as a whole and cannot be matched to particular students.

Sources Told Different Stories

All six students said they enjoyed the lessons and the teacher said students were focused, yet the observer recorded drowsiness, passivity, noise, and boredom, and students themselves described boredom and confusion when probed. (Zhai, 2025) recommended alternative measures such as observations or interviews rather than relying on self-report alone, and these divergences show why. The questionnaire total, a student's agreement that they enjoyed the lessons, and what an observer sees are different kinds of evidence about engagement, and they need not agree. Each source captures something different: the questionnaire records students' general perceptions of their reading engagement, observation records visible behavior in lessons, student interviews record retrospective perceptions, and the teacher interview records a professional interpretation. Disagreement among them therefore does not mean that one source is wrong; it shows that engagement has aspects that no single instrument captures.

Context Mattered as Much as the Strategy

Engagement is shaped by teacher practices, text difficulty, and classroom interaction(Abdelhalim, 2017; Protacio, 2017). The present data are consistent with that view. The teacher's scaffolding appears in every session, the lessons were affected by a shortened timetable and by last-period fatigue, and several of the liveliest moments occurred in activities the teacher added. Engagement here cannot be attributed to Read-STOP-Write Strategy alone. A further distinction concerns implementation. The teacher delivered lessons planned and prepared with the researcher, who also gave feedback after each lesson, so what was observed is this teacher's implementation, supported by that feedback, and the present evidence cannot show whether a different implementation of Read-STOP-Write would have produced different engagement.

The Routine's Demands Are Visible in the Data

(Strong et al., 2026) reported a decline in motivation to read nonfiction after Read-STOP-Write Strategy and attributed it to text difficulty. The present questionnaire totals did not decline, but the measures differ and the comparison should not be pressed. What the present data do show are the kinds of strain that concern raised: a Summarize stage that consumed most of the first lesson, a TIDE sheet that most students said they did not understand when first introduced, a student who said there were too many stages, and a student who described being dizzy from thinking. Sequencing and organizers may reduce the number of elements a learner handles at once (Sweller, 2010), but this depends on how the organizers are introduced.

Working memory also contributes more to reading performance as tasks move from words to sentences to whole texts (Escobar et al., 2025), which fits the strain students reported at the Summarize and Organize stages, although that study involved Spanish-reading fourth graders in Chile and did not examine instruction. The observer's

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

recommendation of a worked example before the TIDE sheet, and the improvement seen by sessions 4 and 5 once students had met the organizers, are consistent with this idea, though the study did not test it. Cognitive load was not measured in this study, so it is offered as a possible interpretation, not as an established mechanism.

Motives for Engaging

Most students gave reasons about finishing tasks, such as finishing quickly or because they had to. Such motives need not mean shallow engagement: among Saudi EFL students in a university preparatory year, extrinsic motivation was the strongest predictor of reading strategy use and comprehension (Kassem & Alqahtani, 2023). The present study did not measure motivation type or comprehension in this class, so it cannot say whether completion-oriented motives led to deeper processing here.

Language Demands and Scaffolding

Difficulty with writing in one's own words, which the teacher linked to limited vocabulary, resembles findings from Ethiopian secondary EFL classrooms, where students leaned on rote repetition and dictionary lookup and where vocabulary learning strategies and motivation were associated with reading comprehension(Taye & Lodebo, 2026). The bilingual presentations the teacher allowed when students seemed burdened also find some support: among Chinese university students, first-language questions served as cognitive scaffolds and showed no sign of interfering with second-language reading (Chen et al., 2026).

These studies used correlational and laboratory designs with different populations, so they suggest plausible mechanisms rather than explanations for this class. Taken together, these studies and the present observations suggest that limited vocabulary is not merely background: it appeared to bear on comprehension, finding main ideas, paraphrasing, writing, and willingness to take part, so linguistic proficiency interacts with the strategy. A routine developed in L1 elementary classrooms may therefore need more scaffolding when transferred to lower-proficiency EFL learners, such as a mini-lesson on main ideas and worked examples of the organizers.

