BILINGUAL PUBLISHING GROUP

**Forum for Linguistic Studies**

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

# The Impact of Cognitive Task Complexity and Task Sequence on L2 Speaking Performance: A Technology-Mediated TBLT Study

***Wanyi Zhang*** * , ***Justin James Bartlett***

*Faculty of Humanities, Srinakharinwirot University, Bangkok 10110 , Thailand*

## ABSTRACT

This study investigates how varying cognitive complexity levels of tasks and three predetermined task-sequence orders influence second language (L2) learners' spoken performance within a technology-mediated task-based language teaching (TMTBLT) environment. Participants completed three monologic crime-reporting tasks, systematically designed to represent Simple, Middle, and Complex cognitive demands, presented in ascending, descending, or interleaved mixed sequences. Analysis focusing on syntactic complexity, lexical diversity, accuracy, and fluency demonstrated a partial inverted-U trend. Specifically, the Middle-level task frequently produced the most balanced complexity, accuracy, lexis, and fluency (CALF) profile across several linguistic dimensions, although this pattern was not consistently observed across all measures. In contrast, the Complex task typically enhanced accuracy but simultaneously constrained lexical diversity. Task sequencing exerted a noticeable effect primarily on fluency outcomes, with ascending sequences facilitating progressive fluency improvements over time. Conversely, beginning with the most challenging task initially diminished speaking speed but ultimately triggered notable recovery in fluency performance towards task completion. These findings lend empirical support to Robinson's Cognition Hypothesis and Skehan's attentional trade-off model, underscoring the potential for moderately challenging tasks, coupled with thoughtfully structured sequencing, to optimize oral proficiency development in technology-enhanced classroom contexts. Future research should incorporate a neutral baseline task and functional adequacy assessments to further elucidate these observed patterns and extend their generalizability.

***Keywords:*** Cognitive Task Complexity; Task Sequence; Technology-mediated TBLT; L2 Oral Performance; Cognitive Hypothesis; Limited Attentional Capacity

*CORRESPONDING AUTHOR:

Wanyi Zhang, Faculty of Humanities, Srinakharinwirot University, Bangkok 10110 , Thailand, Email: wanyi.zhang@g.swu.ac.th

ARTICLE INFO

Received: 28 April 2025 | Revised: 9 June 2025 | Accepted:17 June 2025 | Published Online: 18 July 2025

DOI: https://doi.org/10.30564/fls.v7i7.9735

CITATION

Zhang, W., Bartlett , J.J., et al., 2025. The Impact of Cognitive Task Complexity and Task Sequence on L2 Speaking Performance: A Technology-Mediated TBLT Study. Forum for Linguistic Studies. 7(7): 918–933. DOI: https://doi.org/10.30564/fls.v7i7.9735

# 1. Introduction

In language education, understanding how task design influences learners' spoken performance is crucial for developing effective teaching strategies. This need has become even more pronounced with the growing popularity of mobile and online language learning applications. As instructors integrate technology-mediated tasks into their curricula, questions arise about how task-related factors, particularly cognitive task complexity and task sequence, might affect second language (L2) speaking outcomes. Addressing these questions can provide a foundation for optimizing Technology-Mediated Task-Based Language Teaching (TMTBLT) practices, both in classroom-based and purely virtual contexts.

This study aligns with Technology-Mediated Task-Based Language Teaching (TMTBLT), which emphasizes authentic, goal-oriented tasks facilitated by digital tools[1]. Despite growing interest in TMTBLT, the combined impact of task complexity and sequencing on learners' oral performance remains underexplored. Cognitive task complexity, defined as the mental or attentional effort required by tasks, likely influences learners' spoken output by imposing varying cognitive loads. Clarifying how complexity impacts syntactic and lexical complexity, accuracy, and fluency (CALF) is crucial for designing tasks that effectively support language development.

Similarly, task sequence could shape how learners gradually adapt to or become fatigued by the demands of each successive task, thus influencing their speaking performance over time. Although prior research in classroom-based Task-Based Language Teaching (TBLT) has highlighted the role of sequence design, there is limited empirical insight into how sequence order interacts with cognitive load in technology-mediated environments. Therefore, building upon the theoretical and empirical literature, this study examines how cognitive task complexity and task sequence might shape different aspects of spoken L2 performance (CALF) in a TMTBLT setting.

Research Questions:

1. How, and to what extent, does increased cognitive complexity of TMTBLT tasks affect various aspects (CALF) of spoken L2 performance?
2. How, and to what extent, do different task sequences of TMTBLT tasks affect various aspects (CALF) of spoken L2 performance?

By focusing on these questions, the study aims to clarify the interplay between cognitive complexity, task order, and speaking outcomes, ultimately informing the design of more effective technology-mediated language tasks.

In order to anchor these research questions in a solid theoretical and empirical context, the following literature review explores the foundational studies on cognitive task complexity, task sequencing, and TMTBLT frameworks. This overview lays the groundwork for the subsequent methodological choices and analysis.

# 2. Literature Review

This section establishes the theoretical and research foundation for the study. It begins by outlining recent developments in language teaching, particularly Technology-Mediated Task-Based Language Teaching (TMTBLT) and its role in technology-assisted learning. It then explores how cognitive task complexity and task sequence interact with L2 speaking performance, drawing on the Cognition Hypothesis and the Limited Attentional Capacity Hypothesis. Additionally, it examines the influence of task order on language output through the SSARC Model.

## 2.1. Traditional Techniques for SLA

In second language acquisition, traditional methods like Audio-Lingual Method (ALM), Grammar-Translation Method (GTM), and Silent Way (SW) each have distinct strategies and limitations: ALM aids pronunciation but limits expression; GTM strengthens grammar but may impede interaction; and SW fosters autonomy but lacks sufficient guidance[2–5]. To overcome these issues, Communicative Language Teaching (CLT) emerged, emphasizing real social interaction and enhancing fluency and communicative competence, yet was criticized for insufficient accuracy training, leading to the integration of form-focused instruction[6–8]. Task-Based Language Teaching (TBLT) further evolved from CLT, prioritizing engagement through meaning-focused tasks[9].

## 2.2. Task-Based Language Teaching (TBLT) and Technology-mediated TBLT (TMTBLT)

Task-Based Language Teaching (TBLT) is a learner-centered instructional framework that emphasizes meaningful, goal-oriented tasks (e.g., booking a hotel) to enhance language use and development [10–12]. TBLT effectively supports communicative competence, oral fluency, and accuracy; promotes learner interaction; and reduces anxiety compared to traditional methods, though it may neglect accuracy development or challenge learners with limited vocabulary [13–15].

Integration of modern technology with TBLT (TMTBLT), through mobile devices or learning management systems, enhances motivation, oral skills, and provides authentic digital skills practice, reducing anxiety and promoting engagement [1,16–18]. However, successful implementation requires appropriate technological resources and refined task-design strategies.

This study explores cognitive task complexity within TMTBLT, guided by Robinson's Cognition Hypothesis (CH). Research indicates that higher task complexity might decrease interaction but enhance accuracy in simpler tasks, with limited effects on linguistic complexity in technology-mediated contexts [19–21]. Moreover, computer-mediated communication (CMC) features like asynchronous interaction and planning time, may challenge CH's predictions [22]. Given the evolving nature of TMTBLT into immersive and intelligent CALL environments [23,24], this study aims to examine how technology shapes task complexity and to identify effective measures for evaluating complexity in TMTBLT.

