

Effects of Thematic Clustering Vocabulary Presentation on EFL Writing Complexity: A Quasi-Experimental Study in a Chinese Senior High School

Yutong Du^{1*}

¹ School of Foreign Languages and Literature, Beijing Normal University, Beijing, China

*Correspondence: Yutong Du

Received: 26 May 2026; **Accepted:** 23 June 2026; **Published:** 18 August 2026

Abstract:

Vocabulary is often presented through isolated explanation and memorization in EFL classrooms, which may limit students' ability to use it flexibly in writing. This study examines the effects of thematic clustering vocabulary presentation on Chinese Grade 10 students' EFL writing complexity, accuracy, and fluency. A quasi-experimental study was conducted in two intact classes in a senior high school in Suzhou, China. The experimental class received a unit-based intervention in which friendship-related vocabulary was organized through a framework of core quality, behavioral expression, and emotional value. The comparison class received conventional vocabulary presentation. Pre-test CAF measures showed no significant baseline differences between the two groups. After the intervention, students completed the same diary writing task, and their compositions were analyzed with clauses per T-unit, corrected type-token ratio, errors per 100 words, and words per T-unit. Post-test results showed that the experimental class significantly outperformed the comparison class in syntactic and lexical complexity, with medium effect sizes. No significant differences were found in accuracy or fluency. The findings suggest that thematic clustering vocabulary presentation may support more complex written production, especially in clausal subordination and lexical diversity, although longer-term instruction may be needed to improve accuracy and fluency.

Keywords:

Thematic Clustering, Vocabulary Presentation, EFL Writing, Writing Complexity, Quasi-experiment, Chinese Senior High School

1. Introduction

Vocabulary is central to second language writing because written expression depends not only on grammatical control, but also on the ability to retrieve, select, and organize lexical resources for communicative purposes. For English as a foreign language learners, a lack of usable vocabulary often leads to repetitive word choice, limited elaboration, and underdeveloped ideas in writing. This is especially evident in senior high school classrooms, where learners are expected to express nuanced meanings in relation to textbook themes and writing tasks.

In actual EFL classrooms, however, vocabulary teaching often stays at the level of word meaning, pronunciation, example sentences, and memorization. This kind of teaching is useful, but it may not be enough for writing. For example, a student may know the word *honest*, but still write only “My friend is honest” or “He tells the truth.” The difficulty is not simply that students know too few words. More often, they do not know how to connect a word with actions, situations, and personal meanings in a written text.

Thematic clustering vocabulary presentation offers a possible response to this problem. Unlike semantic clustering, which groups words that are semantically and syntactically similar, thematic clustering organizes words around a shared topic, event, or cognitive scene. Tinkham [1] found that semantic clustering may hinder the learning of new L2 vocabulary, while thematic clustering may facilitate vocabulary learning by allowing words to be connected through meaningful contextual relations. This distinction is important for classroom writing instruction. A theme-based vocabulary framework may help learners move from isolated adjectives to more complete expressions of behavior and meaning.

The Chinese senior high school English curriculum also provides a relevant pedagogical context for this approach. The General Senior High School Curriculum Standards state that English curriculum content includes thematic contexts, text types, language knowledge, cultural knowledge, language skills, and learning strategies. Thematic contexts provide the topics and situations around which teaching and learning content is organized [2]. The Standards also emphasize thematic meaning and contextualized language use, which makes thematic clustering vocabulary presentation relevant to writing instruction in the Chinese senior high school context.

EFL writing research has increasingly examined writing instruction, writing quality, and student texts. Yet school-level EFL writing remains less fully explored than college-level writing. Geng et al. [3] note that although EFL writing has received abundant attention, writing in primary and secondary school settings has been examined less often. This suggests the need for more classroom-based studies of younger EFL writers. The present study therefore situates itself within this underexplored school-level context.

The present study investigates thematic clustering vocabulary presentation in a unit-based writing task from the Integrated Skills section of Unit 3, *Getting along with others*, in the Yilin edition senior high school English textbook. The teaching topic was describing the qualities of a good friend. The experimental class was guided to organize vocabulary through a three-part framework of core quality, behavioral expression, and emotional value. The comparison class received conventional vocabulary presentation. The study then compared the two classes’ writing performance through complexity, accuracy, and fluency measures.

