COGNITIVE TASK COMPLEXITY, EMOTION AND L2 PERFORMANCE IN ENGLISH: A STUDY FROM TMTBLT PERSPECTIVE

WANYI ZHANG

Graduate School Srinakharinwirot University

2024

-

WANYI ZHANG

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปรัชญาดุษฎีบัณฑิต สาขาวิชาภาษาและการสื่อสารสากล
คณะมนุษยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2567

COGNITIVE TASK COMPLEXITY, EMOTION AND L2 PERFORMANCE IN ENGLISH: A STUDY FROM TMTBLT PERSPECTIVE

WANYI ZHANG

An Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of DOCTOR OF PHILOSOPHY

(Ph.D. (Language and Global Communication))

Faculty of Humanities, Srinakharinwirot University

2024

THE DISSERTATION TITLED

COGNITIVE TASK COMPLEXITY, EMOTION AND L2 PERFORMANCE IN ENGLISH: A STUDY FROM TMTBLT PERSPECTIVE

BY

WANYI ZHANG

HAS BEEN APPROVED BY THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DOCTOR OF PHILOSOPHY
IN PH.D. (LANGUAGE AND GLOBAL COMMUNICATION) AT SRINAKHARINWIROT UNIVERSITY

(Assoc. Prof. Dr. Chatchai Ekpanyaskul, MD.)

Dean of Graduate School

ORAL DEFENSE COMMITTEE

.... Major-advisor
(Lecturer Dr.Justin James Bartlett)

.... Chair
(Assoc. Prof. Dr.Nattama Pongpairoj)

.... Committee
(Assoc. Prof. Dr.Sugunya Ruangjaroon)

.... Committee
(Assoc. Prof. Dr.Nuntana Wongthai)

.... Committee
(Asst. Prof. Dr.Sakulrat Worathumrong)

Title	COGNITIVE TASK COMPLEXITY, EMOTION AND L2 PERFORMANCE IN ENGLISH: A STUDY FROM TMTBLT PERSPECTIVE
Author	WANYI ZHANG
Degree	DOCTOR OF PHILOSOPHY
Academic Year	2024
Thesis Advisor	Lecturer Doctor Justin James Bartlett

This study investigated the effects of cognitive task complexity and task sequencing in Technology-Mediated Task-Based Language Teaching (TMTBLT) on learners' emotions, specifically foreign language fear (FLF) and foreign language enjoyment (FLE), and their subsequent influence on L2 English speaking performance. Sixty Chinese undergraduate students participated in three computer-mediated oral reporting tasks of varying complexity levels (low, medium, and high), arranged in counter-balanced sequences. Spoken performance was evaluatedusing complexity, accuracy, lexical complexity, and fluency (CALF) measures. Additionally, participants' emotional responses were assessed immediately after task completion. Results identified an optimal complexity zone, wherein medium complexity tasks produced the highest lexical diversity and syntactic complexity. Conversely, high-complexity tasks negatively impacted fluency and increased levels of FLF. Furthermore, a task sequence progressing from simple to complex, consistent with Robinson's SSARC model, was more effective in preserving accuracy and fluency compared to reverse or mixed sequencing. Positive correlations were observed between FLE and both lexical and syntactic complexity, whereas FLF negatively correlated with fluency. These findings empirically distinguish FLF from traditional anxiety constructs in language learning contexts. Overall, this research refines the Cognition Hypothesis, underscores the pedagogical importance of sequencing tasks from simple to complex, and emphasizes the necessity of designing technology-mediated tasks that account for emotional factors in digital language education.

Keyword : cognitive task complexity, task sequencing, technology-mediated TBLT, foreign language fear, foreign language enjoyment, CALF, second language acquisition

ACKNOWLEDGEMENTS

I extend my sincere gratitude to everyone who contributed academically and personally to this dissertation.

Foremost, I am deeply indebted to my advisor, Ajarn Justin James Bartlett, whose invaluable guidance shaped this research from its earliest stages. His support, both academic and personal, greatly enriched my doctoral experience.

I also thank my dissertation committee members, Ajarn Sugunya Ruangjaroon and Ajarn Nuntana Wongthai. Their insightful feedback significantly strengthened this research, particularly by suggesting the incorporation of Foreign Language Enjoyment (FLE) and Robinson's SSARC model. I am additionally grateful for their practical guidance and genuine concern for my well-being.