## 2.3. The Hypothesis of Cognitive Task Complexity

Do task complexity and the sequence of procedures affect language learning? Over the past two decades, TBLT scholars have debated how task complexity and task sequence influence language development [25]. Two principal frameworks dominate this discussion: Robinson's Cognition Hypothesis (CH) [26] and Skehan's Limited Attentional Capacity Model (LAC Model) [27]. Both link cognitive processes to task design but diverge on whether learners can simultaneously prioritize complexity, accuracy, and fluency in L2 output.

### 2.3.1. Skehan's Limited Attentional Capacity Model (Trade-off Theory)

Skehan's Limited Attentional Capacity Model (LAC Model) characterizes task complexity through cognitive complexity, code complexity, and communicative stress [28]. According to this model, increased cognitive load raises task demands, though familiarity with task content can mitigate difficulty [27]. Similar to Robinson's Cognition Hypothesis (CH), LAC acknowledges working memory limitations, but whereas CH claims that higher complexity directs attention to accuracy and structural complexity at fluency's expense, the LAC Model suggests learners must prioritize complexity, accuracy, or fluency due to limited attentional resources [27]. Thus, CH views complexity as enhancing accuracy and structural complexity, while LAC Model emphasizes inevitable trade-offs among linguistic dimensions.

### 2.3.2. Robinson's Cognition Hypothesis

Robinson's Cognition Hypothesis analyzes L2 performance through the Triadic Componential Framework (TCF), which includes (1) *cognitive task complexity*, (2) *task conditions*, and (3) *task difficulty* [29,30]. Within *cognitive task complexity*, two variable types emerge:

**Resource-directing variables** (e.g., ±few elements, ±here and now) channel cognitive resources toward specific linguistic features, often yielding more complex and accurate but less fluent performance.

**Resource-dispersing variables** (e.g., ±prior knowledge, ±planning) broaden cognitive demands, potentially reducing complexity, accuracy, and fluency in more demanding tasks.

Task conditions center on interactional factors, such as participation structures (one-way/two-way) or outcome types (convergent/divergent), while task difficulty involves learner-specific traits like motivation, proficiency, and aptitude.

Recent SLA research has extensively investigated how task complexity influences fluency, accuracy, and complexity [22,26,31–34]. Xu, Zhang, and Gaffney [34] found no significant differences in accuracy, fluency, or syntactic complexity under different complexity conditions, though functional adequacy and lexical complexity improved. In oral production, Sasayama and Izumi [32] observed gains in syntactic complexity but losses in fluency and accuracy, partially aligning with both Robinson's CH and Skehan's LAC. Similarly, Michel [35] reported that more complex tasks enhance lexical diversity in both monologic and dialogic tasks among Dutch ESL learners, though other linguistic dimensions remained unaffected. These mixed results underscore the nuanced ways in which task complexity can shape various aspects of L2 performance.

### 2.4. Robinson's SSARC Model

Robinson's SSARC Model (Simplify, Stabilize/Automatize/Restructure, Complexity) is grounded in his Cognition Hypothesis and specifies that tasks should be sequenced (1) solely based on cognitive factors and (2) in an order from simple to complex [36]. Practically, the model recommends three steps [37]:

Simplify & Stabilize (SS): Task versions remain simple on both resource-directing and resource-dispersing dimensions (e.g., few elements, prior knowledge).

Automatize (A): Task versions are simple on resource-directing but more demanding on resource-dispersing dimensions.

Restructure & Complexify (RC): Task versions are complex on both resource-directing and resource-dispersing dimensions.

Empirical studies [38–40] often compare sequences such as SC versus CS or random. While cognitive complexity has shown significant effects on L2 performance, manipulating sequence alone tends to yield less pronounced impacts. In this study, because task sequence can produce varied outcomes, the present study explicitly manipulates three distinct sequences: simple-middle-complex, complex-middle-simple, and middle-complex-simple to investigate how different orders of tasks shape L2 performance.

## 3. Methodology and Experiment

Having established the study's theoretical foundation in Section 2, this section details the practical procedures and rationale, including methodology, participants, instruments, materials, procedures, coding/scoring, and data analysis.

### 3.1. Participants

Sixty L1 Chinese undergraduate English learners (39 females, 21 males) voluntarily participated in an on-campus experiment at a university in China. All were native Mandarin speakers, aged 19 to 23, and each held an IELTS speaking band score of at least 5, ensuring readiness for the tasks. Recruitment took place via email, which included a background survey and consent confirmation; afterward, the participants were randomly assigned to three groups of 20.

No proficiency-based control group was employed because the primary focus is on cognitive task complexity and its influence on learners' output, rather than on evaluating TMTBLT efficacy. Each participant completed three tasks of increasing cognitive complexity in different sequences; the details of which are presented in the following subsection.

### 3.2. Materials

This section outlines three main components used in the study:

1) Computer-Based TMTBLT Tasks
2) Camera Monitors
3) Performance Measuring Scale

Each of these elements is discussed in detail below, along with the rationale for including three levels of task complexity and different task sequences.

#### 3.2.1. Computer-Based TMTBLT Tasks

*Task Environment:* Sixty participants were each assigned to a private room equipped with a computer preloaded with a custom task program. This setup aimed to minimize external distractions and ensure a standardized environment for every participant. The software interface guided learners step by step through the tasks, displayed the on-screen prompts, and recorded their responses.

*Task Nature and Design:* In this study, a computer-based conversational program was developed to simulate a real-life scenario in which participants act as witnesses reporting a crime to a simulated police operator. Before each interactive session, participants watched a CCTV video depicting a crime scene, complete with details of the offender(s), victim, location, timing, and key incident events. After viewing, they were prompted by the computer to report the incident and respond to follow-up questions from the simulated police operator. A second interactive phase followed, where participants were connected to another simulated officer who requested additional information to clarify or expand upon the incident details.

The seemingly atypical crime-reporting scenario was selected for two pedagogical reasons. First, incident reporting is a frequent and high-stakes communicative task for international students and expatriate professionals (e.g., lost passports, accommodation break-ins). Second, cognitive research in TMTBLT indicates that tasks with significant stakes direct learners' attention to both propositional content and linguistic form, enhancing processing depth and supporting lasting L2 development [41]. Additionally, the macro-structure of these tasks (problem description, narrative reconstruction, causal inference) aligns with common academic genres such as project debriefings and case-study presentations, thereby facilitating their transfer to mainstream academic discourse.

To systematically adjust cognitive load, each task varied in the number and diversity of elements—"Who", "When", "Where", and "What"—across simple (S), middle (M), and complex (C) levels. These adjustments involved adding characters, actions, or settings ( "+elements") in line with Robinson's Cognition Hypothesis (CH) principles. Below is a brief overview of the three complexity tiers.