The focus of this article is writing complexity, because thematic clustering may help learners enrich topic-related vocabulary and extend sentence-level expression. Accuracy and fluency are also analyzed, but they are interpreted cautiously because a short-term vocabulary intervention may not directly lead to grammatical accuracy or production fluency.

The study was guided by two research questions. RQ1 asks what effects thematic clustering vocabulary presentation has on Grade 10 EFL learners’ syntactic and lexical complexity in writing. RQ2 asks what effects thematic clustering vocabulary

presentation has on Grade 10 EFL learners' writing accuracy and fluency, compared with conventional vocabulary presentation.

2. Literature Review

2.1. Thematic clustering in vocabulary presentation

Research on vocabulary presentation has long questioned the assumption that words should be taught in semantically similar sets. Semantic clustering refers to the practice of presenting words that share a common superordinate category, such as body parts, colors, or occupations. This practice is common in textbooks and classroom materials because it appears orderly and convenient. Yet its cognitive value is not self-evident. Tinkham [1] argues that grouping semantically and syntactically similar words may increase interference, since learners have to remember both the new forms and the subtle distinctions among similar items. His study provided evidence that semantic clustering hindered L2 vocabulary learning, while thematic clustering showed a more facilitative effect.

Waring [4] reached a similar conclusion through a replication study. He found that learning new words in semantic sets could interfere with vocabulary learning. Nation [5] also warned against teaching closely related lexical sets without sufficient contextual support, arguing that related words learned together may compete with one another in memory. These studies suggest that vocabulary organization matters because lexical items are learned through particular patterns of connection and processing.

Thematic clustering differs from semantic clustering because it does not rely on similarity within a single lexical category. It organizes words around a shared theme, situation, or event. In a friendship-themed unit, for instance, honest, sincere advice, tell the truth, respect principles, and recognize shortcomings may belong to the same thematic field, although they differ in word class, syntactic function, and semantic scope. Such organization may provide learners with a coherent route for processing vocabulary because items are linked by situational logic rather than similarity alone. In writing instruction, this distinction has practical value. Thematic clustering can guide students to construct a topic-related expression network instead of using target words in isolation.

2.2. Vocabulary resources and EFL writing performance

Vocabulary knowledge contributes to writing performance because writing requires productive control of words in context. Laufer and Nation [6] proposed the Lexical Frequency Profile as a measure of productive vocabulary in L2 written production, showing that the distribution of word frequency levels in learner texts can reflect productive vocabulary knowledge. Later research has continued to show that lexical features are related to writing quality. Kyle and Crossley [7], for example, found that lexical sophistication indices were associated with holistic writing scores in independent writing tasks, although their predictive value varied across task types. Their findings suggest that word choice, lexical range, and phrase-level features form part of the broader construct of writing proficiency.

Lexical diversity is another relevant construct. Jarvis [8] notes that lexical diversity generally refers to the variety of words used in a language sample and is often operationalized through relationships between types and tokens. He also distinguishes

lexical diversity from lexical sophistication and lexical richness, which is useful for the present study because the lexical measure used here is corrected type-token ratio. In this article, lexical complexity is therefore discussed mainly in terms of lexical diversity rather than overall vocabulary quality.

For senior high school EFL writers, lexical diversity is closely tied to the ability to express ideas without excessive repetition. In a writing task about friendship, a learner who repeatedly writes good, kind, helpful, and like may complete the task but may not demonstrate mature lexical control. By contrast, a learner who connects reliable with face difficulties, offer help, feel comforted, and sustain friendship demonstrates a wider vocabulary range. The present study therefore examines whether one way of organizing vocabulary may influence measurable dimensions of written performance in a short-term classroom intervention.

2.3. CAF measures in L2 writing research

Complexity, accuracy, and fluency have been widely used to describe second language performance. Skehan [9] treated the three dimensions as important indicators for analyzing language performance under different learning and task conditions. In L2 writing research, CAF measures allow researchers to examine not only whether a text is better in a general sense, but also how it differs in linguistic structure, error control, and production quantity.