Time and Pacing

Time was another condition that shaped what students could do. The stages that came last (Organize, Plan, and Write) were repeatedly the ones lost when earlier stages overran (sessions 1 and 2) or when the lesson was shortened (session 4), and students described the lesson as too long or wanted more time. This suggests that pacing and implementation fidelity may influence what the strategy can visibly do in a classroom: a strategy whose later stages are not reached cannot show effects attributable to them. Whether a different schedule or shorter texts would change this was not tested.

Dimensions Intertwined

Conversation-analytic work on shared reading among upper-secondary vocational students found that the engagement dimensions intermingle, and that a Swedish as a second language group showed less stance-taking, possibly because of greater comprehension demands (Gustafsson et al., 2026). The present observations show a similar intertwining. Presentations drew on behavioral engagement (volunteering), cognitive engagement (stressing similarities and differences), affective engagement

(enthusiasm, nervousness), and social engagement (audience questions and correction) at once. Confusion in sessions 1 and 2 appeared to limit participation across several dimensions at the same time. The contexts differ, so these parallels are suggestive only.

Social Engagement Took Different Forms

Students' discussion of word meanings and supporting details, peer correction, and audience support were joined by copying and being given answers. The teacher's statement that some students were reluctant to ask peers for help did not match the peer help the observer recorded in four of six sessions, although two sessions were mainly individual work. In the validation of the EFL-REQ, social engagement showed a small negative effect on examination scores (Li et al., 2026). The present study did not examine achievement, but the mixture of help and copying is a reminder that a high level of peer interaction does not tell us what kind of interaction it is.

Instructional Influence on Engagement

A study of adult EFL learners at a diplomatic institute found higher engagement after AI-supported instruction than in a comparison condition(AlTwijri & Abdelhalim, 2026). The learners, intervention, and design differ from this study, and the contrast is offered only to show that engagement is open to instructional influence, which this study neither confirms nor refutes for Read-STOP-Write Strategy.

Study Limitations

Several limitations constrain what the findings can support, and they were taken into account in the interpretation above.

Design and context. The study involved one school, one teacher, and two intact classes over six sessions. Qualitative data were collected only in the experimental class, so the study cannot say how engagement looked in the control class or whether the engagement seen in the experimental class was distinctive. The increase in questionnaire totals in both classes shows that factors other than Read-STOP-Write Strategy, such as repeated testing or the passage of time, may have been at work. The classes were assigned on the basis of a previous English test, and other pre-existing differences were not examined. The ANCOVA also showed unequal residual variances, although the conclusion did not change with robust standard errors.

The questionnaire. The main analysis used total scores; dimension scores were analyzed only as secondary analyses, and the subscales are short (four or five items) with modest reliability at pretest (α = .67–.76), so dimension-level results should be read with caution. The questionnaire captures students' reported engagement with English reading in general, rather than their engagement specifically during Read-STOP-Write Strategy lessons, and it was administered at two time points, before and after the intervention. The Indonesian version was validated through expert judgment and showed high internal consistency for the total score, but it was not piloted or factor-analyzed with lower-secondary learners, and an item written for university exam contexts had to be reworded, so evidence for its validity with this age group is limited.

Observation. Observations were written by a single observer, who was also the researcher and the interviewer. The reports are narrative and unscored, the proportions they cite are estimates, and the protocol was designed to document implementation as well as student responses. The presence of an observer may have influenced behavior. The teacher added activities that are not part of Read-STOP-Write Strategy, which makes

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

it impossible to separate their effects from the strategy's. The researcher also prepared the lesson plans and materials, discussed the implementation with the teacher beforehand, and gave feedback after each lesson, so the researcher was not a neutral outsider: the feedback may have shaped later lessons, and the observer had an interest in the strategy working.

Interviews. Six students and one teacher were interviewed. Many student responses were brief and elicited through yes/no questions or questions that included wording suggested by the interviewer, which limits the strength of claims about subjective experiences such as enjoyment and confidence. In some interviews, the interviewer referred to students' score changes, which may have influenced their explanations. One student also gave inconsistent responses about which step of the strategy was the most difficult. The possibility of social desirability bias should also be considered, as students may have provided responses they believed the researcher expected.