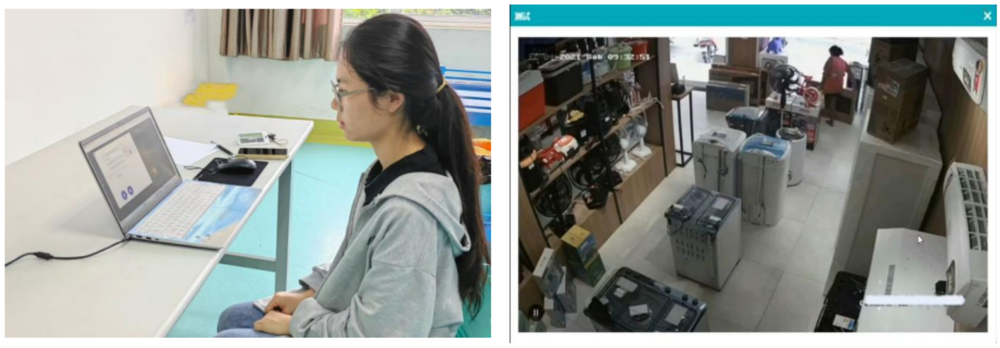
Following Robinson's [30] Triadic Componential Framework, our manipulation specifically targets cognitive complexity in the reasoning dimension (**Table 1**). The complex task introduces additional critical elements, such as multiple suspects, conflicting motives, and contradictory timelines, requiring participants to infer the most plausible scenario from competing information. Therefore, the manipulation is primarily resource-directing (±reasoning) rather than merely resource-dispersing (±prior knowledge). To confirm this characterization, three TBLT experts independently rated each task's reasoning versus memorization demands using a 5-point bipolar scale (1 = purely mnemonic, 5 = predominantly inferential), achieving strong inter-rater reliability (ICC = 0.83). The complex task was rated significantly higher in inferential demand (M = 4.27, SD = 0.49) compared to the medium (M = 3.10, SD = 0.52) and simple tasks (M = 2.05, SD = 0.47); $p < 0.01$, partial $\eta^2 = 0.07$, supporting the validity of the complexity manipulation. However, some short-term memory load inevitably co-occurs with reasoning demands; this limitation is mentioned further in Section 4.

To facilitate understanding, **Figure 1** and **2** include a photo of one participant during the experiment and screenshots of the task as displayed on the computer.

Throughout each interaction, participants' verbal responses were recorded and later transcribed for performance analysis. Although no immediate feedback was provided in this experimental design, real classroom scenarios could incorporate corrective or formative feedback to help learners refine their descriptive accuracy and fluency.

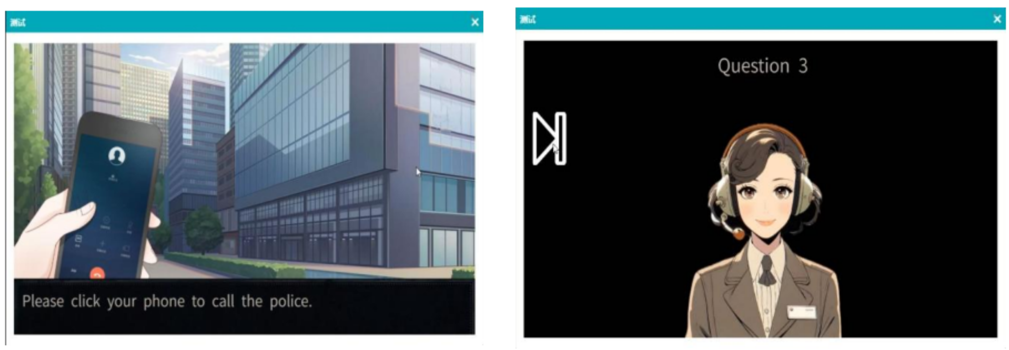
**Table 1.** The Overview of the Three Tasks.

| | Scenario | Details | Key Features |
|---|---|---|---|
| Simple | A daytime theft in an unattended grocery store | A slightly overweight woman with dark skin, wearing a pink top and black pants, seizes the opportunity to steal a new, boxed television | Minimal sequence of events, fewer interacting parties, and a straightforward timeline |
| Middle | A late-night break-in at a bicycle shop near a crossroads | A middle-aged man in an orange jacket and dark pants arrives in an engineering vehicle, deliberately shatters the storefront window, then attaches two new bicycles with a rope and drives away | More elaborate plot, multi-step actions, emphasis on timing (night), and an expanded setting (e.g., crossroads) |
| Complex | Multiple characters and sequential events involving a motorcycle and a speeding car | A motorcycle is knocked over, three distinctly dressed young men arrive in a white car at high speed, the motorcyclist flees, abandoning valuables like a wallet and phone, and the men take both the items and the motorcycle before leaving the scene | Rich interactions (knocking over, fleeing, looting), multiple individuals, and dynamic transitions in a single event chain |



(a) (b)

**Figure 1.** (**a**) One participant testing the procedure before the task began; (**b**) Screenshot of the simple task video.




(a) (b)

**Figure 2.** (**a**) Instructions from the task procedure; (**b**) The interface for conversations with the police operator.

*Rationale for Three Complexity Levels:*

1) Capturing Non-Linear Patterns. When complexity is only "low" or "high," researchers may overlook potential rise-and-fall trajectories in performance (fluency, for instance) that occur between those extremes. Introducing a middle level helps capture a more nuanced progression: learners might initially improve, reach an optimal point, then experience a drop-off if the task becomes too demanding.
2) Differentiated Task Demands. By varying elements such as the number of details to process, the familiarity of the topic, or the mental operations required, each complexity level targets distinct cognitive load conditions. This approach enables a clearer view of how incremental increases in mental effort affect syntactic and lexical complexity, accuracy, and fluency (CALF).

Besides video delivery, the program managed cognitive load through intentional interface design. A single-pane recorder eliminated distractions from chat windows and menus to avoid split attention, and automated voice prompts reminded learners of key information, externalizing part of the memory load. To further optimize Technology-Mediated TBLT, we recommend incorporating a timed planning phase that delays recording briefly to encourage idea organization and providing optional auto-generated transcripts with highlighted lexical features for reflective practice without increasing task demands.

### 3.2.2. Task Sequences and Group Assignments

In addition to varying complexity, the study also manipulates task order. A total of 60 participants were randomly divided into three groups of 20, each completing the tasks in a different sequence:

Group 1: S →M →C

Group 2: C →M →S

Group 3: M →C →S

**Figure 3** shows the design of the study:

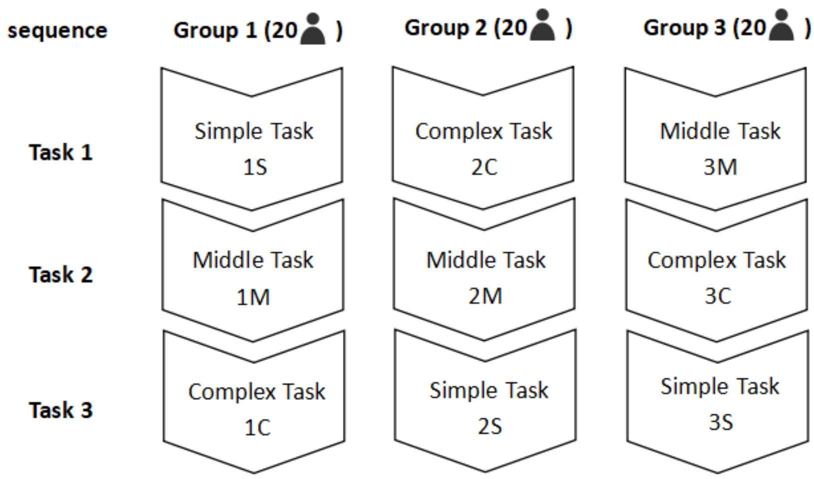



**Figure 3.** The Design of the Study.

*Why Multiple Sequences?*

1) Real-World Simulation

TBLT underscores realistic language use, but in actual contexts, tasks do not follow a predictable "simple-to-complex" progression. Allowing varied orders (e.g., encountering a hard scenario before an easier one) mirrors the unpredictability learners might face outside the classroom, thus enhancing external validity.

2) Avoiding Participant Adaptation

If tasks always progressed strictly from simple to complex (S–M–C), participants might adapt their strategies, diluting the cognitive load differences. The fixed mixed order (e.g., C–M–S) can help capture a more genuine cognitive response to each task's demands, without allowing too much "acclimation" to a fixed pattern.

3) Interacting Variables

The fixed mixed condition used a predetermined interleaved order (M-C-S), following principles recommended in skill-learning pedagogy [42] to complement ascending and descending sequences. Examining interactions between cognitive complexity and task sequence may reveal how learners adapt strategically, indicating potential trade-offs, such as prioritizing accuracy initially and fluency subsequently.