Complexity is often divided into syntactic and lexical complexity. In this study, syntactic complexity is measured by clauses per T-unit. This index reflects the amount of clausal subordination within a T-unit and has been widely used in L2 writing research [10]. It has also been applied in studies of Chinese adolescent EFL writers, where syntactic complexity indices including clauses per T-unit have been found to relate to writing quality [11]. The L2 Syntactic Complexity Analyzer developed by Lu [12] provides automated measures of syntactic complexity, including clauses per T-unit. Lexical complexity is measured by corrected type-token ratio, which was selected because it reduces the influence of text length more effectively than raw type-token ratio.

Accuracy concerns the extent to which learner language conforms to target language norms. Chandler [13] used errors per 100 words as a measure of accuracy in L2 writing research, and this measure is adopted in the present study. Fluency in this study is operationalized as words per T-unit. This measure reflects the amount of written production within a T-unit, although it cannot fully represent writing speed or cognitive automaticity.

Pallotti [14] cautions that CAF measures should not be equated directly with language development or communicative adequacy. CAF dimensions describe properties of language performance, but they do not exhaust the quality of writing. This warning is important for the present study. Higher C/T and C-TTR values may indicate greater syntactic and lexical complexity, but they do not automatically mean that a learner's overall writing ability has improved. Similarly, a non-significant result in accuracy or fluency should not be treated as a failed outcome.

Based on this view, the present study uses CAF measures as a focused analytic lens. The purpose is not to claim broad writing development, but to examine whether thematic clustering vocabulary presentation is associated with differences in specific aspects of students' written performance. This approach connects classroom

vocabulary instruction with observable textual outcomes while keeping interpretive claims within the scope of the design.

3. Materials and Methods

3.1. Research design

This study adopted a classroom-based quasi-experimental design. Since the research was conducted in a real senior high school classroom, random assignment of individual students was not feasible. Two intact Grade 10 classes were therefore selected, with one class assigned as the experimental class and the other as the comparison class. This design was appropriate for examining different vocabulary presentation approaches within a regular instructional context [15].

The broader research period lasted from September to December 2024, during which the instructional intervention was embedded in the teaching of Unit 3, *Getting along with others*, in the Yilin edition of the senior high school English textbook. The intervention was designed as a short-term unit-based intervention lasting two weeks and spanning eight 45-minute lessons. Both classes were taught by the same English teacher during the unit, which helped reduce teacher-related variation in instructional style, classroom management, and feedback practices.

3.2. Participants and pre-test

The study involved two parallel Grade 10 classes in a senior high school in Suzhou, China. Each class had 40 students, so the total number of participants was 80. Both classes used the same Yilin senior high school English textbook and followed the same school teaching schedule. One class was assigned as the experimental class, and the other was used as the comparison class. Since the study was carried out in an ordinary school setting, the original class structure was kept unchanged.

Before the unit-based intervention, the same 80 students completed a routine classroom writing task. The task was an email of advice on a familiar interpersonal topic, and it was marked with the school's 15-point writing rubric. The two classes were close in their overall writing scores, and the difference was not significant, $t(78) = -0.22$, $p = .830$, $d = -0.05$, 95% CI $[-0.70, 0.56]$. To check the baseline more closely, four CAF indices were also calculated for the pre-test texts: clauses per T-unit, corrected type-token ratio, errors per 100 words, and words per T-unit. The pre-test and post-test were both completed in class under timed conditions. Students did not use dictionaries, textbooks, smart devices, or peer discussion during either task.

3.3. Instructional intervention

The instructional intervention was embedded in the complete teaching cycle of Unit 3, *Getting along with others*, from the Yilin Senior High School English Textbook. Thematic clustering vocabulary presentation was integrated across the unit in forms appropriate to each lesson type, such as pre-reading prediction and post-reading consolidation. The focal lesson analyzed in this study was Lesson 7, the writing production lesson, which served as the immediate instructional context for the post-test writing task.

Both the experimental and comparison classes followed the same 45-minute lesson structure, with 15 minutes allocated to vocabulary work as a pre-writing activity. This activity was positioned after students had read Patrick's diary entry as reading input

and before they began writing Amy's diary entry as written output. The main instructional difference lay in how vocabulary was organized and practiced during this 15-minute pre-writing phase.