My appreciation extends to my classmates and colleagues for their insightful comments and moral support during our academic seminars and discussions, making this journey more rewarding.

Finally, I warmly acknowledge my family and close friends for their unconditional emotional support, understanding, and unwavering faith in me throughout these doctoral years.

Thank you all sincerely for your indispensable guidance, encouragement, and kindness.

WANYI ZHANG

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background

As an English teacher, it is natural for me to observe each of my students and keep seeking an effective way to support them improving their language performance. Once in a normal oral English class, I randomly asked a boy to read aloud a conversation in front of his classmates. He looked nervous and uncomfortable, reading out with tremble and low voice: *What 'Timi' is it?* With no doubt, he made a mistake. The 'Time' was incorrectly read as 'Timi'. Later, I realized the reason for making this mistake is because there was a popular mobile game among young group, and every time the game is started, the logo of the game company will be showed and heard, that is what the boy said 'Timi' (rhymes with 'Timmy'). The whole class burst into laugh loudly after his incorrect reading. Although he felt uncomfortable when he was standing as the picked one, he breathed a sigh of relief and recovered as usual when I said "please sit down" to him. Something seems to appear and then disappear during this period, and it absolutely affected the student's behaviors and performance. Therefore, based on this impressive experience, I have several considerations as follow.

First, the behaviors and performance of the students went to an abnormal way when he was picked: The deviant behaviors refer to his hesitation, trembling, higher heart rate, slow speaking speed and low accuracy. Here follows several questions: What kind of emotion did he produce? Did he feel anxious or fear when he was picked? Why did he show those specific behaviors? Second, the pronunciation of the 'time' seems to be influenced by a mobile game he likes and changed into 'timi'. Does this mean the mobile-based methods and the game-based approaches are relatively attractive and effective for learners? The behaviors, the performance and the mistake seem to be connected with his emotion and the learning method. Among the above questions and guesses, I decided to make the current study and try to find answers.

Moreover, the English self-learning Apps started to be popular few years ago. Students tried to do self-learning in order to mend the limitation of in-class language learning. However, the effectiveness of various Apps are of varying quality. Are task sequence, complexity and feedback of the Apps logical and effective? If not, what is the best range of the task complexity? Can the mobile-based language learning really decrease learners' negative emotion and boost their positive emotion?

Notably, the scenario described above is not unique to my classroom—it reflects a broader phenomenon in language education. Around the world, many foreign language learners suffer from debilitating classroom emotions such as anxiety or fear, which can hinder their participation and performance. Researchers have long observed that negative emotions (often labeled foreign language anxiety) correlate with poorer language outcomes and can even discourage learners from speaking up (Horwitz et al., 1986; Horwitz, 2017). At the same time, today's educational landscape is increasingly digital. Educators are grappling with how to integrate technology into language teaching in ways that enhance learning rather than exacerbate stress. Technology-mediated language learning can, when grounded in sound methodology, be as effective as face-to-face instruction, and it offers new opportunities to engage learners through games, apps, and online tasks. The question is how to harness these tools to reduce negative emotions and improve performance. In a globalized era that demands communicative competence, finding this balance is crucial. Thus, this study addresses a timely challenge in modern education: optimizing task design (complexity and sequencing) and emotional support in technology-mediated language learning to improve learner outcomes. In the next section, an overview of the study will be presented.

1.2 Overview of the study

In order to address the questions outlined in the previous section, I chose to situate all elements in the context of technology-mediated task-based language teaching (TMTBLT), an emerging approach for language acquisition (Gonzales-Lloret & Ortega, 2014). Within TMTBLT, various aspects remain to be examined—i.e., task design, task sequence, and the method's effectiveness for learner performance. For task sequence, the specific order in which tasks are performed can produce distinct outcomes in L2 performance, as learners may adapt over time or become fatigued under successive tasks. Meanwhile, regarding task design, cognitive task complexity is a key factor referring to the inherent difficulty of the task itself, which can either diminish or enhance learners' language performance given the limitations of their cognitive load. In this study, set within TMTBLT, I explore how cognitive task complexity, task sequence, foreign language fear, foreign language enjoyment, and learners' speaking performance—assessed through complexity (syntactic and lexical), accuracy, and fluency—interconnect.