### 3.2.3. Camera Monitors

Video Recording: Each private room was also equipped with a discreet camera monitor. This measure served two purposes: 1) Behavioral Observation: The video feed allowed the researcher to note participants' expressions, gestures, and potential signs of anxiety or disengagement. 2) Verification of Task Engagement: Ensuring participants genuinely completed each stage without external assistance. 3) All participants provided informed consent for video recording, and measures were taken to safeguard their confidentiality (e.g., secure storage, restricted access).

### 3.2.4. Performance Measuring Scale

Consistent with previous TBLT studies [33,43,44], this research adopted CALF dimensions (syntactic and lexical complexity, accuracy, fluency) to assess learners' oral performance. These dimensions form a robust framework widely recognized in ISLA research. Building on Skehan's [45] inclusion of complexity alongside accuracy and fluency [46], the CALF approach explicitly highlights lexical complexity. The sub-dimensions applied in this study, tailored specifically to L2 oral tasks, are summarized in the **Table 2**.

Rather than imposing fixed language proficiency targets, we focused on relative differences across the three participant groups, highlighting how task complexity and sequence might alter performance in complexity, accuracy, and fluency. Following each task, participants' spoken data were transcribed and analyzed according to this CALF framework. Section 4: *Results and Analysis* details how these analyses were conducted and discusses the resulting performance outcomes.

**Table 2.** The CALF Measurement for the Study.

| Category | Subcategory | Measure |
|---|---|---|
| Complexity | Syntactic complexity | Mean length of AS-unit (MLAS)<br>Mean length of clause (MLC) |
| | Lexical complexity | Type-token ratios (TTR) |
| Accuracy | Accuracy and comprehensibility | Error-free clause ratio (EFCR) |
| Fluency | Productivity | Total number of words (excluding repetitions, repairs, reformulations and false-starts) (TNW)<br>Total number of syllables (excluding repetitions, repairs, reformulations and false-starts) (TNS) |
| | Speech rate | Number of words per minutes (WPM)<br>Number of syllables per second (SPS) |

### 3.2.5. Vertical and Horizontal Analyses

We explore cognitive task complexity and sequencing through both vertical (e.g., 1S →1M →1C in Group 1) and horizontal (e.g., 1S vs. 2S vs. 3S across groups) comparisons. Vertical analysis shows how escalating complexity affects learners' L2 output, while horizontal analysis examines whether placing the same complexity level in different positions (first, second, or third) influences performance. Together, these perspectives clarify whether complexity and sequence act independently or interactively, and they also reveal any "cognitive desensitization" from repeated exposure to the same complexity level.

Since the study involves more than two conditions (S-M-C, C-M-S, M-C-S), Analysis of Variance (ANOVA) was chosen for statistical comparisons to simultaneously evaluate multiple groups, reducing inflated Type I errors that occur with multiple t-tests. ANOVA also suits continuous outcome measures such as type–token ratio and error-free clause ratio. Specifically, a one-way ANOVA contrasts complexity levels (S, M, C), whereas a two-way ANOVA tests main effects and interactions between task complexity and sequence order. When omnibus F tests were significant ($p < 0.05$), post-hoc comparisons (Tukey HSD) identified specific group differences. Effect sizes (partial $\eta^2$) accompany significant findings to quantify magnitudes, consistent with APA 7 guidelines [47]. Descriptive statistics (means, standard deviations) supplement inferential results to illustrate learner performance across conditions.

## 4. Results and Analysis

In a single-day experiment, each participant completed three speaking tasks, and audio data from all 60 participants was recorded and transcribed using speech-to-text software. Audio files were transcribed using Whisper v2.1 (medium model). Two trained coders independently assessed 10% of the transcripts for functional adequacy. Inter-rater reliability was high (Cohen's $\kappa = 0.92$), indicating strong agreement. Discrepancies were resolved through consensus.

In the following sections, each group's performance is compared across the S–M, M–C, and S–C cognitive task complexity (CTC) levels. By examining p-values, we assess whether different CTC levels significantly influence L2 performance, referred to here as the "task-level effect".

Additionally, tasks S1-S2-S3, M1-M2-M3, and C1-C2-C3, representing the same CTC levels performed by Groups 1, 2, and 3, respectively, are compared to explore how task order might affect L2 performance, termed the "task-order effect". These analyses together offer a nuanced view of how both task complexity and sequencing shape participants' language outcomes.

### 4.1. Lexical Complexity: Type-Token Ratio

Type-Token Ratio (TTR) measures lexical variety by dividing the number of distinct words (types) by total words (tokens) [48]. A higher TTR indicates greater vocabulary diversity, while a lower TTR suggests more repetition. TTR is widely used to evaluate vocabulary richness and linguistic variation in both written and spoken outputs. **Figure 4** visually depicts the trend in Type-Token Ratio (TTR) for the three groups based on the average TTR values for each group.

Group 1 (SMC). For the task-level effect, TTR shows a clear downward trend with rising task complexity: from S to M ($p = 0.074$), from M to C ($p = 0.007$), and from S to C ($p < 0.001$, $\eta^2 = 0.07$). As cognitive load increases, participants appear to reduce their lexical variety, likely focusing on accuracy or fluency. The task-order effect is also significant ($p$ = 0.001), with TTR steadily declining from S →M →C. This pattern suggests that as tasks progress in complexity, learners devote more resources to structural or accuracy-related demands, constraining their overall vocabulary range.

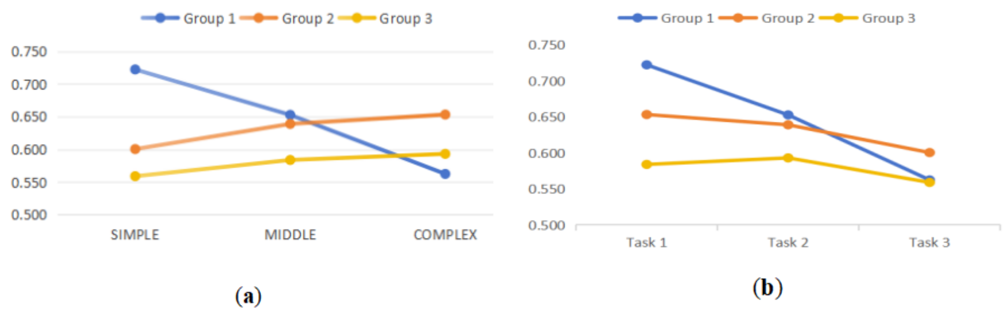



**Figure 4.** (**a**) The TTR Trend by CTC Level; (**b**) The TTR Trend by Task Sequence.

Group 2 (CMS). For the task-level effect, all comparisons remain statistically non-significant: S to M ($p$ = 0.268), M to C ($p$ = 0.682), and S to C ($p$ = 0.106). These results indicate that increased task complexity does not substantially affect TTR for Group 2. However, the task-order effect proves significant ($p$ = 0.004, $\eta^2$ = 0.10), with a steady TTR decline from C to M to S. This suggests that starting with the highest complexity may lead participants to remain cautious or fatigued as tasks become easier, resulting in a net reduction in lexical variety.

Group 3 (MCS). For the task-level effect, none of the pairwise comparisons: S to M ($p$ = 0.462), M to C ($p$ = 0.798), or S to C ($p$ = 0.375) reach significance, implying that rising task complexity does not reliably influence TTR in this group. Likewise, the task-order effect is non-significant ($p$ = 0.142), as TTR shows only mild fluctuations among Middle, Complex, and Simple tasks. Thus, Group 3 exhibits no clear evidence that task complexity or sequence meaningfully alters lexical variety.

