Impact of Task Complexity on Lexical Complexity in CET4 Writing Insights from LCA Measurement Dimensions

Changlin Li[1,2], Nik Aloesnita Nik Mohd Alwi[1]*, Mohammad Musab Azmat Ali[1]

[1]Centre for Modern Languages, Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia
[2]Academic Affairs Office, Hebei Minzu Normal University, China

Abstract

To investigate how task complexity cognitive factors influence the lexical complexity in essay writing, this study was guided by Robinson's Cognition Hypothesis (CH) and Skehan's Limited Attentional Capacity Model (LACM), and examined the effects of task complexity on lexical complexity in undergraduates' essay writing. Using Lu's Lexical Complexity Analyser (2010, 2012), the study analysed essays written by 60 undergraduates across 25 dimensions of lexical complexity, following the SSARC grouping method by Robinson and Gilabert (2007). The Wilcoxon signed-rank test evaluated the relationship between pre-existing standardised writing scores (PS) and the essay outputs under the experimental condition (QM), focusing on lexical density, lexical sophistication, and lexical variation dimensions. Findings indicate that verb sophistication and the Uber index are significantly associated with higher proficiency, supporting Robinson's CH, which links cognitive resource allocation in complex tasks to improved language proficiency. Noun variation also emerged as significant, aligning with recent studies emphasising lexical diversity in writing. However, measurements like lexical density and type/token ratio showed no significant differences, suggesting a limited role. The study highlights the importance of sophisticated lexical use; however, it also acknowledges that cognitive resource limitations, as posited by Skehan's LACM, may constrain the deployment of more complex lexical structures.

Keywords: *Task Complexity, Influence, Lexical Complexity, CET4 Writing*

Introduction

Task-Based Language Teaching and Learning (TBLT) represents a dynamic and interrelated domain within applied linguistics and language education (Robinson, 2001a;

Robinson & Gilabert, 2007; Robinson, 2007, 2011c; Robinson et al., 2014; Skehan & Foster, 1997; Skehan, 1998, 2009, 2011; Skehan & Foster, 2008). A pivotal strand of inquiry within this field concerns task complexity and its impact on language production. This line of research has garnered sustained scholarly attention in applied linguistics and second language acquisition (Ellis, 2018a, 2018b; Robinson, 2001a; Robinson, 2007, 2011c; Robinson et al., 2014; Sauer & Ellis, 2019; Skehan, 1998, 2009, 2018). In recent years, research has examined the impact of varying levels of task complexity on language learners' writing proficiency, with a particular focus on lexical complexity, fluency, overall language use, etc. The previous research has contributed to this field of study (Abdi Tabari, et al., 2023; Allaw, 2021; Awwad & Tavakoli, 2022; DeCoursey, 2023; Golparvar & Rashidi, 2021; Gomez Laich & Taguchi, 2023; Güvendir & Uzun, 2023; Hsu, 2020; Khatib & Farahanynia, 2020; Kim, 2022a; Lee, 2020b; Soh, 2020, 2022; Soh et al., 2020; Xu et al., 2022). Although the influence of task complexity on learners' language proficiency, particularly in terms of lexical complexity, has been widely examined, existing studies have yielded varied conclusions due to differing research contexts and methodological approaches. Consequently, this study seeks to investigate the impact of task complexity on essay writing within the College English Test Band 4 (CET4) (NCETC, 2016) context, with a specific focus on lexical complexity as assessed through the dimensions operationalised by the Lexical Complexity Analyser (LCA) (Lu, 2010, 2012).

Within the frameworks of TBLT, task complexity encompasses a range of factors that shape learners' performance and mediate the cognitive and linguistic demands necessary for successful task completion. Research on task complexity, particularly under Skehan's Limited Attentional Capacity Model (LACM) (Skehan & Foster, 1997; Skehan, 1998, 2009, 2011, 2013; Skehan & Foster, 2008) and Robinson's Cognition Hypothesis (CH) (Robinson, 2001b, 2005; Robinson & Gilabert, 2007; Robinson, 2015, 2022b), has shown that attentional resources play a pivotal role in determining how learners manage task demands. However, there is still limited research exploring how task complexity impacts lexical complexity in college English test (CET) writing, particularly when it explores the influence of task complexity on lexical complexity in CET essay writing using an automated measurement approach such as the Lexical Complexity Analyser (LCA) (Lu, 2010, 2012) from 25 dimensions. Accordingly, this study investigates how cognitive factors of task complexity, specifically "+PT, –FE" and "–PT, +FE", influence lexical complexity in CET4 essay writing. Grounded in the principles of TBLT, the research explores undergraduate writing performance by drawing on Skehan's LACM, Robinson's CH, and lexical measurement via Lu's LCA. The study aims to inform the design of college English writing tasks and to enhance instructional practices in undergraduate writing pedagogy.

Literature Review
This section reviews the impact of task complexity on essay writing under Skehan's LACM (Skehan & Foster, 1997; Skehan, 1998, 2009, 2011, 2014, 2018; Skehan & Foster, 2001, 2008) and Robinson's CH (Robinson, 2001a, 2001b, 2003, 2005; Robinson & Gilabert, 2007; Robinson, 2007, 2011a, 2011b, 2011c; Robinson et al., 2014), and prior

research mainly in the past decade.

A Brief Review of TBLT and the Cognitive Factors of Task Difficulty
Since TBLT is an instructional approach emphasising using meaning-focused, real-world tasks to promote language learning (Ellis, 2018a; Ellis, 2005; Robinson, 2001a; Robinson & Gilabert, 2007; Robinson, 2007, 2011c; Skehan & Foster, 1997; Skehan, 1998, 2009, 2011; Skehan & Foster, 2008), it draws on theoretical insights from language teaching to inform its principles and practices. Thus, language teaching research has employed TBLT as an essential framework for elucidating how disparate task conditions and types can facilitate or impede language development. Previous research has examined the relationship between cognitive load and task complexity, with particular attention to how these factors impact learners' language performance.

The Limited Attentional Capacity Model (LACM) (Skehan & Foster, 1997; Skehan, 1998, 2009, 2011, 2013; Skehan & Foster, 2008), which proposed that learners have a finite pool of attentional resources. Language production in task-based contexts necessitates a balance between language complexity, accuracy, and fluency due to the constraints on attentional capacity and working memory. It further asserted that focusing on one aspect, such as complexity or accuracy, might detract from fluency. The theory claimed that an emphasis on certain elements could result in a decline in performance in other areas. Overall, these dynamics may provide insights into language instruction strategies that effectively balance complexity, accuracy, and fluency. Studies have investigated the impact of planning time on managing cognitive load during complex tasks. Bui and Huang (2020a) found that learners given pre-task planning time performed better in accuracy and complexity than those with no planning time. The researchers concluded that planning time allows learners to reduce cognitive load by organising their thoughts and preparing linguistic resources. Similarly, Révész et al. (2019) demonstrated that providing learners with additional planning time helped them manage the cognitive demands of complex tasks, resulting in improved performance across CAF dimensions. These findings support Skehan's LACM, suggesting that planning time enables learners to balance complexity, accuracy, and fluency.