As mentioned, the current study will implement tasks under the context of TMTBLT (Gonzales-Lloret & Ortega, 2014). TMTBLT was developed from the traditional TBLT, which initially originates from the Communicative Language Teaching (CLT) approach (Habermas, 1970; Hymes, 1971). The developing progress and relative histories would be presented in detail in Chapter 2. On the other hand, based on Robinson's Cognitive Hypothesis (CH) (Robinson, 2001a) and Skehan's Limited Attentional Capacity (LAC) (Skehan, 1998), these two frameworks lead to different ideas among the relationship between cognitive task complexity and learners' performance. In short, LAC said that because human's attention (cognitive load) is limited, the higher cognitive complexity will force learners to manage extra works and they will, therefore, not have enough cognitive capacity to dispose language performance, and this will lead to worse performance. However, CH supposed that learners' performance will be separately affected by higher level of cognitive complexity: the complexity, the accuracy and fluency will increase.

Further, the current study will explore the relationship between cognitive task complexity and performance based on Robinson's Triadic Componential Framework (TCF) (Robinson, 2005). This framework systematically provides three main dimensions of elements that might influence final performance in TBLT, i.e. cognitive task complexity, task difficulty, and task conditions. They respectively correspond to the level of the task requirement for learners' cognitive ability, the interactive factors and learners' related factors. These items would be explained in detail in Chapter 2. Specifically, in the current study, the way of manipulating the cognitive task complexity will refer to the TCF in the dimensions from resource-directing variables, particularly, +/- few elements.

Thus, the above sections state: 1) the study would be implemented in the context of TMTBLT; 2) the two main hypotheses (CH and LAC) for the connection between cognitive task complexity and language performance; 3) the specific method of manipulating cognitive task complexity (+/— elements in the TCF); and 4) the notion that task sequence may also shape learners' L2 performance by influencing how they approach successive tasks. However, these aspects alone cannot fully explain the teaching experience mentioned at the beginning of this paper—namely, what kind of emotion the picked student produced, whether he felt anxious or fearful when chosen to speak, and why he exhibited those specific behaviors. Consequently, learners' emotions, feelings, and behaviors also play an important role in the present study. In particular, I argue that the negative emotion displayed by the picked student is "fear" rather than "anxiety", as commonly labeled in previous research. This argument is presented in detail in Chapter 2. In the next section, the emotion-related aspect of this study will be introduced.

Specifically, the current study will concern the relationship between emotion, performance and complexity. In detail, the emotion refers to both positive and negative emotion: enjoyment and fear. However, previous related studies tend to confuse foreign language *anxiety* (FLA) (Horwitz et al., 1986;) and foreign language *fear (FLF)*, which will be criticized by the current study. The current study will argue that foreign language fear as an independent and distinct concept from the perspective of the general psychological theory of fear and anxiety (APA Dictionary of Psychology, n.d.), the Three-Dimensional Taxonomy of Achievement Emotions (Pekrun, 2007), and the three modes of defense concept in neurobehavioral science (Timberlake & Lucas, 2019). Moreover, the present study also provides evidence and examples for the argument that foreign language fear could appear during the ongoing process of language learning.

In conclusion, this study will rest on a *multi-faceted framework* that includes emotion, task complexity, task sequence, and L2 performance. Here, emotion refers specifically to FLE (Foreign Language Enjoyment) and FLF (Foreign Language Fear); task complexity addresses the cognitive demands placed on learners; task sequence highlights how the order of tasks may further shape performance; and L2 performance is measured through participants' complexity, accuracy, and fluency (CALF).

In the next chapter, all relevant theories, frameworks, and prior research will be laid out, along with the theoretical research questions guiding this study, offering a comprehensive foundation for the forthcoming empirical investigation.

1.3 Significance and Motivation

In general, the current study focuses on two main variables: cognitive task complexity and learner emotion. These two elements have been debated in the literature without an agreed conclusion, and they lie at the heart of the observed classroom challenges. The importance of this investigation extends beyond one classroom. Language educators worldwide are grappling with how task demands and emotions interact, especially as technology becomes more prevalent in education. Many L2 learners experience intense affective responses—often termed anxiety—that can significantly impede their classroom performance. Finding ways to reduce such negative emotions (or reinterpreting them properly as fear) is increasingly seen as important for fostering better learning environments. Meanwhile, digital learning tools and tasks are being adopted on a large scale, prompting questions about how to use them to enhance engagement and enjoyment rather than induce stress (Ortega & González-Lloret, 2014). These broader educational trends underscore the significance of the present study's focus.