## 4.2. Syntactic Complexity: Mean Length of AS-Unit and Clause

Mean Length of AS-unit (MLAS) and Mean Length of Clause (MLC) are key syntactic complexity measures in language proficiency assessment. MLAS averages words or syllables per AS-unit, indicating language sophistication, while MLC measures clause length to reflect syntactic elaboration. Both metrics gauge language development in spoken and written forms. **Figure 5** visually presents the tendency of the Syntactic complexity (both for MLAS and MLC) for the three groups.

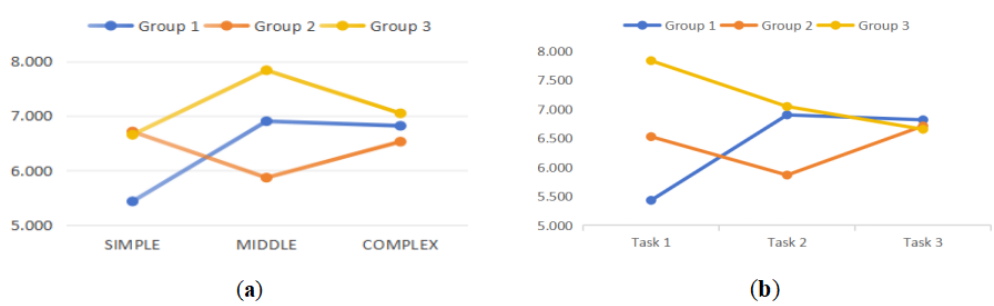



**Figure 5.** (**a**) The MLAS Trend by CTC Level; (**b**) The MLAS Trend by Task Sequence.

Group 1 (SMC). MLAS climbs sharply from S to M ($p$ = 0.003, $\eta^2$ = 0.06) and remains higher at C than at S ($p$ = 0.015), but shows no extra gain from M to C ($p$ = 0.697). The task sequence itself is significant ($p$ = 0.002, $\eta^2$ = 0.09):

length peaks at the middle task, then slips slightly under the heaviest load while staying above the baseline. Overall, moderate cognitive demands trigger the greatest syntactic expansion, whereas further complexity tempers, but does not erase this benefit.

Group 2 (CMS). For the task-level effect, MLAS falls from S to M ($p$ = 0.1404), nudges upward from M to C ($p$ = 0.297), and stays virtually flat between S and C ($p$ = 0.769); none of these differences reach significance, showing that syntactic length scarcely responds to changes in cognitive load. The task-order effect is likewise non-significant ($p$ = 0.117), as the C →M →S progression yields only minor, inconsistent fluctuations. In sum, Group 2 maintains a broadly stable AS-unit length regardless of both task complexity and sequence.

Group 3 (MCS). For the task-level effect, none of the pairwise comparisons reach significance: MLAS rises slightly from S to M ($p$ = 0.069), falls from M to C ($p$ = 0.145), and shows no clear change between S and C ($p$ = 0.515). In contrast, the task-order effect is significant ($p$ = 0.035): MLAS starts highest in the first (Middle) task, drops at Complex, and declines further at Simple, yielding an overall downward trajectory across the M →C →S sequence.

**Figure 6** visually presents the tendency of the Syntactic complexity (both for MLAS and MLC) for the three groups.

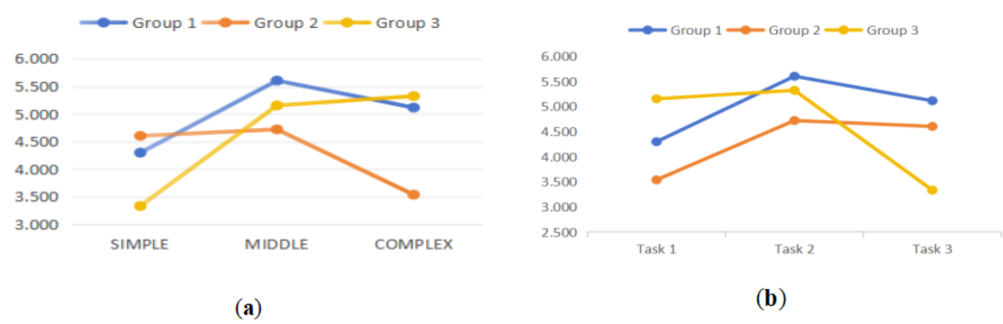



**Figure 6.** (**a**) The MLC Trend by CTC Level; (**b**) The MLC Trend by Task Sequence.

For Group 1 (SMC), the task-level effect shows that mean clause length (MLC) jumps significantly from S to M ($p$ = 0.002, $\eta^2$ = 0.11), slightly increases from M to C ($p$ = 0.220), and is marginally higher at C than at S ($p$ = 0.053). The task-order effect is likewise significant ($p$ = 0.022): clauses lengthen most in the middle task and then ease back under the heaviest load. Together, these findings indicate that moderate complexity prompts the greatest clause elaboration, while additional cognitive demands temper, but do not erase this syntactic gain.

For Group 2 (CMS), the task-level effect shows that mean clause length is unchanged between S and M ($p$ = 0.779), but drops sharply from M to C ($p$ < 0.001, $\eta^2$ = 0.08) and is likewise shorter at C than at S ($p$ = 0.016). Thus, the highest cognitive load curtails syntactic elaboration, whereas moderate load does not. The task-order effect is also highly significant ($p$ < 0.001, $\eta^2$ = 0.06): MLC rises steeply when the sequence moves from the opening Complex task to the subsequent Middle task, then slips slightly at the final Simple task but remains above the starting point. Together, these findings indicate that Group 2's clauses lengthen once the initial strain of the most complex task is removed, but elevated cognitive demands, whether encountered first or in pairwise comparison, consistently compress syntactic length.

For Group 3 (MCS), the task-level effect shows that mean clause length (MLC) rises sharply from S to M ($p$ < 0.001, $\eta^2$ = 0.09) and remains significantly higher at C than at S ($p$ = 0.007, $\eta^2$ = 0.10), but it does not change between M and C ($p$ = 0.673). The task-order effect is extremely pronounced ($p$ < 0.001, $\eta^2$ = 0.11): MLC holds steady from the first (M) to the second (C) task, then drops steeply in the final Simple task. Taken together, these findings suggest that moderate and high complexity both support richer clause elaboration, but when the sequence ends with an easier task, participants revert to markedly shorter clauses, highlighting a strong influence of task order alongside the broader complexity gains.

## 4.3. Accuracy: Error-Free Clause Ratio

The Error-Free Clause Ratio (EFCR) is a linguistic metric used to evaluate the accuracy of language use in both spoken and written discourse. An error-free clause is one that contains no grammatical, syntactic, or lexical errors. A higher EFCR indicates a greater proportion of accurate language production, making it a valuable measure for assessing linguistic accuracy, particularly in language proficiency evaluations. **Figure 7** visually present the tendency of the accuracy (EFCR) for the three groups.

For Group 1 (SMC), the task-level effect shows that EFCR rises from Simple to Middle ($p = 0.261$) and from M to C ($p = 0.347$) without reaching significance at either step, whereas the overall gain from S to C is significant ($p = 0.025$, $\eta^2 = 0.08$) indicating noticeably higher clause-level accuracy at the highest cognitive load compared with the lowest. The task-order effect is likewise significant ($p = 0.041$, $\eta^2 = 0.09$), with accuracy improving steadily across the S →M →C sequence. Together, these findings suggest that while incremental improvements are too small to detect between adjacent tasks, the cumulative effect of rising complexity, and its placement later in the sequence, leads participants to produce significantly more error-free clauses by the final, most demanding task.