The Cognition Hypothesis (CH) (Robinson, 2001b, 2005; Robinson & Gilabert, 2007; Robinson, 2015, 2022b) emphasises that more demanding tasks, such as those with more elements or higher reasoning demands, can push learners to produce more advanced language. It explored the impact of task demands on language output, utilising general metrics such as accuracy, fluency, and complexity. For instance, metrics typically involve the error percentage per speech unit (C-unit) or the number of subordinate clauses within a C-unit, among other factors. These comprehensive output measures supported the appropriateness of their application and indicated that increased task complexity generally leads to greater accuracy and lexical complexity across various task dimensions when compared to less complex tasks. Research on the number of elements in a task has shown that increasing the number of elements significantly increases cognitive load. Trebits and Kormos (2019) examined how the number of task elements affected learners' performance on written tasks. They found that tasks with more elements decreased fluency as learners struggled to manage the additional cognitive demands. However, they

also concluded that learners could produce more complex language structures, which supports Robinson's CH. Meanwhile, Kang and Lee (2019) also explored the impact of task elements on cognitive load, finding that tasks with a higher number of elements led to a greater focus on accuracy at the expense of fluency. Researchers have concluded that learners tend to prioritise grammatical accuracy to cope with the heightened cognitive demands imposed by increased task complexity. Further studies have investigated the influence of task complexity and planning time on writing performance. Findings consistently indicate that the provision of planning time positively affects learners' writing, as it facilitates more effective idea organisation and more efficient management of linguistic resources. Increasing task complexity, particularly by adding more elements, generally results in greater lexical complexity. However, this can lead to reduced accuracy due to cognitive overload (Abdi Tabari, 2016). The interaction between these variables indicates that planning time is significant for learners in managing the increased cognitive demands of complex tasks.

Research has increasingly recognised the intricate relationship between task complexity and the multiple dimensions of writing proficiency. The impact of task complexity on lexical complexity has been elucidated by Abdi Tabari et al. (2023), who emphasised higher levels in complex tasks. Meanwhile, the research (Abdi Tabari et al., 2023) explored the effects of verb argument construction on syntactic complexity and emphasised the significance of complex tasks in developing sophisticated essays. A deeper investigation into the domain of anxiety has demonstrated that elevated levels of writing anxiety may impede working memory, consequently reducing complexity (Güvendir & Uzun, 2023). Furthermore, Gomez Laich and Taguchi (2023) demonstrated the beneficial effect of complex tasks on rhetorical strategies, particularly compared to more straightforward essay tasks.

Empirical Studies on Task Complexity and Lexical Outcomes

Indeed, research consistently demonstrated that task complexity, often operationalised through planning time and the number of task elements, influences writing performance (Ellis, 2018a; Kim, 2022a; Skehan, 2009). Within the task complexity factors, Skehan's LACM hypothesises that learners possess finite cognitive resources, which they allocate to different linguistic dimensions, such as complexity, accuracy, and fluency (Skehan, 1998, 2009). When tasks become more demanding, learners may enhance lexical complexity at the expense of accuracy. In contrast, Robinson's CH proposes that cognitively demanding tasks can enhance linguistic complexity and accuracy, provided learners receive adequate support and structure (Robinson, 2001a, 2005; Robinson, 2022a).

Moreover, previous studies indicate that planning time usually increases lexical variety as learners have opportunities to retrieve more advanced lexicon (Bui & Huang, 2020b). The number of task elements has been shown to stimulate sophisticated verb use and diverse noun phrases (Abdi Tabari, Johnson, et al., 2023; Güvendir & Uzun, 2023). Lexical density and adjective variation appear less reactive to short-term changes in task conditions, possibly due to constraints such as word count limits (Soh, 2022; Xu et al., 2022).

Furthermore, automated analysis tools, including Lu's LCA (Lu, 2010, 2012), enable researchers to capture multiple facets of lexis, including type/token ratio, verb sophistication,

and the Uber index, thus providing a nuanced view of lexical development (Crossley, 2020; Lu, 2010, 2012). However, discrepancies often emerge between post-intervention scores (QM) and pre-intervention scores (PS), indicating that learners may overestimate their lexical use (Lee, 2020b). Overall, task complexity and planning time are crucial for advancing lexical complexity, although variables such as proficiency and individual strategies also shape outcomes. More research should utilise quantitative methods to investigate further how cognitive factors and task design promote or constrain linguistic development.

However, extant research highlights the impact of task complexity on lexical complexity; many studies oversimplify the role of planning time by treating it as a uniform, minute-based construct (Kim, 2022b). More nuanced analyses investigate how learners utilise planning time; some may concentrate on refining lexical choices, whereas others merely highlight the content. Such micro-level insights could explain why certain TTR metrics remain unaffected for specific learners (Bui & Huang, 2020b). Additionally, the discrepancy between the pre-intervention scores (PS) and post-intervention scores (objective measurements) (QM) of lexical sophistication warrants deeper inquiry. Existing research indicates that learners often perceive their vocabulary as more advanced than objective data suggests; however, the underlying factors contributing to this discrepancy, such as overconfidence or anxiety, remain unexplored (Güvendir & Uzun, 2023; Lee, 2020a). Bridging the gap could enable targeted pedagogical interventions that connect standardised testing with measured lexical proficiency.

Simultaneously, numerous studies emphasised measures, such as overall TTR, rather than examining how distinct categories, including modifiers and discipline-specific nouns, respond to complexity changes (Xu et al., 2022). Such omissions may be insufficient to account for context-specific lexical requirements. While Skehan's LACM and Robinson's CH remain central theories, future research would benefit from an integrated model that accounts for individual cognitive factors. This could more accurately capture the variability in how learners leverage heightened task complexity to produce complex lexical output.

In summary, primarily guided by Skehan's LACM and Robinson's CH, with support from an automated analytical LCA tool (Lu, 2010, 2012), research on the impact of task complexity on lexical complexity continues to develop. Although previous research indicates that tasks with moderate complexity and sufficient planning time promote advanced lexis. However, critical questions remain regarding genre-related constraints, individual learner differences, and the interplay between objective measurements and learners' subjective perceptions. This underscores the importance of continued empirical exploration and a closer investigation between theoretical propositions and practical, data-driven insights into TBLT research. Indeed, the impact of task complexity on CET4 writing remains underexplored, especially using multi-dimensional lexical analyses.