Furthermore, with respect to the emotion variable, the current study will argue that what previous studies called "foreign language anxiety" is not actually anxiety, but a form of fear. This distinction will be substantiated through theoretical and empirical evidence (as outlined in Chapter 2). The motivation for clarifying this concept is not only academic but also practical: mislabeling fear as anxiety (or vice versa) could lead teachers to adopt less effective strategies for helping students. In sum, the study is driven by both the need to resolve theoretical debates in the field and the need to address real-world educational challenges in today's technology-enhanced, emotion-aware language classrooms.

1.3.1 The Cognitive Task Complexity and The Performance

Among TBLT, the two mainstream hypotheses for the relationship between cognitive task complexity and performance are Robinson's Cognitive Hypothesis (CH) (Robinson, 2001a) and Skehan's Limited attentional Capacity (LAC) (Skehan, 1998). LAC believes that within human's limited cognitive load, the higher degree of cognitive task complexity would lead to worse language performance, while the CH brings a relatively systematic framework and argues that different dimensions of the performance would be positively and negatively affected when the level of cognitive complexity is increased. Plenty of studies attempted to explore this relationship, however, results are various. The reason might be, that throughout the range of variation, there might be a point, where the performance starts to increase, and then reaches the peak, and then starts to decrease as cognitive complexity increases. To avoid this potential problem, in this study the performance would be measured under three levels of cognitive complexity: Low Complexity, Middle, and High Complexity. In Chapter 3, Methodology, the concrete experiment design would be illustrated. In brief, first, the current question as to which theory – LAC or CH – is correct has not been answered. Also, TMTBLT is a relatively new pedagogy with more possible and potential to be explored. More information would be given in Chapter 2 and Chapter 3.

1.3.2 The Cognitive Task Complexity and Emotion

Apart from the performance, the current study tends to explore learners' emotional changes while acquiring a foreign language. Scholars' attention used to be paid to negative emotions. People tried to understand the effects on language learning and performance brought by negative emotion, and both pros and cons were assumed to be produced by negative emotion, especially anxiety, in foreign language classrooms. However, whether the emotion named 'anxiety' in these studies is really anxiety is debatable, both from the standard psychological definition and from behavioral perspectives. I will argue that anxiety has been conflated with fear in the SLA literature. This argument will be completely demonstrated in Chapter 2.

On the other hand, although positive emotions begin to be explored more in recent years, there are still few studies which relate them to: first, the enjoyment emotion in the context of TMTBLT, and second, the possible connection between cognitive complexity and enjoyment. Obviously, for some learners, they might feel unhappy when facing higher degrees of task complexity, whereas, for some others, they would enjoy the complex tasks more as they prefer to challenge. Therefore, this phenomenon is worth exploring deeper. Furthermore, it might not be logical to say that the fear and enjoyment are opposite to each other. Actually, they can exist at the same time and in the same situation. Through the discussion and results of the present study, the answer of the above hypothesis from the experiment in the current study.

In conclusion, in the empirical study section, this work will provide evidence to support CH or LAC, or both, in the context of TMTBLT, and also analyze the psychological aspects of both foreign language fear and foreign language enjoyment produced under the condition of three levels of cognitive complexity. In addition, in the theoretical study section, a barrier will be built between FLA and FLF, to clearly distinguish these two concepts.

1.4 Aims and Research Questions

The aim of this thesis is to understand the relationship between the variables of task complexity, emotion, language performance among university-level second language English learners.

Specifically, the study aims to:

Major Aim (Empirical)

This study aims to describe and measure how a computer-mediated language learning task's cognitive complexity and task sequence influence fear, enjoyment, and language performance, by exploring the following questions:

For Task Complexity

Q1. In what ways (increase/decrease/none), and to what extent (greater/lesser), does task complexity affect fear-in in foreign language learning among the sample?

Q2. In what ways (increase/decrease/none), and to what extent (greater/lesser), does task complexity affect enjoyment in foreign language learning among the sample?