**Figure 7.** (**a**) The EFCR Trend by CTC Level; (**b**) The EFCR Trend by Task Sequence.

For Group 2 (CMS), the task-level effect shows no significant change in EFCR across any complexity step—S to M ($p = 0.920$), M to C ($p = 0.371$), and S to C ($p = 0.360$) all exceed the .05 threshold—indicating that accuracy remains steady regardless of cognitive load. Consistently, the task-order effect is non-significant ($p = 0.638$), with only minor, non-systematic fluctuations as the sequence moves from complex to simpler tasks. Together, these results suggest that Group 2 maintains a stable clause-level error rate, unaffected by either task complexity or the order in which tasks are presented.

For Group 3 (MCS), the task-level effect shows no significant change in EFCR between adjacent steps—S to M ($p = 0.384$) and M to C ($p = 0.243$) both exceed .05—whereas the overall jump from S to C is significant ($p = 0.028$, $\eta^2 = 0.11$), indicating higher clause-level accuracy at the greatest cognitive load than at the lowest. In contrast, the task-order effect is non-significant ($p = 0.247$); accuracy rises modestly in the second (C) task before dipping again in the final (S) task, leaving no reliable sequence-driven change. Altogether, Group 3's accuracy improves only when the complexity gap is large, and this improvement is not consistently shaped by task order.

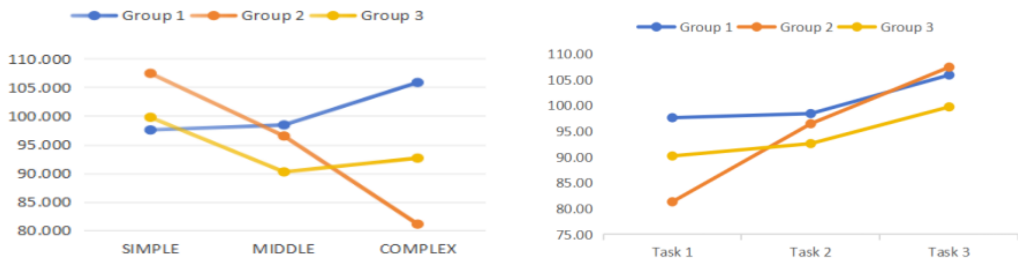
## 4.4. Fluency: Number of Words per Minute & Number of Syllables per Second

Fluency in this study is captured through two complementary speech-rate metrics. Words per minute (WPM) gauges how many words a speaker produces in sixty seconds; higher WPM signals faster, more fluent delivery, whereas lower WPM reflects a slower, more deliberate pace. Syllables per second (SPS) refines this view by counting syllables articulated each second, offering a fine-grained index of articulation speed. Together, these measures provide a straightforward but robust picture of production tempo: higher values on either metric denote greater fluency and rapid articulation, while lower values indicate a more measured speech rate.

**Figure 8** visually presents the tendency of the fluency (both for WPM and SPS) for three groups.

For Group 1 (SMC), the task-level effect shows no significant change in speech rate: WPM rises only marginally from S to M ($p = 0.904$) and from M to C ($p = 0.404$), and the overall S-C contrast is likewise non-significant ($p = 0.361$).

The task-order effect is also non-significant ($p = 0.387$), as the gradual uptick across the S →M →C sequence never surpasses chance-level variation. Taken together, these results indicate that Group 1's speaking speed remains essentially stable, with no reliable acceleration in response to increasing cognitive task complexity or its placement within the sequence.




(**a**) (**b**)

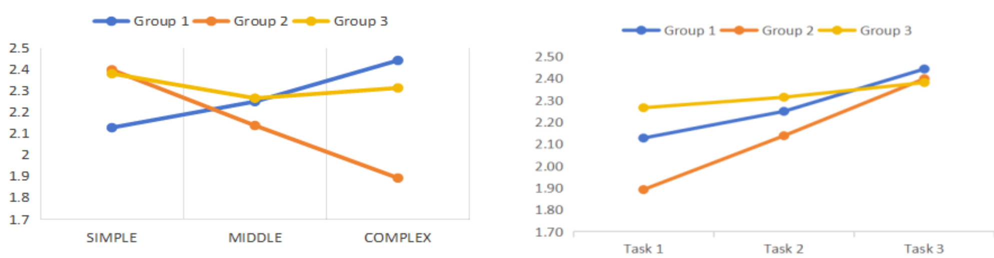
**Figure 8.** (**a**) The WPM Trend by CTC Level; (**b**) The WPM Trend by Task Sequence.

For Group 2 (CMS), the task-level effect shows no significant change in WPM between S and M ($p = 0.187$) or M and C ($p = 0.080$), while speech rate drops significantly when the extremes are compared, with markedly slower delivery at C than at S ($p = 0.004$, $\eta^2 = 0.09$). By contrast, the task-order effect is highly significant ($p = < 0.001$, $\eta^2 = 0.12$): speakers begin the sequence with their slowest rate during the initial Complex task, accelerate during the M task, and reach their fastest WPM on the final S task. Taken together, these findings indicate that Group 2's fluency remains fairly steady across adjacent complexity levels but slows noticeably at the highest load, while the descending complexity order (C →M →S) enables a clear, cumulative acceleration in speech rate.

For Group 3 (MCS), the task-level effect shows no significant change in speech rate: WPM differs neither between S and M ($p = 0.277$) nor between M and C ($p = 0.764$), and the overall S–C contrast is likewise non-significant ($p = 0.429$). The task-order effect is also non-significant ($p = 0.449$), as the modest fluctuations across the M →C →S sequence do not exceed chance. Together, these results indicate that Group 3's speaking speed remains essentially stable, with no reliable adjustment in response to either task complexity or its ordering.

The tendency of the fluency (both for WPM and SPS) for three groups is presented in **Figure 9**.

For Group 1 (SMC), the task-level effect on speech tempo shows a borderline rise in syllables per second from S to M ($p = 0.071$), no additional increase from M to C ($p = 0.547$), but a significant overall gain when S is compared with C ($p = 0.032$). The task-order effect is likewise significant ($p = 0.041$, $\eta^2 = 0.09$): SPS climbs steadily across the S →M →C sequence, indicating that participants speak faster as tasks grow more demanding in order. Together, these results suggest that moderate complexity initiates a partial acceleration which becomes statistically clear only when the lowest and highest loads are contrasted, while the ascending sequence itself reinforces this cumulative boost in articulation speed.




(**a**) (**b**)

**Figure 9.** (**a**) The SPS Trend by CTC Level; (**b**) The SPS Trend by Task Sequence.

For Group 2 (CMS), the task-level effect shows no reliable change in syllables per second between S and M ($p$ = 0.068) or M and C ($p$ = 0.095); however, speech rate drops significantly when the extremes are compared, with markedly slower articulation at C than at S ($p < 0.001$, $\eta^2 = 0.07$). By contrast, the task-order effect is significant ($p$ = 0.003, $\eta^2 = 0.08$): SPS rises steadily across the C →M →S sequence, so participants finish the session speaking fastest on the final, easiest task. Together, these findings indicate that Group 2 slows noticeably under the heaviest cognitive load, and accelerates cumulatively as task demands lighten in order.

For Group 3 (MCS), neither complexity nor order alters articulation speed in a reliable way. The task-level effect shows that SPS remains statistically unchanged across all comparisons: S to M ($p$ = 0.316), M to C ($p$ = 0.388), and S to C ($p$ = 0.620). Consistently, the task-order effect is non-significant as well ($p$ = 0.449), with only minor, chance-level fluctuations across the M →C →S sequence. In short, Group 3 maintains a stable syllable-per-second rate regardless of task complexity or its position in the sequence.