In addition, the teacher who delivered the intervention was also interviewed about it, so her account partly evaluated her own practice. Interview codes could not be linked to individual students' questionnaire records; therefore, the qualitative findings cannot be directly associated with specific individual patterns of change. The interviews were translated from Indonesian into English by the researcher. Coding was conducted by a single researcher, without independent coding by a second coder. These limitations were considered when interpreting the interview findings.

5. Conclusion

In one ninth-grade EFL classroom, Read-STOP-Write was accompanied by engagement that was real but conditional. Students volunteered, reread, used dictionaries, helped one another, and presented their work, especially when they could complete tasks they could manage. They also became confused at demanding steps, tired in late or early lessons, and sometimes worked to finish, to copy, or to guess. The study does not show that Read-STOP-Write raised engagement compared with conventional instruction. It shows that engagement during the strategy fluctuated with task complexity, language demands, scaffolding, time, fatigue, and peer interaction, and that these conditions mattered as much as the strategy itself.

Total questionnaire scores rose in both classes, but after controlling for pretest scores the classes did not differ at posttest (p = .528) or in gains (p = .492), nor in any of the four engagement dimensions, and individual change was more varied in the experimental class.

Suggestions

The following implications are tentative, because they rest on one classroom, and each is tied to the finding that supports it.

For teachers. (a) Plan Read-STOP-Write Strategy across several lessons and protect time for each stage, since the stages that came last (Organize, Plan, Write) were lost when earlier stages overran (sessions 1, 2, and 4). (b) Model the strategy explicitly: teach finding the main idea in a short mini-lesson before Summarize and show a fully worked example of each organizer before students use it, since main-idea and TIDE tasks needed heavy scaffolding in sessions 1 and 2. (c) Use shorter texts at first so that a full cycle can be finished in one lesson, given the time lost in sessions 1, 2, and 4. (d) Structure peer work, for example by assigning roles for discussing word meanings and supporting details, since

students reported both helpful peer support and copying. (e) Monitor participation and fatigue, because in some lessons a few students gave most oral answers and sleepiness was recorded in early-morning and last-period lessons. (f) Presentations with audience questions drew strong participation and may be worth keeping, though their effect on engagement should be considered separate from the strategy.

For research. Future studies should use an engagement measure validated for lower-secondary EFL learners with subscales long enough to report dimension-level scores reliably; observe comparison classes as well as intervention classes with a scored protocol and more than one observer; keep interviewees linked to their questionnaire records; and use neutral, open interview questions. Longer implementations would show whether early confusion gives way to steadier engagement. Future work could also measure critical thinking, which the teacher mentioned cautiously and which has been shown to predict reading comprehension among seventh-grade English language learners in India (Paige et al., 2024).

6. References

Abdelhalim, S. M. (2017). Developing EFL students' reading comprehension and reading engagement: Effects of a proposed instructional strategy. *Theory and Practice in Language Studies*, *7*(1), 37–48. https://doi.org/10.17507/tpls.0701.05

AlTwijri, L., & Abdelhalim, S. M. (2026). Enhancing reading engagement in adult EFL classrooms through AI-supported instruction: A mixed-methods study. *System*, *136*. https://doi.org/10.1016/j.system.2025.103906

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE Publications.

Chen, X., Ye, Z., & Chi, R. (2026). L1-assisted L2 reading comprehension: Evidence from the differential impacts of L1 pre- and post-reading questions on L2 reading comprehension. *Reading and Writing*, *39*(5), 1–32. https://doi.org/10.1007/s11145-026-10798-1

Cho, E., Toste, J. R., Lee, M., & Ju, U. (2019). Motivational predictors of struggling readers' reading comprehension: The effects of mindset, achievement goals, and engagement. *Reading and Writing*, *32*(5), 1219–1242. https://doi.org/10.1007/s11145-018-9908-8

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research*. SAGE Publications.

Escobar, J.-P., Espinoza, V., & Valle, M. (2025). Exploring the relation between reading comprehension and executive functions: Does the level of reading comprehension test matter? *Learning and Instruction*, *99*. https://doi.org/10.1016/j.learninstruc.2025.102193

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, *74*(1), 59–109.