Research Objectives
This study aims to explore how variations in planning time (PT) and number of task elements (FE), two cognitive factors frequently highlighted by Robinson's CH, affect the lexical complexity of undergraduates' CET essay writing. Specifically, it examines three task conditions: "+PT, –FE" (10 minutes of pre-task planning, 2 elements); "–PT, –FE" (no pre-task planning, 2 elements); and "–PT, +FE" (no pre-task planning, 4 elements). While all three groups were originally included to observe how combined or separate manipulations of PT and

FE may shape performance, this paper primarily compares "+PT, –FE" versus "–PT, +FE" for theoretical reasons (i.e., contrasting more time-based planning and more element-based complexity). The third group, "–PT, –FE" serves as an additional, relatively lower-complexity task condition for reference.

Research Hypothesis

Given prior empirical evidence indicating that varying levels of task complexity may exert differential effects on lexical complexity, this study formulates the following research hypotheses.

RH$_0$: There is no significant difference in lexical complexity in CET4 essay writing between the "+PT, –FE" and "–PT, +FE" task conditions.

RH$_1$: Undergraduates exhibit greater lexical complexity in "+PT, –FE" than in "–PT, +FE", as measured by Lu's LCA (2010, 2012).

Method

This study adopts a single-treatment experimental design and employs quantitative methods, with a within-subject repeated measure (PS versus QM), to investigate the influence of task complexity on the lexical complexity of undergraduate essay writing. Following Crossley (2020), the research employs the Lexical Complexity Analyser (LCA) (Lu, 2010, 2012) for the assessment of lexical complexity in students' writing, analysing 25 dimensions of lexical complexity, thereby providing multi-dimensional insights into lexical development under varying cognitive demands.

Participants

Following the requisite approval from the experimental university, 60 non-English major undergraduates were selected to participate in the study and stratified randomly assigned to one of three groups, with 20 individuals allocated to each group. The participants exhibited an intermediate level of English proficiency, as evidenced by their previous CET writing scores and standardised institutional English test results. The participants had an average age of approximately 18.50 years. Notably, none of the participants had had the opportunity to study abroad and that their experience of learning English ranged from 10 to 13 years, with an average of 11.5 years of English learning experience.

Research Design and Task Conditions

The study investigates "+PT, –FE" and "–PT, +FE" task conditions to evaluate the effects of task complexity on lexical performance (Robinson & Gilabert, 2007). The principal aim is to assess the impact of task complexity on undergraduate students' CET writing. The study may offer insights into how cognitive factors affect lexical output when task complexity increases, providing valuable insights for English writing instruction (See Table 1).

Table 1

Overview of Research Design

Task Condition	Planning Time (PT)	Number of Elements (FE)	Lexical Complexity Focus	Author (Year)
G1: (+PT, –FE) N=20	10 minutes	2 elements	Lower element complexity, additional planning time	(Chao, 2015; Crossley, 2020; Qin, 2021; Robinson & Gilabert, 2007)
G2: (–PT, –FE) N=20	No	2 elements	Lexical performance under intermedia support	
G3:(–PT, +FE) N=20	No	4 elements	Lexical performance under increased cognitive load	

Note: G2 serves as a reference with intermediate-complexity demands.

Independent Variables (IVs)

Planning Time (PT): + PT, 10 minutes of pre-task planning; – PT, no planning time, immediate start under time constraints.

Task Elements (FE): +FE, containing four arguments, thereby increasing cognitive load; –FE, containing two arguments.

Dependent Variables (DVs)

Lu's LCA automated tool (Lu, 2010, 2012) will be employed to assess the lexical complexity of essay writing. Representative measures include lexical density, lexical sophistication, and lexical variation. The LCA autonomously evaluates each essay and stores the results for subsequent statistical analysis.

Task Procedure

Pre-existing Score (PS)

Baseline data were obtained from participants' prior institutional writing assessments, which functioned as a quasi-pretest measure. Participants completed a standardised English essay writing task, which required them to produce a 120–180-word essay incorporating two specified task elements. The task was administered with a 30-minute writing limit and an additional 10-minute planning phase. The resulting test data were analysed using the LCA (Lu, 2010, 2012) measurement dimensions.

Experimental Task (QM)

The study employed a random sampling technique to assign 60 participants into three groups of 20. It also employed a one-time experimental approach, comprising two 30–40-minute essay writing tasks, entitled "Selecting a Roommate" by Chao (2015).

Group 1 (+PT, –FE): with a 10-minute pre-task planning period and completed a task comprising two elements, to assess students' lexical complexity performance.

Group 2 (–PT, –FE): completed a two-element task without pre-task planning time, further testing students' adaptive lexical performance under intermediate support.

Group 3 (–PT, +FE): completed a four-element task without pre-task planning, used to observe cognitive load influence on lexical complexity.

Under Robinson's Cognition Hypothesis and SSARC task model (Robinson, 2001b, 2005; Robinson & Gilabert, 2007; Robinson, 2015, 2022b), the tasks commenced with

"+/– PT and +/– FE", with the outcomes evaluated using the LCA (Lu, 2010, 2012) automated tool. Group 1 was required to select two roommates from a pool of four and identify two benefits associated with each within 30 minutes. An additional 10 minutes were allotted for planning to assess the participants' lexical performance. Meanwhile, Group 2 completed the same task without the planning time. Furthermore, Group 3 was presented with a more complex task: selecting four roommates from six and listing four advantages within the same time frame, minus the planning period. Subsequently, Group 1 was allotted a 40-minute timeframe, while Groups 2 and 3 were permitted a 30-minute window. Groups 1 and 2 were required to complete a two-element task, while Group 3 was tasked with a four-element one. This structure permitted the evaluation of the manipulation effect, considering the task's complexity by controlling the "+/– PT" and "+/– FE" dimensions and examining the influence of essay writing instruction on lexical complexity. Moreover, potential confounding variables, such as gender and age, were controlled to ensure an accurate assessment of the treatment's impact.

Instruments

This study utilised the LCA (Lu, 2010, 2012) to evaluate lexical complexity and employed SPSS v.25 to perform Wilcoxon signed-rank tests for paired data comparisons. The task design was grounded in Robinson's CH and SSARC grouping model (Robinson & Gilabert, 2007). All essays (QM) were analysed via Lu's LCA (2010, 2012), which computes 25 lexical complexity metrics (e.g., lexical density, sophistication, variation). Each participant's QM output was then compared to their own PS using non-parametric tests. Although three groups were included, the main comparative analysis focused on the baseline (+PT, –FE) and high-complexity (–PT, +FE) conditions.