### 4.5. Analysis and Discussion

The results highlight distinct patterns of learner responses to increased cognitive demands, aligning clearly with theoretical frameworks. Group 1 mirrored Robinson's Cognition Hypothesis (CH), enhancing syntactic complexity and accuracy as tasks became more challenging, but experiencing a marked decrease in lexical variety under maximal load. This suggests syntax gains persist only until attentional resources become fully occupied, at which point fluency and lexical diversity decline first.

Group 2, in contrast, immediately slowed down and produced shorter clauses as complexity increased, without compensatory improvements in accuracy or vocabulary. This pattern aligns closely with Skehan's[27] Limited Attentional Capacity model and echoes findings from Xu et al.[34] and Donate[33]. Group 3 exhibited a mixed profile, displaying modest gains in accuracy and productivity only when complexity differences were substantial. This intermediate pattern resembles findings from Rahimi & Zhang[49] and Frear & Bitchener[50].

The sharp reduction in lexical variety observed under high cognitive load across groups supports Skehan's model. Learners prioritize cognitive resources for conceptualizing complex messages and maintaining basic fluency, sacrificing lexical diversity due to the extra effort involved in retrieving varied vocabulary.

Furthermore, task sequence notably influenced performance. Group 1, progressing from simpler to more complex tasks, improved fluency, productivity, and accuracy, consistent with Robinson's SSARC principle of gradually increasing task complexity. Group 2, initially facing high complexity, displayed fluency improvements once task difficulty decreased, but accuracy gains did not recover fully, indicating sequencing can aid fluency recovery but is less effective for accuracy repair. Group 3, encountering an ascending-then-descending complexity sequence, showed variable outcomes with minor fluency gains offset by later reductions in syntax and productivity, reflecting prior observations by Baralt[38] and Malicka[40].

Unlike Baralt's[22] findings with interactive synchronous tasks, our monologic video-report tasks aligned clearly with CH predictions. The discrepancy stems primarily from task modality and interactional demands: Baralt's tasks dispersed attention through simultaneous reading, typing, and negotiating meaning, whereas our monologic, scripted narratives allowed learners to focus attention specifically on reasoning processes. Thus, absolute cognitive complexity remains central to shaping spoken performance, though sequencing can modulate specific CALF dimensions, particularly fluency.

Regarding the connection between Task-level Effect and Task-order Effect, a comparative analysis reveals how task complexity and sequencing jointly influence CALF dimensions. Group 1 effectively managed the increasing complexity, enhancing syntax, accuracy, and fluency, but experienced a decline in lexical variety during the most demanding final task, likely due to cognitive fatigue. Conversely, Group 2's immediate exposure to high complexity reduced fluency and syntax; despite subsequent easier tasks boosting fluency, initial cognitive strain prevented gains in accuracy and vocabulary. Group 3 showed intermediate outcomes: moderate complexity occasionally improved syntax or accuracy, but inconsistent fluency and productivity indicated sensitivity to fluctuations in task demands. Overall, lexical variety was most susceptible under peak difficulty or fatigue; syntactic improvements relied on available cognitive resources, fluency primarily responded to task sequence, and accuracy benefited only when cognitive demands and sequencing aligned favorably.

## 5. Conclusions

The results emphasize that task design, particularly cognitive complexity and sequencing, significantly shapes oral performance in technology-mediated TBLT. Moderate cognitive demands consistently yielded optimal syntactic length and accuracy without compromising fluency. For example, Group 1 reached peak productivity at intermediate complexity, highlighting the importance of moderately challenging, well-scaffolded tasks.

Task sequencing also influenced learner outcomes. A progressive sequence (simple to complex) allowed some learners to steadily build fluency, productivity, and accuracy. Conversely, initiating with more complex tasks enabled others to subsequently boost fluency and clause length when demands eased. Thus, instructors might strategically alternate between tasks that stretch learners' abilities and those that consolidate their gains, optimizing overall language development.

We acknowledge that increasing task complexity inevitably raises short-term memory load. However, expert ratings confirmed that the tasks primarily stimulated reasoning processes aligned with Robinson's resource-directing category.

Despite these insights, several limitations constrain the study's generalizability. The small, single-site sample, exclusive reliance on monologic computer-mediated tasks, limited complexity steps, and overlapping effects of complexity and sequence complicate causal attribution. Additionally, the absence of a neutral baseline task makes it challenging to distinguish true sequence effects from practice or fatigue effects. Finally, the analysis focused solely on traditional CALF metrics, overlooking direct measures of communicative effectiveness.

Future research should address these limitations by broadening participant pools, embedding tasks in interactive classroom contexts, introducing finer-grained complexity increments; using neutral baseline tasks; tracking learners longitudinally; applying advanced statistical models; and incorporating measures of functional adequacy for a more comprehensive assessment.

## Author Contributions

Author Contributions: Conceptualization, W.Z.; methodology, W.Z. and J.B.; software, W.Z.; validation, W.Z.; formal analysis, W.Z.; investigation, W.Z.; data curation, W.Z.; writing—original draft preparation, W.Z.; writing—review and editing, W.Z. and J.B.; supervision, J.B.; project administration, W.Z. and J.B. All authors have read and agreed to the published version of the manuscript.

## Funding

This work received no external funding.

## Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the School of Culture, Tourism and International Communication, Yunnan Open University (9 May 2024).

## Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

## Data Availability Statement

Data Availability Statement: The raw datasets generated and analyzed during the current study are available from the corresponding author on reasonable request at Google Drive via the following link: https://drive.google.com/drive/folders/1W957MfDfctG7fBrjySLTukHA4D_Il_7x

## Acknowledgments

In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Acknowledgments: The authors gratefully acknowledge the insightful guidance and continued support of Professor Justin James Bartlett throughout all stages of this research. We also extend our sincere thanks to the dedicated student volunteers at Yunnan Open University who assisted in organizing participant schedules and collecting data for the experiment.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