Gustafsson, A. W., Malmström, M., & Ohlsson, A. (2026). Identifying indicators of reading engagement among upper-secondary students during shared reading sessions. *L1-Educational Studies in Language and Literature*, *26*, 1–30. https://doi.org/10.21248/l1esll.2026.26.1.1187

Kassem, H. M., & Alqahtani, D. A. (2023). Motivation, strategy use, and comprehension in foreign language reading: The case of Saudi EFL learners at the preparatory year. *Journal of Language Teaching and Research*, *14*(5), 1290–1301. https://doi.org/10.17507/jltr.1405.17

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Lee, Y., Jang, B. G., & Conradi Smith, K. (2021). A Systematic Review of Reading Engagement Research: What Do We Mean, What Do We Know, and Where Do We Need to Go? *Reading Psychology*, *42*(5), 540–576. https://doi.org/10.1080/02702711.2021.1888359

Li, J., Ma, W., & Gan, Z. (2026). Development and validation of the English as a Foreign Language Reading Engagement Questionnaire. *Language Teaching Research*. https://doi.org/10.1177/13621688261443730

Özata, H., & Rızaoğlu, F. (2026). Reading with a purpose: How eye movements and working memory inform comprehension across L1 and L2 English readers. *System*, *141*. https://doi.org/10.1016/j.system.2026.104085

Protacio, M. S. (2017). A case study exploring the reading engagement of middle grades English learners. *RMLE Online*, *40*(3), 1–17. https://doi.org/10.1080/19404476.2017.1280586

Ronimus, M., Tolvanen, A., & Hautala, J. (2022). The roles of motivation and engagement in computer-based assessment of children's reading comprehension. *Learning and Individual Differences*, *98*. https://doi.org/10.1016/j.lindif.2022.102197

Shehzad, M. W., Zahoor, M., Sarfraz, S., & Razzaq, S. (2023). The relationships between reading mindsets, reading engagement, and reading comprehension performance in the Saudi EFL context. *Indonesian Journal of Applied Linguistics*, *13*(1), 99–113. https://doi.org/10.17509/ijal.v13i1.58276

Strong, J. Z. (2020). Investigating a text structure intervention for reading and writing in grades 4 and 5. *Reading Research Quarterly*, *55*(4), 545–551. https://doi.org/10.1002/rrq.356

Strong, J. Z. (2023). Investigating the effects and social validity of an informational text structure intervention for reading and writing in grades four and five. *Reading Psychology*, *44*(7), 820–852. https://doi.org/10.1080/02702711.2023.2202172

Strong, J. Z., Tortorelli, L. S., Anderson, B. E., & Cho, E. (2026). Development of a multicomponent reading and writing intervention in grades 4-5. *Reading Psychology*, *47*(1), 26–61. https://doi.org/10.1080/02702711.2025.2526464

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Taye, T., & Lodebo, T. (2026). The relationship between vocabulary learning strategies, reading comprehension, and motivation among secondary school EFL learners. *Social Sciences & Humanities Open*, *13*. https://doi.org/10.1016/j.ssaho.2026.102516

Wang, Y., & Hu, J. (2025). Relationships between reading engagement features and students' digital reading performance: An international large-scale study. *Learning and Instruction*, *99*. https://doi.org/10.1016/j.learninstruc.2025.102157

Yulia, M. F., Sulistyo, G. H., & Cahyono, B. Y. (2020). Affective engagement in academic reading: What EFL student teachers reveal. *International Journal of Evaluation and Research in Education*, *9*(3), 791–798. https://doi.org/10.11591/ijere.v9i3.20635

Zhai, K. (2025). Exploring the relationships between academic buoyancy, engagement, and achievement in English reading among EFL learners. *SAGE Open*, 1–13. https://doi.org/10.1177/21582440251357156

Zhang, X., Hui, S. Y., Lam, W. I., Ng, H. W. R., Chow, K. W., & Cheong, C. M. (2026). Mapping the association between reading engagement features in reading comprehension: A network analysis method. *Reading and Writing*, *39*(5), 1–28. https://doi.org/10.1007/s11145-026-10831-3