Ethical Consideration

The approval was obtained from the Academic Affairs Office of the experimental university. This involved completing the research checklist, obtaining the necessary documentation, and securing approval for the research. In addition, the I-test system, an internally authorised platform for English proficiency assessment, was used. Consent was also obtained from all 60 participants.

Data Analysis

To evaluate the same participant's writing performance before and after the new complex task, this study employed the Wilcoxon Signed-Rank Test. Despite the three-group design, each participant's PS (existing exam score text, processed by LCA) was paired with their new essay's metrics (QM). This design justifies a within-subject analysis, as the same individuals are compared in a "before–after" sense, albeit with different tasks. Pairing was established based on individual scores obtained before and after the intervention. As a non-parametric statistical procedure, the Wilcoxon signed-rank test is particularly suited to small sample sizes with non-normal distributions, enabling the detection of statistically significant within-subject differences. The PS were derived from participants' prior standardised writing assessments, with lexical complexity assessed

using Lu's (2010, 2012) LCA. The QM were obtained following participants' completion of essays under the "–PT, +FE" task condition. Participants' QM and PS scores were paired and compared, with differences ranked by absolute value and signed to reflect either improvement or decline. Positive ranks signified greater lexical complexity in QM output, negative ranks indicated reduced performance, and ties reflected no change between conditions. The test statistic Z and associated p-values were used to determine statistical significance, with the Bonferroni correction applied to account for the 25 separate lexical complexity dimensions under analysis. Given 25 lexical measures, a strict Bonferroni threshold would require $p < 0.002$ for significance (0.05/25). However, this study adopts an exploratory standpoint, reporting original p-values at 0.05 while acknowledging potential Type I error. This study provide both uncorrected $p < 0.05$ and note whether findings also survive more stringent criteria.

Results

The Wilcoxon Signed-Rank test was applied to 25 measurement dimensions using Lu's LCA (Lu, 2010, 2012) to examine the impact of task complexity on lexical complexity. The findings are outlined by comparing baseline (PS) and post-intervention (QM) results, with attention to potential lexical diversity and sophistication shifts. The subsequent examination determines whether the observed p-values reveal substantial changes under cognitively demanding conditions. The investigation focused particularly on verb sophistication indices, type/token ratios, and the Uber index, as these had emerged as sensitive measures. The collected data serve as a foundation for understanding how planning time and task elements impact students' lexical performance in essay writing, offering insights into the factors that shape lexical usage.

Table 2

Wilcoxon Signed-Rank Test: Selected Lexical Indices (PS versus QM)

Measurement Dimensions	Z	Asymptotic Significance (Two-Tailed) (p=0.05)
Lexical Density (LD)	-0.213[b]	0.831
Lexical Sophistication (LS1)	-1.782[b]	0.075
Lexical Sophistication (LS2)	-1.014[b]	0.311
Verb Sophistication (VS1)	-2.176[b]	0.030
Verb Sophistication (VS2)	-2.100[b]	0.036
Corrected VS1 (CVS1)	-2.100[b]	0.036
Number of Different Words (NDW)	-1.784[b]	0.074
NDW (first 50 words) (NDWZ)	-0.892[b]	0.372
NDW (expected random 50) (NDWERZ)	-0.306[c]	0.760
NDW (expected sequence 50) (NDWESZ)	-1.244[b]	0.214
Type/Token Ratio (TTR)	-1.370[b]	0.171
Mean Segmental TTR (MSTTR)	-0.120[b]	0.905
Corrected TTR (CTTR)	-1.988[b]	0.047
Root TTR (RTTR)	-1.988[b]	0.047
Bilogarithmic TTR (logTTR)	-1.653[b]	0.098
Uber Index (Uber)	-2.191[b]	0.028
Verb Variation (VV1)	-0.767[b]	0.443
Squared VV1 (SVV1)	-1.682[b]	0.093
Corrected VV1 (CVV1)	-1.581[b]	0.114
Lexical Word Variation (LV)	-1.257[b]	0.209
Verb Variation (VV2)	-0.776[b]	0.438

Measurement Dimensions	Z	Asymptotic Significance (Two-Tailed) (p=0.05)
Noun Variation (NV)	-2.023[b]	0.043
Adjective Variation (ADJV)	-0.631[c]	0.528
Adverb Variation (ADVV)	-1.620[b]	0.105
Modifier Variation (MODV)	-0.070[c]	0.944

Note. a. Wilcoxon signed-rank test; b. Based on positive ranks; c. Based on negative ranks; d. mainly focus on G3 PS and QM comparisons.

Table 2 represents the Wilcoxon Signed-Rank test outcomes for the 25 dimensions of lexical complexity, comparing PS versus QM primarily in the "+PT, –FE" and "–PT, +FE" contexts, a particular lexical measure, including lexical density (LD), lexical sophistication (LS1, LS2), type/token ratio (TTR), etc. Meanwhile, the Z-values and p-values indicate whether a significant difference emerges, $p < 0.05$, under different task complexities. Based on the experimental data, lexical density (LD) assesses the ratio of content words, nouns, verbs, adjectives, and adverbs to total words, indicating how "information-heavy" a text is. A non-significant Z (-0.213^b, $p = 0.831$) suggests minimal sensitivity to changes in planning time or task elements. Lexical sophistication (LS1, LS2) captures using less frequent or more advanced words; near significant or non-significant results ($p = 0.075$, 0.311) imply partial or inconsistent responsiveness to task manipulations. Verb sophistication (VS1, VS2, CVS1), with Z (-2.176^b, -2.100^b, -2.100^b; $p = 0.030$, 0.036, 0.036), significantly increased under conditions that allowed more cognitive focus, e.g., with planning time, reflecting heightened attention to complex verb forms. The type/token ratio (TTR) and its variants (CTTR, RTTR) gauge lexical diversity. Significant differences for CTTR and RTTR ($p = 0.047$) indicate that these indices respond when learners have enough time or fewer elements to engage with varied vocabulary. The Uber index (UBER)Z (-2.191^b, $p = 0.028$) similarly shows heightened lexical richness under more amenable task conditions, supporting that cognitively taxing tasks can stimulate advanced vocabulary use if appropriately managed. Noun variation (NV) Z (-2.023^b, $p = 0.043$) measures demonstrate sensitivity to how learners allocate attentional resources. In contrast, adjective variation (ADJV) remains non-significant ($p = 0.528$), suggesting that specific lexical categories may be less likely to alternate. Taken together, these findings illustrate that lexical complexity measures are differentially sensitive to task manipulations involving planning time and the number of elements. Sophisticated verbs, corrected TTR, and Uber index notably increase under supportive conditions (+PT, –FE), while other metrics like LD or specific variation indices show limited changes. Consequently, task design that balances complexity against available planning time proves essential to fostering a broad range of lexical complexities in CET essay writing. However, under Bonferroni < 0.002, these are not statistically significant. This suggests an exploratory but noteworthy trend aligned with Robinson's CH (2005), where increased complexity (either from +PT or +FE) can drive more sophisticated verb use.