[1] González-Lloret, M., Ortega, L., 2014. Technology-mediated TBLT: Researching Technology and Tasks (Vol. 6). John Benjamins: Amsterdam, Netherlands.
[2] Thornbury, S., 2000. How to Teach Grammar, 1st ed. Longman: Harlow, UK.
[3] Nita, S.A., Syafei, A.F.R., 2012. Involving Audio-Lingual Method (ALM) and Communicative Language Teaching (CLT) in teaching speaking skill at junior high school. Journal of English Language Teaching. 1(1), 65–73.
[4] Chang, S.C., 2011. A contrastive study of grammar-translation method and communicative approach in teaching English grammar. English Language Teaching. 4(2), 13–18. DOI: https://doi.org/10.5539/elt.v4n2p13
[5] Gattegno, C., 1977. English — The Silent Way [videotape series]. Silent Way Video Company: New York, NY, USA.
[6] Hymes, D.H., 1971. Pidginization and Creolization of Languages (Ed. D.H. Hymes). Cambridge University Press: Cambridge, UK.
[7] Spada, N., 2007. Communicative Language Teaching: Current Status and Future Prospects. In: Cummins, J., Davison, C. (eds.). International Handbook of English Language Teaching. pp. 271–288. Springer: New York, USA.
[8] Norris, J.M., Ortega, L., 2000. Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. Language Learning. 50, 417–528.
[9] Nunan, D., 2003. The impact of English as a global language on educational policies and practices in the Asia-Pacific region. TESOL Quarterly. 37(4), 589–613.
[10] Van den Branden, K., Bygate, M., Norris, J.M., 2009. Task-Based Language Teaching: A Reader. John Benjamins: Amsterdam, Netherlands.
[11] Richards, J.C., Rodgers, T.S., 2014. Approaches and Methods in Language Teaching, 3rd ed. Cambridge University Press: Cambridge, UK.
[12] Samuda, V., Bygate, M., 2008. Task Research from a Pedagogical Perspective. In Bygate, M., Skehan, P., Swain, M. (eds.). Tasks in Second Language Learning. Longman: Harlow, UK. pp. 133–191.
[13] Gilabert, R., Manchón, R., Vasylets, O., 2016. Mode in theoretical and empirical TBLT research: Advancing research agendas. Annual Review of Applied Linguistics. 36, 117–135.
[14] Angelini, M.L., García-Carbonell, A., 2019. Developing English speaking skills through simulation-based instruction. Teaching English with Technology. 19(2), 3–20.
[15] Ganta, T.G., 2015. The strengths and weaknesses of task-based learning (TBL) approach. Scholarly Research Journal for Interdisciplinary Studies. 3(16), 2760–2771.
[16] Eslami, Z.R., Kung, W.T., 2016. Focus-on-form and EFL learners' language development in synchronous computer-mediated communication: Task-based interactions. The Language Learning Journal. 44(4), 401–417.
[17] Calderon, O., Sood, C., 2020. Evaluating learning outcomes of an asynchronous online discussion assignment: A post-priori content analysis. Interactive Learning Environments. 28(1), 3–17.
[18] Smith, B., González-Lloret, M., 2021. Technology-mediated task-based language teaching: A research agenda. Language Teaching. 54(4), 518–534.
[19] Nik, N., 2010. Examining the Language Learning Potential of a Task-Based Approach to Synchronous Computer-Mediated Communication [Unpublished doctoral dissertation]. Victoria University of Wellington: Wellington, New Zealand.
[20] Nik, N., Adams, R., Newton, J., 2012. Writing to learn via text chat: Task implementation and focus on form. Journal of Second Language Writing. 21, 23–39.
[21] Adams, R., Nik, N., 2014. Prior Knowledge and Second-Language Task Production in Text Chat. In González-Lloret, M., Ortega, L. (eds.). Technology-Mediated TBLT: Researching Technology and Tasks. John Benjamins: Amsterdam, Netherlands. pp. 51–78.
[22] Baralt, M., 2013. The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. Studies in Second Language Acquisition. 35(4), 689–725.
[23] Heift, T., Schulze, M., 2007. Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues. Routledge: London, UK.
[24] Peterson, M., 2011. Towards a research agenda for the use of three-dimensional virtual worlds in language learning. CALICO Journal. 29(1), 67–80.
[25] Robinson, P., 2011. Task-based language learning: A review of issues. Language Learning. 61, 1–36.
[26] Robinson, P., 2001. Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In Robinson, P. (ed.). Cognition and Second Language Instruction. Cambridge University Press: Cambridge, UK. pp.

287–318.
[27] Skehan, P., 1998. A cognitive Approach to Language Learning. Oxford University Press: Oxford, UK.
[28] Skehan, P., 1996. A framework for the implementation of task-based instruction. Applied Linguistics. 17(1), 38–62.
[29] Robinson, P., 2005. Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. International Review of Applied Linguistics. 43, 1–32.
[30] Robinson, P., 2007. Criteria for classifying and sequencing pedagogic tasks. In García Mayo, M.P. (ed.). Investigating Tasks in Formal Language Learning. Multilingual Matters: Clevedon, UK. pp. 7–26.
[31] Albert, A., 2011. When individual differences come into play: The effect of learner creativity on simple and complex task performance. In Robinson, P. (ed.). Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance. John Benjamins: Amsterdam, Netherlands. pp. 239–226.
[32] Sasayama, S., Izumi, S., 2012. Effects of task complexity and pre-task planning on EFL learners' oral production. In Shehadeh, A., Coombe, C. (eds.). Task-Based Language Teaching in Foreign Language Contexts. John Benjamins: Amsterdam, Netherlands. pp. 23–42.
[33] Donate, Á., 2018. Cognitive Task Complexity, Foreign Language Anxiety and L2 Performance in Spanish: A TBLT perspective [Doctoral dissertation]. Georgetown University: Washington, DC, USA.
[34] Xu, T.S., Zhang, L.J., Gaffney, J.S., 2023. A multidimensional approach to assessing the effects of task complexity on L2 students' argumentative writing. Assessing Writing. 55, 100690.
[35] Michel, M.C., 2011. Effects of task complexity and interaction on L2 performance. In Robinson, P. (ed.). Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance, Vol. 2. John Benjamins: Amsterdam, Netherlands. pp. 141–174. DOI: https://doi.org/10.1075/tblt.2.12ch6
[36] Robinson, P., 2010. Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In Pütz, M., Sicola, L. (eds.). Cognitive Processing in Second Language Acquisition. John Benjamins: Amsterdam, Netherlands. pp. 243–268.
[37] Robinson, P., 2022. The Cognition Hypothesis, the Triadic Componential Framework and the SSARC model: An instructional design theory of pedagogic task sequencing. In the Cambridge Handbook of Task-Based Language Teaching. Cambridge University Press: Cambridge, UK. pp. 205–225.
[38] Baralt, M., 2014. Task complexity and task sequencing in traditional versus online language classes. In Baralt, M., Gilabert, R., Robinson, P. (eds.). Task Sequencing and Instructed Second Language Learning. Bloomsbury: London, UK. pp. 95–122.
[39] Levkina, M., Gilabert, R., 2014. Task Sequencing in the L2 Development of Spatial Expressions. In Task Sequencing and Instructed Second Language Learning. Bloomsbury: London, UK. pp. 37–70.
[40] Malicka, A., 2020. The role of task sequencing in fluency, accuracy, and complexity: Investigating the SSARC model. Language Teaching Research. 24(5), 642–665.
[41] Robinson, P., 2015. The Cognition Hypothesis, second-language task demands, and the SSARC model of pedagogic task sequencing. In Bygate, M. (ed.). Domains and Directions in the Development of TBLT. John Benjamins: Amsterdam, Netherlands. pp. 87–122.
[42] Pan, S.C., Rickard, T.C., 2018. Transfer of test-enhanced learning: Meta-analytic review and synthesis. Psychological Bulletin. 144(7), 710–737.
[43] Kormos, J., 2011. Speech Production and the Cognition Hypothesis. In Robinson, P. (ed.). Second Language Task Complexity. John Benjamins: Amsterdam, Netherlands. pp. 39–60.
[44] Sasayama, S., 2016. Is a "complex" task really complex? Validating the assumption of cognitive task complexity. The Modern Language Journal. 100(1), 231–254.
[45] Skehan, P., 1989. Language testing part II. Language Teaching. 22(1), 1–13.
[46] Housen, A., Kuiken, F., 2009. Complexity, accuracy and fluency in second language acquisition. Applied Linguistics. 30(4), 461–473.
[47] Lakens, D., 2022. Sample size justification. Collabra: Psychology. 8(1), Article 33267.
[48] Richards, J.C., Schmidt, R.W., 2013. Longman dictionary of language teaching and applied linguistics, 4th ed. Routledge: London, UK.
[49] Rahimi, M., Zhang, L.J., 2019. Writing task complexity, students' motivational beliefs, anxiety and their writing production in English as a second language. Reading and Writing. 32(3), 761–786.
[50] Frear, M.W., Bitchener, J., 2015. The effects of cognitive task complexity on writing complexity. Journal of Second Language Writing. 30, 45–57. DOI: https://doi.org/10.1016/j.jslw.2015.08.005