In the experimental class, vocabulary was presented through thematic clustering using three instructional tools: a vocabulary classification framework, mind map or tree diagram templates, and a sentence upgrading worksheet. The teacher presented three core adjectives describing friendship qualities as central nodes of the vocabulary network: supportive, honest and open, and patient. Each core quality was extended into behavioral expressions and emotional or evaluative values. For example, supportive was linked with stand by me in trouble, share my feelings, and make me feel understood; honest and open was linked with give sincere advice, tell the truth, and build trust; patient was linked with listen to my complaints, wait without frustration, and make me feel comforted. Students then used the templates and sentence upgrading worksheet to expand the core qualities into fuller written expressions before beginning the diary writing task.

In the comparison class, the same 15 minutes were used for conventional vocabulary review and sentence-level practice. The teacher explained word meanings and usage, provided example sentences from the textbook, and asked students to practice using the words in isolated sentences. Students were not guided to organize vocabulary thematically or to connect core qualities with behavioral expressions and emotional values.

Table 1. *Overview of the instructional intervention.*

Aspect	Experimental class	Comparison class
Unit duration	Two weeks, eight 45-minute lessons	Two weeks, eight 45-minute lessons
Focal lesson	Lesson 7, writing production lesson	Lesson 7, writing production lesson
Vocabulary slot	15-minute pre-writing activity	15-minute pre-writing activity
Main vocabulary focus	Supportive; honest and open; patient	Supportive; honest and open; patient
Vocabulary processing	Thematic clustering through core quality, behavioral expression, and emotional value	Conventional review of word meanings, example sentences, and sentence-level practice
Materials	Vocabulary classification framework; mind map or tree diagram; sentence upgrading worksheet	PPT explanation; textbook examples; conventional sentence practice
Writing task	Diary entry from Amy's perspective	Same diary entry from Amy's perspective

3.4. Writing task

After the unit-based instruction, students in both classes completed the same post-test writing task. They were asked to write a diary entry from Amy's perspective based on the unit context. The writing prompt asked students to describe the qualities that had helped repair and sustain the friendship. The task was aligned with the Integrated Skills section of the textbook and required students to express friendship qualities through both description and reflection.

The post-test was administered as a routine classroom writing task under the same conditions for both classes. Students completed the task within 15 minutes, with a target length of 80-100 words. The teacher provided no immediate linguistic hints or

corrective feedback during the test. All scripts were anonymized and assigned numerical codes before computer-aided analysis and manual error checking.

The writing task was treated as part of the regular classroom teaching sequence rather than as an external examination. This helped maintain the authenticity of the task while keeping the testing conditions consistent across the two classes. Both groups completed the task individually within the same time limit and without dictionaries, textbooks, smart devices, or peer consultation. The teacher did not provide sentence-level hints, corrective feedback, or additional vocabulary support during the writing period. All scripts were collected immediately after the task and anonymized before analysis. In this way, the writing samples reflected students' immediate written production after the pre-writing vocabulary activity, while reducing the possible influence of external assistance or post-task revision.

3.5. Measures and data analysis

Students' pre-test and post-test compositions were analyzed in terms of complexity, accuracy, and fluency. For syntactic complexity, the texts were processed with the L2 Syntactic Complexity Analyzer. This study used clauses per T-unit as the main syntactic index, because it captures the extent to which students combine clauses within a T-unit. This index was suitable for the present writing task, where students needed to explain qualities, actions, and emotional effects in connected sentences. Lu [12] describes L2SCA as a tool that can automatically calculate a range of syntactic complexity indices in second language writing, including clauses per T-unit.

Lexical complexity was measured by corrected type-token ratio. This index was used to examine lexical diversity in students' compositions. Compared with raw type-token ratio, corrected type-token ratio is less affected by text length and is therefore more suitable for comparing short student texts. Accuracy was measured by the number of grammatical and lexical errors per 100 words. Fluency was measured by words per T-unit, which served as a proxy measure of written fluency.

Table 2. Measures used for analyzing writing performance.