Lexical Density (LD) and Sophistication (LS1, LS2)
This section succinctly highlights negative, positive, and tied ranks for lexical density and sophistication, illustrating how previous standard tests and quantitative measures diverge when tasks differ in planning time and task elements (See Table 3).

Table 3

Completed Rank Sums for Lexical Complexity Measures (LD, LS1, LS2）

Measure	Negative Ranks (n)	Rank Average	Sum of Ranks	Positive Ranks (n)	Rank Average	Sum of Ranks	Tied Ranks (n)
QM-LD - PS-LD	6	6.50	39.00	10	3.30	33.00	4
QM-LS1 - PS-LS1	10	3.80	38.00	2	2.00	4.00	8
QM-LS2 - PS-LS2	10	6.20	62.00	8	3.50	28.00	2

In Table 3, for lexical density (LD), the larger negative-rank average 6.50 but smaller positive-rank average 3.30 suggests that, although more participants show positive ranks (n=10), those with opposing ranks demonstrate a higher magnitude of difference. Meanwhile, lexical sophistication (LS1) exhibits primarily opposing ranks (n=10), implying cases where PS surpasses QM, but its average rank of 3.80 is relatively moderate. Finally, Lexical Sophistication (LS2) also shows the most negative ranks (n=10), though the sum of ranks is notably high, 62.00, indicating a substantial group difference. These findings imply that lexical complexity measures (LD, LS1, LS2) are sensitive to task manipulations involving planning time and the number of task elements.

Type/Token Ratio (TTR) and its Variants
This section examines type/token ratio (TTR) variations, revealing predominant opposing ranks across measures. While MSTTR is more balanced, CTTR, RTTR, and LOGTTR show a more substantial "PS > QM" tendency.

Table 4

Completed Rank Sums for Lexical Complexity Measures (TTR)

Measure	Negative Ranks (n)	Rank Average	Sum of Ranks	Positive Ranks (n)	Rank Average	Sum of Ranks	Tied Ranks (n)
QM-TTR - PS-TTR	12	5.67	68.00	6	3.67	22.00	2
QM-MSTTR - PS-MSTTR	8	5.88	47.00	10	4.30	43.00	2
QM-CTTR - PS-CTTR	16	5.88	94.00	4	4.00	16.00	0
QM-RTTR - PS-RTTR	16	5.88	94.00	4	4.00	16.00	0
QM-LOGTTR - PS-LOGTTR	14	4.21	59.00	2	6.50	13.00	4

Experimental data in Table 4 compare the type/token ratio (TTR) and its variants (MSTTR, CTTR, RTTR, LOGTTR) between QM and PS conditions. TTR (QM–PS): A higher number of opposing ranks (n=12) suggests that TTR values under PS generally exceed QM, though some participants show the opposite trend (n=6). MSTTR: Balanced negative (n=8) and positive (n=10) ranks imply that MSTTR does not uniformly favour one condition. CTTR, RTTR: Marked opposing ranks (16 each) indicate that PS outperforms QM for most participants. LOGTTR: Mostly negative ranks (n=14) also point toward higher values for PS. These results highlight that TTR metrics are sensitive to task manipulations, possibly due to variations in planning time, the number of elements, or both affecting lexical diversity.

Uber Index and Lexical Word Diversity
This analysis of UBER measures reveals a strong "PS > QM" trend, with nine opposing ranks and only one positive rank. These findings suggest that participants generally perform their lexical diversity as more significant than what is quantitatively measured. See Table 5.

Table 5

Completed Rank Sums for Lexical Complexity Measures (UBER)

Measure	Negative Ranks (n)	Rank Average	Sum of Ranks	Positive Ranks (n)	Rank Average	Sum of Ranks	Tied Ranks (n)
QM-UBER - PS-UBER	18	5.44	98.00	2	6.00	12.00	0

In Table 5, "QM-UBER - PS-UBER" compares QM and PS scores on the Uber Index, which gauges advanced lexical sophistication. The opposing ranks (n=18, average=5.44) imply that, for most participants, PS exceeded QM, indicating a higher previous standard test lexical richness than the quantitatively measured one. By contrast, positive ranks (n=2, average=6.00) suggest only two instances where QM surpassed PS, and tied ranks=0 show no identical scores. These findings illustrate that the Uber index is sensitive to task manipulations, especially regarding planning time and the number of elements. When learners have sufficient planning time or fewer task elements, their quantitatively measured lexical sophistication (QM) may align more closely with or exceed previous standard test richness (PS). Conversely, when time is limited or tasks are more complex, previous standard assessment appears to outstrip what is quantitatively captured.

Verb Diversity and Lexical Word Variation (VV1, VV2, SVV1, CVV1, LV, NV)
This analysis of lexical variation measures, VV1, SVV1, CVV1, LV, VV2, and NV, indicates a predominant "PS > QM" trend, with negative ranks consistently outnumbering positive ranks. These findings suggest that participants generally perform their lexical variation as higher than quantitative assessments indicate.

Table 6

Completed Rank Sums for Lexical Complexity Measures (VV1, VV2, SVV1, CVV1, LV, NV)

Measure	Negative Ranks (n)	Rank Average	Sum of Ranks	Positive Ranks (n)	Rank Average	Sum of Ranks	Tied Ranks (n)
QM-VV1 - PS-VV1	12	5.83	70.00	8	5.00	40.00	0
QM-SVV1 - PS-SVV1	14	6.29	88.00	6	3.67	22.00	0
QM-CVV1 - PS-CVV1	14	6.14	86.00	6	4.00	24.00	0
QM-LV - PS-LV	14	4.71	66.00	4	6.00	24.00	2
QM-VV2 - PS-VV2	12	4.83	58.00	6	5.33	32.00	2
QM-NV - PS-NV	14	5.64	79.00	4	2.75	11.00	2

In Table 6, opposing ranks (n=14 for SVV1) indicate instances where PS scores surpassed QM scores, suggesting that learners' previous standard test lexical variation exceeded the quantitative calculations. Positive ranks (n=8 for VV1) occurred when QM outperformed PS, implying more measured variation than the previous standard test. The sum of ranks shows the magnitude of these differences, while tied ranks represent identical scores. Overall, the firm opposing ranks across VV1, SVV1, CVV1, LV, VV2, and NV imply that previous standard test lexical variation often outstrips its quantitative counterpart. These findings demonstrate that lexical variation measures are sensitive to task manipulations, such as planning time and the number of elements. When planning time is restricted or task elements increase, learners' previous standard test richness may

diverge from measured outcomes, reflecting the interplay between cognitive load and vocabulary choices.