Q3. In what ways (improve/not improve), and to what extent (greater/lesser), does task complexity affect foreign language performance (L2 speaking) among the sample?

For Task Sequence

Q4. In what ways (increase/decrease/none), and to what extent (greater/lesser), does task sequence affect fear-in in foreign language learning among the sample?

Q5. In what ways (increase/decrease/none), and to what extent (greater/lesser), does task sequence affect enjoyment in foreign language learning among the sample?

Q6. In what ways (improve/not improve), and to what extent (greater/lesser), does task sequence affect foreign language performance (L2 speaking) among the sample?

(*Independent variables*: cognitive complexity and task sequence. *Dependent variables*: fear (F), enjoyment (E), language competence (LC), and language performance (LP). *Controlled variables*: duration, age, English competency, setting.)

Minor Aim (Theoretical):

To argue for a distinction between anxiety and fear, and for a further distinction between fear-in and fear-of foreign language. It will be argued that this distinction is important for language teachers so that they can treat students appropriately.

Question 1: Is foreign language anxiety fear or anxiety?

Question 2: How does anxiety differ from fear?

Question 3: How does fear-in differ from fear-of?

Question 4: Why are these distinctions useful?

The theoretical part of aims in the thesis will be illustrated and undertaken in Chapter two, therefore in chapter three – for methodology – only the empirical research questions will be explored independently.

1.5 Glossary of Terminology and Definitions

Technology-Mediated Task-Based Language Teaching (TMTBLT)

An approach to language teaching that integrates technology with task-based instruction. Developed from traditional TBLT and the field of computer-assisted language learning (CALL), it represents a new and important research direction in applied linguistics (Ortega & González-Lloret, 2014). TMTBLT emphasizes that tasks and digital tools should be meaningfully integrated to facilitate L2 acquisition.

Robinson's Cognitive Hypothesis (CH) & Triadic Componential Framework (TCF)

Robinson's Cognition Hypothesis (Robinson, 2001b) predicts how learners' L2 performance changes under different levels of cognitive task complexity. Essentially, it posits that more complex tasks can lead to more complex and accurate language use (due to greater cognitive effort), though possibly less fluency. To detail task features, (Robinson, 2005) proposed the Triadic Componential Framework, a taxonomy for categorizing task characteristics along three dimensions: cognitive task complexity (features of the task itself, e.g. number of elements, reasoning demands), task conditions (interactional factors, e.g. one-way vs two-way tasks), and task difficulty (learner factors, e.g. proficiency, affective variables). The TCF is considered a workable model for systematically designing and comparing tasks in research and pedagogy.

Skehan's Limited Attentional Capacity (LAC) model

Skehan's LAC model (Skehan, 1998) explains performance limitations under complex tasks by noting that working memory resources are limited. Learners cannot pay attention to all aspects of language simultaneously with equal focus. Skehan enumerated three main factors influencing task difficulty: cognitive complexity, code complexity, and communicative stress. Under LAC, when cognitive demands increase, learners tend to prioritize certain aspects (often meaning and fluency) at the expense of others (accuracy or linguistic complexity). This leads to trade-offs in performance. The LAC model implies that task designers must consider attentional limits, balancing tasks so that learners are not overwhelmed.

Cognitive Task Complexity

This refers to the inherent mental challenge posed by a task's design. (Robinson, 2001b) defines cognitive task complexity as "the inherent cognitive demands of the task imposed on the learners by the structure of the task." In Skehan's framework, task complexity is affected by cognitive factors (like reasoning demands), linguistic factors (code complexity such as vocabulary and syntax difficulty), and communicative factors (time pressure, modality, etc.). In Robinson's TCF, cognitive task complexity refers specifically to features of the task itself and is divided into resource-directing variables (elements that direct learners' attention to language form or meaning, e.g. +/- few elements, here-and-now vs there-and-then) and resource-dispersing variables (elements that disperse attentional resources, e.g. planning time, prior knowledge). In summary, a "complex" task is one that imposes greater cognitive processing load on the learner (Ellis, 2003; Robinson, 2001a).