Construct	Operationalization	Index
Syntactic complexity	Clausal subordination	Clauses per T-unit
Lexical complexity	Lexical diversity	Corrected type-token ratio
Accuracy	Error rate	Errors per 100 words
Fluency	Length within T-units	Words per T-unit

Pigai was used for the first round of error checking. After that, the researcher read each composition again and checked the flagged errors in context. The final error count included four types of problems: grammar errors, unsuitable word choices, spelling errors, and punctuation errors that affected syntactic segmentation, such as missing full stops or comma splices. Repeated errors involving the same word form were counted once. Some items marked by Pigai were not counted as errors after manual checking, such as proper nouns that the system did not recognize or diary-style expressions that were acceptable in context.

To make the error counts more reliable, another trained rater checked 20% of the compositions. The sample included eight texts from the experimental class and eight from the comparison class. The two raters agreed on 87.5% of the checked cases. For the remaining cases, the two raters discussed the context and reached a final decision before the error rates were calculated.

SPSS 26.0 was used for statistical analysis. Independent-samples t-tests were conducted to compare the experimental and comparison classes on pre-test and post-test CAF measures. Levene's tests were used to examine the homogeneity of variance assumption. When equal variances were not assumed, Welch's t-test results were reported. Cohen's d was calculated to indicate the magnitude of group differences, and 95% confidence intervals were reported for the mean differences. The significance level was set at $p < .05$.

4. Results and Discussion

4.1. Results

4.1.1. Baseline comparability

Before discussing the post-test results, the pre-test CAF data were checked to see whether the two classes started from a similar point. Table 3 shows the results. The two classes did not differ significantly in C/T, C-TTR, errors per 100 words, or W/T. This means that the experimental and comparison classes were close not only in their overall writing scores, but also in the four text-level measures used in this study.

Table 3. Independent-samples t-test of pre-test CAF measures.

Measure	Group	M	SD	t(df)	p	d	95% CI
C/T	Comparison	1.41	0.18	-0.25 (78)	.804	-0.06	[-0.09, 0.07]
	Experimental	1.42	0.17				
C-TTR	Comparison	5.75	0.50	0.18 (78)	.861	0.04	[-0.21, 0.25]
	Experimental	5.73	0.52				
Accuracy	Comparison	5.52	2.10	-0.09 (78)	.930	-0.02	[-0.94, 0.86]
	Experimental	5.56	1.95				
Fluency	Comparison	11.25	2.20	-0.10 (78)	.923	-0.02	[-1.09, 0.99]
	Experimental	11.30	2.45				

Note. Confidence intervals and Cohen's d are based on the mean difference between the comparison and experimental classes.

4.1.2. Post-test CAF performance

The post-test results are presented in Table 4. The experimental class obtained higher mean scores than the comparison class in both clauses per T-unit and corrected type-token ratio. The differences were statistically significant for syntactic complexity, $t(78) = -2.30$, $p = .024$, and lexical complexity, $t(78) = -2.69$, $p = .009$. Effect sizes were medium for both syntactic complexity ($|d| = 0.51$) and lexical complexity ($|d| = 0.60$), based on Cohen's [16] conventions.

Table 4. Independent-samples t-test of post-test CAF measures.

Measure	Group	M	SD	t(df)	p	d	95% CI
C/T	Comparison	1.48	0.18	-2.30 (78)	.024	-0.51	[-0.17, -0.01]
	Experimental	1.57	0.17				
C-TTR	Comparison	5.82	0.50	-2.69 (78)	.009	-0.60	[-0.52, -0.08]
	Experimental	6.12	0.50				
Accuracy	Comparison	5.28	2.45	1.64 (55.38)	.107	0.37	[-0.16, 1.56]
	Experimental	4.58	1.15				
Fluency	Comparison	11.62	2.34	-1.92 (78)	.058	-0.43	[-2.36, 0.04]
	Experimental	12.78	3.02				

Note. The Welch result is reported for accuracy because the equal variance assumption was not met. Confidence intervals and Cohen's d are based on the mean difference between the comparison and experimental classes.