Adjective, Adverb, and Modifier Variation (ADJV, ADVV, MODV)
This analysis of adjectival, adverbial, and modal verb variation (ADJV, ADVV, MODV) shows mixed trends, with varying distributions of negative and positive ranks. The findings suggest that perceptions of lexical variation differ across categories.

Table 7

Completed Rank Sums for Lexical Complexity Measures (ADJV, ADVV, MODV)

Measure	Negative Ranks (n)	Rank Average	Sum of Ranks	Positive Ranks (n)	Rank Average	Sum of Ranks	Tied Ranks (n)
QM-ADJV - PS-ADJV	6	2.50	15.00	6	4.50	27.00	8
QM-ADVV - PS-ADVV	12	3.92	47.00	2	4.50	9.00	6
QM-MODV - PS-MODV	8	4.38	35.00	8	4.63	37.00	4

In Table 7, QM versus PS examines how learners perform adjectives (ADJV), adverbs (ADVV), and modifiers (MODV) variations compared to measured indices. Opposing ranks occur when PS exceeds QM, while positive ranks indicate QM outpacing PS. For ADJV, an equal split of negative (n=6) and positive ranks (n=6) suggests similar previous standard test versus measured adjective variation frequencies, with eight tied ranks indicating frequent alignment. In ADVV, predominantly negative ranks (n=12) highlight that previous standard test adverb variation often surpasses measured values, whereas just two positive ranks show limited instances of higher measured variation. Meanwhile, MODV exhibits an equal distribution of negative (n=8) and positive ranks (n=8), implying a balanced discrepancy or agreement between perception and measurement. These lexical complexity measures demonstrate sensitivity to task conditions, especially planning time and the number of elements, which can amplify or reduce differences between the previous standard test and measured variation.

Discussion
The findings of this study, particularly concerning verb sophistication (VS1, VS2, CVS1) and the Uber index, align with recent research emphasising the impact of planning time and task complexity on lexical production in essay writing. Specifically, the observed increase in verb sophistication under conditions featuring planning time or fewer elements corroborates Kim's (2022a) assertion that learners perform better lexically when afforded the additional cognitive capacity to organise and refine the complex language. Similar studies, such as Güvendir and Uzun (2023), also found that tasks with moderate complexity and supportive conditions, e.g., sufficient planning time, promote advanced lexical usage, reinforcing that an optimal level of challenge can drive lexical gains.

In addition, the significant effects of noun variation (NV) dovetail with those of Abdi Tabari et al. (2023), who emphasise the importance of nominal structures in enhancing lexical variety. Their research shows that as tasks become more cognitively demanding, learners diversify their noun usage to convey more nuanced content. Meanwhile, the non-significant results for lexical density (LD) and specific variation measures, e.g., adjective

variation, resonate with Trebits and Kormos (2019), who argue that some global or category-specific lexical measures may remain stable under varying levels of complexity, especially when tasks are relatively short or thematically constrained.

Moreover, the partial responsiveness of type-token ratio (TTR) and its variants to task changes echoes Kang and Lee (2019), who account that TTR-based measures, though reflective of lexical diversity, may not always capture shifts in advanced lexis or complex verb constructions. Taken together, these parallels underscore the general consistency of our results with existing literature: complex tasks augmented by planning time tend to stimulate advanced lexicon deployment, whereas certain measures, such as LD or specific variation indices, are less sensitive to short-term manipulations in complexity. Consequently, this study contributes to the growing consensus that cognitive support, planning time, task structuring, and the number of elements are pivotal in shaping lexical development in essay writing.

Discrepancies from Other Findings

Despite aligning with much contemporary research, a few discrepancies arise. First, whereas some studies, including Soh (2022), report significant increases in lexical density when tasks grow more complex, our data show LD to be largely unaffected ($Z = –0.213$, $p = 0.831$). One potential explanation lies in task design; tasks featuring confined topics or stricter word-count limits may inherently cap the density of content words, thus dampening any measurable differences. Meanwhile, in specific investigations, Xu et al. (2022) have found adjective variation to escalate under heightened task demands, whereas our results indicate ADJV remains relatively stable ($p = 0.528$). This might reflect the disciplinary or genre-specific nature of our essay prompts, which could emphasise argumentative or narrative structures less reliant on descriptive adjectives.

Additionally, while planning time was generally beneficial, MSTTR did not shift significantly for some participants, contrasting with findings from Bui and Huang (2020), who observed robust gains in lexical diversity with extended pre-task planning. Notably, it is possible that some learners used planning time primarily for conceptual outlining rather than refining linguistic form. These discrepancies underscore the complexity of second language writing research and highlight how task type, learner strategy, and topic constraints can influence the relationship between task complexity and lexical outcomes.

Support for Skehan's LACM and Robinson's CH

The results support LACM by Skehan (1998, 2009, 2011), which posits that learners operate within finite cognitive resources, distributing attention among complexity, accuracy, and fluency. The significant findings in verb sophistication (VS1, VS2) and noun variation (NV) suggest that learners can focus more on lexical expansion under reduced task elements or added planning time, as Skehan's LACM predicts. By contrast, the non-significance for specific measures, such as LD and ADJV, could indicate that once learners invest heavily in complex verbs or nouns, they have fewer attentional resources left for other dimensions, thus revealing a trade-off. Furthermore, the pattern wherein TTR variants (CTTR, RTTR) show some responsiveness, while LOGTTR remains non-significant, may demonstrate LACM's principle that task-induced cognitive load variably

influences different lexical measures. If a task strongly emphasises precision in verb usage, learners might allocate more cognitive effort there, leaving less capacity for exploring a wide range of descriptive adjectives. Consequently, the findings support Skehan's claim that attentional capacity is limited and selectively allocated, affecting how learners respond to specific tasks and lexical demands.

The evidence likewise bolsters Robinson's CH (2001, 2005), which asserts that increased task complexity can spur greater linguistic sophistication when learners are tasked with multiple elements or cognitively challenging prompts. Notably, the significant improvements in verb sophistication (VS1, VS2, CVS1) and the Uber index ($p = 0.028$) align with Robinson's proposition that cognitively intensified tasks encourage learners to tap into more advanced linguistic resources. When combined with planning time, these tasks stimulate the deep processing of vocabulary and the strategic application of more complex lexis. Specific measures, such as noun variation, also reached significance, highlighting CH's stance that complexity can enhance morphosyntactic and lexical components. Meanwhile, the partial or non-significant outcomes for measures like LS1, LS2, and adjective variation underscore Robinson's acknowledgement that not all language features respond uniformly. Factors such as the learner's proficiency level, topic familiarity, and personal strategies can moderate how effectively they leverage additional cognitive demands. Overall, the pattern of results, where advanced lexical features notably increase under more complex or resource-rich tasks, supports CH's central claim that cognitive complexity, correctly harnessed, can influence more elaborate linguistic performance.