Complexity of L2 Performance (CALF framework)

In second language performance research, Complexity, Accuracy, and Fluency (often abbreviated as CAF or CALF when including Lexical complexity) are key measures. Complexity of performance describes how varied and sophisticated the language produced by the learner is. This includes the range of vocabulary used and the use of advanced structures (e.g. subordinate clauses, complex sentence structures). Higher complexity indicates a more elaborate interlanguage system. In broad terms, complexity encompasses the extent, richness, and diversity of L2 performance. For example, a learner who uses a wide variety of verbs and complex sentences is demonstrating high complexity. (Housen & Kuiken, 2009) Accuracy refers to how error-free the language is, and fluency refers to the ease and rate of speech, but these are defined in detail later chapters. Complexity in the CALF framework is crucial for assessing whether increased task demands lead learners to expand their language use or simplify it.

Robinson's SSARC Model (Simplify, Stabilize/Automatize, Restructure, Complexify)

The SSARC model (Robinson, 2010) provides guidance for sequencing tasks in pedagogic contexts. It suggests that instructors should begin with simpler versions of a task and then gradually increase complexity in stages, allowing learners to first Simplify and master the task, then Stabilize/Automatize their performance through practice, Restructure their interlanguage as needed, and finally Complexify the task to approach real-world task demands. In essence, the SSARC model is an implementation of the Cognition Hypothesis: by following a principled sequence from easy to hard, learners can build accuracy and fluency on simpler tasks and then transfer those skills to more complex tasks without being overwhelmed. This model has been considered a theoretically motivated and pedagogically feasible approach to task sequencing in TBLT (Robinson, 2001a).

Foreign Language Anxiety (FLA) & Horwitz's Foreign Language Classroom Anxiety Scale

Foreign language anxiety is a well-known negative emotion specific to language learning contexts, first systematically described by Horwitz, Horwitz, and Cope (1986). Horwitz et al. (1986) defined FLA as a distinct complex of self-perceptions, beliefs, and feelings related to classroom language learning arising from the uniqueness of the language learning process. To measure this construct, Horwitz and colleagues developed the Foreign Language Classroom Anxiety Scale (FLCAS), which has become the most widely used instrument for assessing language anxiety in learners. FLA can manifest as nervousness, fear of speaking in front of others, or worry about making mistakes in a foreign language, and it has been linked to avoidance of participation and performance deficits in numerous studies (Horwitz et al., 1986).

Foreign Language Fear (FLF)

In this thesis, foreign language fear is proposed as an independent and distinct concept (separate from general language anxiety). FLF is defined as a negative emotion occurring during the process of foreign language activities (e.g. while speaking or performing a task) and dissipating once the activity ends. This emotion is typically triggered by real, immediate stimuli in the environment (for example, reactions of peers, fear of receiving a bad score, or a teacher's presence). FLF activates a set of physiological responses (such as a pounding heart, sweating, muscle tension) that correspond to a classic fight-or-flight reaction, motivating the learner to escape from or terminate the ongoing activity as quickly as possible. (This conceptualization draws on general psychology definitions of fear and aligns with fear responses described in neurobehavioral science; see APA Dictionary of Psychology, n.d., and Chapter 2 for detailed justification. Note: FLF is a novel term introduced by the author, so no prior citation is available for this exact concept.)

Foreign Language "Fear-in" vs. "Fear-of"

Within the broader FLF construct, we distinguish between fear-in and fear-of to pinpoint the source or focus of the fear. Fear-in refers to foreign language fear induced by factors external to the language content itself. These are non-language-learning-related triggers, such as a poor teacher-student relationship, a generally intimidating classroom atmosphere, or peer pressure unrelated to the language task. In contrast, fear-of refers to fear arising from language-related situations: for example, a learner who is motivated to learn the language but feels frightened when facing a pop quiz, a speaking test, or having to speak spontaneously in class. In short, fear-in comes from the surrounding context or environment, whereas fear-of comes from the language performance situation itself. (These terms fear-in and fear-of are newly coined in this study to differentiate sources of fear; they will be elaborated in Chapter 2. No external literature has used these exact terms before.)