For accuracy, the experimental class had a lower mean error rate than the comparison class, but the difference was not statistically significant. Because the homogeneity of variance assumption was not met for this measure, Welch's t -test was reported, $t(55.38) = 1.64$, $p = .107$. For fluency, the experimental class produced a higher mean value for words per T-unit, but the difference did not reach statistical significance, $t(78) = -1.92$, $p = .058$.

Overall, the post-test results show a clear but limited pattern. The experimental class significantly outperformed the comparison class in syntactic complexity and lexical complexity. No statistically significant differences were found in accuracy or fluency. The findings therefore answer RQ1 positively for complexity-related measures, while RQ2 shows no significant group difference in accuracy or fluency following the intervention.

4.2. Discussion

4.2.1. Thematic clustering vocabulary presentation and writing complexity

Thematic clustering vocabulary presentation was associated with significant post-test differences in both syntactic and lexical complexity. Since the two groups did not differ significantly on pre-test CAF measures, these post-test differences can be interpreted as classroom-based evidence that the intervention may have supported complexity-related aspects of written production. These findings should therefore be interpreted as evidence of complexity-related differences following the intervention, rather than as evidence of broad writing development.

The classroom data also suggest that the effect of the intervention was closely tied to the nature of the writing task. The post-test required students to describe friendship qualities from Amy's perspective, so students needed to explain not only what kind of friend was valuable, but also how such a friend behaved and why the behavior mattered. This task demand matched the three-part vocabulary framework used in the experimental class. The framework gave students a route from a quality word to an action and then to an emotional or evaluative meaning. This may help explain why the clearest differences appeared in C/T and C-TTR rather than in accuracy or fluency.

In the writing lesson, the experimental class did not review the adjectives only as separate words. The teacher asked students to start from one quality, such as honest, and then think about what this quality looks like in a real friendship. For example, an honest friend may give sincere advice, tell the truth, or point out problems kindly. These links gave students more material to build sentences. Instead of writing only "My friend is honest," they had a clearer path to explain what the friend did and why it mattered. This may be one reason why the experimental class produced more clauses per T-unit in the post-test task.

This result is consistent with Tinkham's [1] argument that thematic clustering facilitates vocabulary learning by allowing words to be connected through meaningful contextual relations. In the present study, the friendship theme provided a shared situational frame within which vocabulary could be organized and retrieved. The instructional focus was not on isolated word recall, but on helping students develop topic-related expression networks for writing.

The increase in lexical complexity can be interpreted in a similar way. The higher corrected type-token ratio in the experimental class suggests that students used a more varied range of words in the post-test task. This aligns with research showing that vocabulary use in writing reflects productive vocabulary knowledge [6] and that lexical features are related to writing quality in independent writing tasks [7]. Some students in the experimental class drew on expressions such as give sincere advice, recognize my own shortcomings, and stand by me in trouble. These expressions suggest that students were not merely repeating adjectives, but extending core qualities into more specific behavioral and evaluative meanings.

The representative excerpts provide a qualitative illustration of the statistical results.

Table 5. Representative excerpts from student writing.

Group	Representative excerpt	Main feature
Experimental class	An honest and open friend always gives me sincere advice instead of just saying nice words, which effectively helps me recognize my own shortcomings.	The core quality is expanded into behavior and personal value. The sentence also includes a relative clause.
Experimental class	My best friend is extremely patient. Whenever I suffer from study stress, he listens to my complaints without any frustration, making me feel deeply comforted.	The sentence shows a situation, behavior, and emotional effect connected with the core quality.
Comparison class	Tom is a loyal friend. He is very loyal to me. When I am sad, he is helpful and helps me a lot.	The writing shows repeated use of similar words and limited topic-related alternatives.

Note. Excerpts were selected from post-test compositions to illustrate typical patterns in each group. Student identifiers were removed.

The contrast across these excerpts helps clarify the textual patterns reflected in the quantitative results. In the first experimental excerpt, the student connects honest and open with sincere advice and recognize my own shortcomings. The sentence also uses a non-restrictive relative clause to explain the value of the friend's behavior. In the second experimental excerpt, the student connects patient with a situation, a behavioral response, and an emotional effect. In the comparison excerpt, the student can use target adjectives such as loyal and helpful, but the writing relies heavily on repetition and short sentence-level description. These examples do not prove the quantitative results, but they illustrate the kinds of textual differences that the C/T and C-TTR measures captured at the group level.