In summary, these results parallel recent studies indicating that appropriately challenging tasks and sufficient planning time foster advanced lexical usage by Kim (2022a) and Trebits and Kormos (2019). They also confirm key tenets of Skehan's LACM, highlighting the selective allocation of finite cognitive resources. Robinson's CH shows that greater complexity triggers deeper lexical processing. However, discrepancies with some investigations illuminate how topic, task type, and learner proficiency can mediate these effects. For practitioners, balancing cognitive demands, such as adding or subtracting task elements with supportive strategies and planning time, can yield meaningful gains in lexical complexity. Future research may explore individual learner differences, varied genres, and longitudinal designs to refine our understanding of optimising task-based writing instruction.

Conclusion

This study investigated how planning time (+PT/–PT) and number of task elements (+FE/–FE), representing cognitive complexity manipulations grounded in Skehan's LACM and Robinson's CH, affect lexical complexity in undergraduates' CET4 essay writing. The analysis of lexical complexity measures in this study provides insights into the relationship between cognitive load, attentional resources, and language proficiency. The findings are consistent with Robinson's CH and Skehan's LACM and provide a nuanced understanding of how linguistic measures such as lexical density, lexical sophistication, and variation in verb and noun usage correlate with cognitive demands. The study outlined significant results for lexical sophistication measures, specifically Verb Sophistication (VS1 and VS2), with p-values of

0.030 and 0.036, respectively. These results support Robinson's claim that more complex linguistic structures require more significant cognitive resources. Similarly, the significance of the lexical sophistication measures (LS1 and LS2) approach suggests a potential link between lexical richness and proficiency. However, lexical density (LD) and type/token ratio (TTR) yield insignificant results, suggesting a weaker relationship with language proficiency, which may challenge some recent studies highlighting the importance of lexical density. Meanwhile, the study examined the rank sums for measures of lexical complexity such as LD, LS1, and LS2. The results reinforce that more sophisticated forms, such as lexical items used in higher proficiency language production, correspond to greater cognitive demands, consistent with Robinson's CH. Conversely, the study highlights type-token ratios (TTR), further suggesting that measures of lexical variation influence cognitive load and proficiency. Furthermore, the study presented rank sums for the Uber Index (UBER), reinforcing findings that sophisticated lexical variation correlates with cognitive load. The study examines variation in different parts of speech (e.g., verb, noun, adjective, adverb). Significant results for noun variation (NV) ($p = 0.043$) suggest that noun usage plays a pivotal role in proficiency, consistent with recent studies on lexical variation and language development. However, a limitation of the analysis is the lack of comprehensive interpretation for some nonsignificant findings, particularly those related to lexical density (LD) and type/token ratio (TTR), crucial variables in lexical studies. To clarify why only the "+PT, –FE" and "–PT, +FE" conditions are examined in detail, this study prioritises a within-group analysis of these specific treatments due to their theoretical significance and limited scope. While the original design included three groups to compare planning time and task elements, data from the specific group were not subjected to detailed analysis. This approach allows for a more focused investigation of the impact of "+PT, –FE" and "–PT, +FE" on lexical complexity, with future research planned to provide a more comprehensive understanding of all experimental conditions. Future research could explore the specific conditions under which these measures may show significance in more detail. Additionally, while the study draws from established models, a deeper examination of the interaction between these lexical measures and other variables, such as task difficulty or learner proficiency levels, would provide a more comprehensive understanding. Moreover, the research sheds light on the role of lexical complexity in language learning. Future studies should aim to address the limitations related to insignificant findings and explore additional variables that could more comprehensively influence CET writing proficiency.

ORCID

https://orcid.org/0000-0002-0793-0230

https://orcid.org/0000-0002-8301-8198

https://orcid.org/0000-0003-3742-5945

Funding

Ethics Declarations

This study adhered to the ethical considerations of the experimental university. All participants gave informed consent, and all research activities were conducted in compliance with institutional ethical guidelines.

Competing Interests

The authors declare no competing interests.

Rights and Permissions

Open Access

References

Abdi Tabari, M. M., Johnson, M. D., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing, 58*, 100783. https://doi.org/10.1016/j.asw.2023.100783

Abdi Tabari, M. M., Khezrlou, S., & Tian, J. (2023). Verb argument construction complexity indices and L2 written production: Effects of task complexity and task repetition. *Innovation in Language Learning and Teaching, 18*(1), 1–16. https://doi.org/10.1080/17501229.2023.2211955

Abdi Tabari, M. M., Lu, X., & Wang, W. (2023). The effects of task complexity on lexical complexity in L2 writing: An exploratory study. *System, 114*, Article 103021. https://doi.org/10.1016/j.system.2023.103021

Abdi Tabari, M. (2016). The effects of planning time on complexity, accuracy, fluency, and lexical variety in L2 descriptive writing. *Asian-Pacific Journal of Second and Foreign Language Education, 1*(1), Article 10. https://doi.org/10.1186/s40862-016-0015-6

Allaw, E. (2021). A learner corpus analysis: Effects of task complexity, task type, and L1 & L2 similarity on propositional and linguistic complexity. *IRAL – International Review of Applied Linguistics in Language Teaching, 59*(4), 569–604. https://doi.org/10.1515/iral-2018-0294

Awwad, A., & Tavakoli, P. (2022). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *IRAL – International Review of Applied Linguistics in Language Teaching, 60*(2), 169–196. https://doi.org/10.1515/iral-2018-0378

Bui, G., & Huang, Z. (2020a). The impact of planning time on complexity, accuracy, and fluency in second language writing. *Journal of Second Language Writing, 48*, 100702. https://doi.org/10.1016/j.jslw.2019.100702

Bui, G., & Huang, Z. (2020b). L2 learners' use of planning time and its impact on complexity, accuracy, and fluency in L2 writing. *System, 90*, 102227. https://doi.org/10.1016/j.system.2020.102227

Crossley, S. A. (2020). A corpus analysis of academic writing and how it informs writing instruction on a university pre-sessional course. *Journal of Writing Research, 11*(3), 415-443. https://doi.org/10.17239/jwr.2020.11.03.03

DeCoursey, C. A. (2023). The second-language essay as cognitive task: Complexity, subjectivity and emotion. *International Journal of English Linguistics, 13*(1). https://doi.org/10.5539/ijel.v13n1p1

Ellis, R. (2018a). *Reflections on task-based language teaching*. Multilingual Matters. https://doi.org/doi:10.21832/9781788920148

Ellis, R. (2018b). Towards a modular language curriculum for using tasks. *Language Teaching Research, 23*(4), 454-475. https://doi.org/10.1177/1362168818765315

Ellis, R. (2005). Planning and task performance in a second language. In Task-based language learning and teaching (pp. 159-175). Oxford University Press

Golparvar, S. E., & Rashidi, F. (2021). The effect of task complexity on integrated writing performance: The case of multiple-text source-based writing. *System, 99*, 102524. https://doi.org/10.1016/j.system.2021.102524

Gomez Laich, M. P., & Taguchi, N. (2023). Effects of task complexity on the learning of genre-specific rhetorical moves and linguistic forms: The case of contrast and argumentative essays. *International Review of Applied Linguistics in Language Teaching, 61*(2), 571–599. https://doi.org/10.1515/iral-2021-0111

Güvendir, M. A., & Uzun, K. (2023). L2 writing anxiety, working memory, and task complexity in L2 written performance. *Journal of Second Language Writing, 60*, 101016. https://doi.org/10.1016/j.jslw.2023.101016

Hsu, W.-C. (2020). The impact of task complexity on patterns of interaction during web-based asynchronous collaborative writing tasks. *System, 93*, 102328. https://doi.org/10.1016/j.system.2020.102328

Kang, S., & Lee, J. H. (2019). Are two heads always better than one? The effects of collaborative planning on L2 writing in relation to task complexity. *Journal of Second Language Writing, 45*, 61–72. https://doi.org/10.1016/j.jslw.2019.08.001

Khatib, M., & Farahanynia, M. (2020). Planning conditions (strategic planning, task repetition, and joint planning), cognitive task complexity, and task type: Effects on L2 oral performance. *System, 93*, 102297. https://doi.org/10.1016/j.system.2020.102297

Kim, Y. (2022a). Effects of task complexity on EFL learners' writing behaviors and performance. *English Teaching, 77*(4), 49–69.

Kim, Y. (2022b). Examining the role of planning time in second language writing tasks: A meta-analysis. *Language Learning & Technology, 26*(2), 35–59. https://doi.org/10.125/llt.20220602

Lee, H. (2020a). An investigation of how task closure and task complexity affect English L2 writing. *Korean Journal of English Language and Linguistics, 20*, 517–540. https://doi.org/10.15738/kjell.20..202009.517

Lee, H. (2020b). Task closure and task complexity effects on L2 written performance. *Journal of Second Language Writing, 50*, 100777. https://doi.org/10.1016/j.jslw.2020.100777

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232.x

NCETC. (2016). *National College English Test Band 4 and Band 6 syllabus (2016 revised edition)*. National Education Examinations Authority.

Qin, X. Y. (2021). *A study of the effects of task complexity on English majors' writing performance* [Master's thesis, Northwest Normal University]. CNKI. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD202201&filename=1021660869.nh

Révész, A., Michel, M., & Lee, M. (2019). The effects of task complexity on linguistic complexity and accuracy: A meta-analysis. *Studies in Second Language Acquisition, 41*(5), 1199–1222. https://doi.org/10.1017/S0272263119000326

Robinson, P. (Ed.). (2001a). *Cognition and second language instruction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524780

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Blackwell.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL – International Review of Applied Linguistics in Language Teaching, 43*(1), 1–32. https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching, 45*(3), 161–176.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (Vol. 2007, pp. 7–27). Multilingual Matters. https://doi.org/10.21832/9781853599286-004

Robinson, P. (2011a). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 3–37). John Benjamins.

Robinson, P. (2011b). *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance*. John Benjamins Publishing Company. https://doi.org/10.1075/tblt.2

Robinson, P. (2011c). Task-based language learning: A review of issues. *Language Learning, 61*(s1), 1–36. https://doi.org/10.1111/j.1467-9922.2011.00641.x

Robinson, P. (2015). The cognition hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (Vol. 8, pp. 87–121). John Benjamins.

Robinson, P. (2022a). The Cognition Hypothesis, task complexity, and L2 development: The SSARC model at 20. In M. del Pilar Garcia Mayo & M. Ahmadian (Eds.), *Recent research in TBLT: Key insights and new avenues* (pp. 25–46). John Benjamins.

Robinson, P. (2022b). The Cognition Hypothesis, the Triadic Componential Framework and the SSARC Model: An instructional design theory of pedagogic task sequencing. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 205–225). Cambridge University Press. https://doi.org/10.1017/9781108868327.013

Robinson, P., Baralt, M., & Gilabert, R. (2014). *Task sequencing and instructed second language learning*. John Benjamins Publishing Company.

Sauer, L., & Ellis, R. (2019). The social lives of adolescent study abroad learners and their L2 development. *The Modern Language Journal, 103*(4), 739–762. https://doi.org/10.1111/modl.12589

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*(3), 185–211. https://doi.org/10.1177/136216889700100302

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press. https://www.degruyter.com/database/COGBIB/entry/cogbib.11126/html

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Skehan, P. (2011). *Researching tasks: Performance, assessment, pedagogy*. Pearson Education.

Skehan, P. (2013). Tasks and language performance assessment. In M. Bygate (Ed.), *Researching pedagogic tasks* (pp. 167–185). Routledge.

Skehan, P. (2014). *Processing perspectives on task performance*. John Benjamins Publishing Company. https://doi.org/10.1075/lllt.39

Skehan, P. (2018). *Second language task-based performance: Theory, research, assessment*. Routledge. https://doi.org/10.4324/9781315629766

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.009

Skehan, P., & Foster, P. (2008). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Soh, S. B. (2020). *The effects of task complexity and task condition on L2 individual writing and peer interaction* [Master's thesis, Universiti Malaya].

Soh, S. B. (2022). Task complexity and pre-writing condition: Exploring Malaysian L2 learners' perceptions on argumentative writing. *3L: Language, Linguistics, Literature, 28*(4), 1–14.

Soh, S. B., Tam, S. S., & Nikitina, L. (2020). The effects of task complexity and peer groupings on individual L2 argumentative writing: A study among Malaysian ESL learners. *Вопросы когнитивной лингвистики*, (1), 22–39.

Trebits, A., & Kormos, J. (2019). The impact of task complexity on linguistic performance: A study on cognitive load and second language production. *Second Language Research, 35*(1), 41–61. https://doi.org/10.1177/0267658317722382

Xu, Y., Zhang, S., & Gaffney, J. (2022). Examining the relative effects of task complexity and cognitive demands on students' writing in a second language. *Studies in Second Language Acquisition, 44*(2), 483–506. https://doi.org/10.1017/S0272263121000310