4.2.2. Limited effects on accuracy and fluency

The intervention was not associated with significant post-test differences in accuracy. Although the experimental class had a lower mean error rate, the difference between the two classes was not statistically significant. This result is understandable because the intervention mainly focused on vocabulary organization and topic-related expression. It did not provide systematic grammar instruction or repeated corrective feedback. Accuracy in L2 writing depends on learners' grammatical control, attention to form, editing ability, and familiarity with target structures. A short-term vocabulary-based intervention may support lexical and syntactic elaboration, but it may not be sufficient to reduce grammatical errors in a measurable way.

This finding also supports a cautious interpretation of CAF measures. Pallotti [14] notes that complexity, accuracy, and fluency are separate performance dimensions. A gain in complexity does not necessarily lead to a gain in accuracy. This is also consistent with Skehan's [9] attention resource model, which suggests that attention to complexity may reduce attention to formal accuracy.

The effect on fluency was also not significant. The experimental class had a higher mean words per T-unit value, but the difference did not reach statistical significance. In this study, words per T-unit was used as a proxy measure of written fluency. It reflects the amount of written production within T-units, but it does not directly measure writing speed, hesitation, or cognitive automaticity. Written fluency is influenced by topic familiarity, planning ability, vocabulary retrieval, syntactic control, and time pressure, which may not change substantially within a short-term unit-based intervention.

The non-significant result for fluency suggests that thematic clustering vocabulary presentation may first influence the organization of lexical resources, while sustained practice may be needed before this ability is reflected in increased written production fluency.

4.2.3. Pedagogical implications

The findings suggest several implications for EFL writing instruction in senior high school classrooms. Vocabulary teaching for writing should not stop at word explanation or memorization. Teachers can guide students to organize vocabulary according to thematic logic, especially when the writing task is closely linked to a unit theme. A three-part framework of core quality, behavioral expression, and emotional value may help students transform abstract adjectives into more concrete and meaningful written expressions. This approach is also aligned with the thematic context-based requirements of the General Senior High School Curriculum Standards [2].

For this kind of thematic vocabulary work to support writing, it needs to lead students from word grouping to sentence making. Words such as supportive, honest, and patient become useful only when students know how to show these qualities in action. In the writing lesson, mind maps and tree diagrams helped students see the links between a quality, a behavior, and an emotional result. Sentence revision tasks then gave them a chance to put these links into longer sentences. This step matters because students are not only learning topic words. They are learning how to turn those words into written expression.

This also means that thematic clustering should not become another kind of word list. If students only copy grouped words, the activity may not help their writing very much. The key is to let students use the grouped words to say something meaningful. For this reason, vocabulary classification needs to be followed by short sentence practice, discussion, and a writing task. In this study, the clustered vocabulary worked as a bridge between the pre-writing input and the diary entry that students later produced.

However, thematic clustering vocabulary presentation should not be expected to improve all aspects of writing within a short period. If teachers aim to improve accuracy, vocabulary work should be combined with form-focused instruction and feedback on common grammatical errors. If they aim to improve fluency, students need repeated writing practice, timed writing opportunities, and familiarity with

different writing tasks. Thematic clustering is therefore best understood as one component of writing instruction, rather than a complete solution.

4.2.4. Limitations

Several limitations need to be acknowledged. The study involved only two intact Grade 10 classes from one senior high school, which limits the generalizability of the findings. Although the pre-test CAF measures showed no significant baseline differences, the use of intact classes means that unmeasured differences between the two groups cannot be completely ruled out.

The intervention was short-term and unit-based, making it unsuitable for capturing long-term writing development or strategy transfer. The pre-test and post-test tasks were related in their interpersonal focus, but they differed in genre: the pre-test was an email of advice, whereas the post-test was a diary entry. Genre differences may have influenced fluency and lexical measures in particular, since diary writing typically allows more personal and emotionally expressive vocabulary than email writing. Future studies could use parallel tasks of the same genre and include delayed post-tests to examine whether the effects are retained.

Finally, the analysis relied on a small set of quantitative indices and representative excerpts. These measures captured important textual features, but they did not fully examine discourse organization, idea development, or learner perceptions. Future research could address these limitations by expanding the sample, extending the intervention period, and combining CAF analysis with interviews, classroom observation, or discourse-based measures.

5. Conclusions

This study investigated whether thematic clustering vocabulary presentation was associated with Grade 10 students' EFL writing performance in a Chinese senior high school. The analysis focused on four CAF measures, including syntactic complexity, lexical complexity, accuracy, and fluency. The experimental class received thematic clustering vocabulary presentation organized through a three-part framework, while the comparison class received conventional vocabulary review.

The results showed that the experimental class significantly outperformed the comparison class in syntactic complexity and lexical complexity after the intervention. Students who received thematic clustering vocabulary presentation produced higher clauses per T-unit and higher corrected type-token ratio in the post-test writing task. These findings suggest that thematic clustering vocabulary presentation may help learners organize topic-related lexical resources and produce more complex written language, especially in clausal subordination and lexical diversity.

No statistically significant differences were found in accuracy or fluency. This indicates that a short-term vocabulary-based intervention may not be sufficient to improve grammatical accuracy or written fluency. Improving accuracy likely requires explicit form-focused instruction alongside vocabulary work, while developing fluency may call for sustained writing practice over a longer period.

The study provides preliminary classroom-based evidence that vocabulary presentation can be linked with unit themes and writing tasks in senior high school EFL instruction. It also suggests a classroom-applicable approach for helping students move from isolated word knowledge to elaborated written expression. Future research

may extend the intervention period, include parallel pre-test and post-test writing tasks, and examine whether thematic clustering can support long-term writing development across different genres, topics, and learner populations.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Acknowledgments

The author would like to acknowledge the participating school, English teacher, and students for their support during data collection.

References

- [1] Tinkham, T. The effects of semantic clustering and thematic clustering on the learning of second language vocabulary. *Second Language Research*, 1997, 13, 138-163. DOI: 10.1191/026765897672376469.
- [2] Ministry of Education of the People's Republic of China. General Senior High School Curriculum Standards: English 2017 Edition, 2020 Revision; People's Education Press: Beijing, China, 2020.
- [3] Geng, F.; Yu, S.; Liu, C.; Liu, Z. Teaching and learning writing in English as a foreign language school education context: A thematic review. *Scandinavian Journal of Educational Research*, 2022, 66, 491-504. DOI: 10.1080/00313831.2021.1897872.
- [4] Waring, R. The negative effects of learning words in semantic sets: A replication. *System*, 1997, 25, 261-274. DOI: 10.1016/S0346-251X(97)00013-4.
- [5] Nation, P. Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, 2000, 9, 6-10.
- [6] Laufer, B.; Nation, P. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 1995, 16, 307-322. DOI: 10.1093/applin/16.3.307.
- [7] Kyle, K.; Crossley, S.A. The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 2016, 34, 12-24. DOI: 10.1016/j.jslw.2016.10.003.
- [8] Jarvis, S. Capturing the diversity in lexical diversity. *Language Learning*, 2013, 63, 87-106. DOI: 10.1111/j.1467-9922.2012.00739.x.
- [9] Skehan, P. A Cognitive Approach to Language Learning; Oxford University Press: Oxford, United Kingdom, 1998.
- [10] Wolfe-Quintero, K.; Inagaki, S.; Kim, H.Y. Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity; University of Hawaii Press: Honolulu, HI, USA, 1998.

- [11]Bi, P.; Jiang, J. Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity. *System*, 2020, 91, 102248. DOI: 10.1016/j.system.2020.102248.
- [12]Lu, X. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 2010, 15, 474-496. DOI: 10.1075/ijcl.15.4.02lu.
- [13]Chandler, J. The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 2003, 12, 267-296. DOI: 10.1016/S1060-3743(03)00038-9.
- [14]Pallotti, G. CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 2009, 30, 590-601. DOI: 10.1093/applin/amp045.
- [15]Mackey, A.; Gass, S.M. *Second Language Research: Methodology and Design*, 2nd ed.; Routledge: New York, NY, USA, 2016.
- [16]Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)