Pekrun's Three-Dimensional Taxonomy of Achievement Emotions

Pekrun (2006, 2007) proposed a taxonomy to classify achievement emotions (emotions linked to academic activities and outcomes) along three dimensions: valence (positive vs. negative emotion), activation level (activating vs. deactivating physiological arousal), and object focus (activity-related vs. outcome-related vs. retrospective emotions). For example, enjoyment in doing a task is a positive, activating, activity-focused emotion; fear of failing an exam is a negative, activating, prospective-outcome emotion. This three-dimensional framework (sometimes referred to as part of Pekrun's Control-Value Theory) helps in understanding where an emotion like foreign language enjoyment or fear falls in terms of its nature and timing relative to learning. Pekrun's taxonomy provides a theoretical basis for why fear and anxiety in the classroom should be seen as different emotions: they may have different valence or appraisals (e.g. control vs. lack of control).

Predatory Imminence Theory (Timberlake & Lucas, 2019)

Over the course of evolution, all animals (including humans) have developed adaptive defensive reactions to threats (Bolles, 1970). (Timberlake & Lucas, 2019) advanced the Predatory Imminence Theory, which describes how an organism's defensive behavior changes as a threat becomes more immediate. They identified three stages or modes of defense: pre-encounter (when no threat is detected, but the organism is in a precautionary state), post-encounter (when a threat is detected and the organism becomes alert and cautious), and circa-strike (when a threat is imminent or attacking, triggering fight-or-flight responses). This theory is relevant to foreign language fear in that a learner's sense of threat (e.g. anticipating being called on in class) can escalate through similar stages — from a mild worry before being called (pre-encounter), to heightened anxiety when the teacher asks a question (post-encounter), to an intense fear reaction when actually having to speak (circa-strike). By drawing this parallel, we can better understand the physiological and behavioral manifestations of fear in language learning situations.

Foreign Language Enjoyment (FLE)

Foreign Language Enjoyment is a positive emotion experienced in the context of learning or using a new language. (MacIntyre, 2016) described FLE as the feeling when learners actively embrace language learning challenges and take pleasure in expanding their proficiency and knowledge. In other words, it's the joy that comes from successfully communicating in the L2, mastering a new element of the language, or engaging in a fun language activity. (Dewaele & MacIntyre, 2016) define FLE as "a complex emotion, capturing interacting dimensions of challenge and perceived success in using the language." This means that learners feel enjoyment partly because the task is challenging but achievable, and they gain a sense of accomplishment from it. FLE has gained attention with the rise of Positive Psychology in SLA, as studies show that higher enjoyment is linked to higher motivation, more willingness to communicate, and better overall achievement in language learning.

1.6 Thesis Structure

In this thesis, the relevance context, theories, frameworks and previous studies will be included in Chapter 2, literature review. This consists of the introduction and development of traditional task-based language teaching to later technology-mediated language teaching. The two main streams of debating the relation between cognitive task complexity and language performance among years: Robinson's Cognitive Hypothesis (CH) and Skehan's Limited Attentional Capacity (LAC). I will survey and present several literature related to CH and LAC and I will then, based on a survey of the literature related to emotion in SLA, build an argument to support a new concept of FLF and a distinct emotion to FLA.

With respect to emotion, both negative emotion (anxiety and fear) and positive emotion (enjoyment) will be discussed. However, the current study argues that the foreign language fear should be distinguish from the foreign language anxiety, which differs from extant studies , as they tend to conflate or mix these two concepts together. On the other hand, With the growth of Positive Psychology (PP), it is anticipated to be vital to pay attention to learners' positive emotions produced during and after language learning. Therefore, the measuring scale for foreign language fear and foreign language enjoyment will be adopted to check participants' emotion changing with the immersion of tasks under different dimensions of manipulation. In brief, in the current study, under the general context of TMTBLT, we would manipulate the cognitive task complexity from the dimension of +/- elements to observe participants' performance from the perspective of lexical complexity, accuracy and fluency, and also their level of both FLF and FLE.

In Chapter 3, the methodology related information will be presented. Generally it will consist of two main sections: the pilot study and the main study. For each of the study, the details of participants, the materials and tools that will be used, the procedures of actualizing the experiments and the methods of collecting and analyzing data will be illustrated. Further, in Chapter 4, the results of all the experiments will be showed, compared and discussed. All the mentioned research questions could be answered in this chapter. Finally, the Chapter 5 will be a conclusion for the thesis, including the further analyzing of the experiment results, the significance the results bring and the potential implementation of the study for the Second Language Acquisition field.

CHAPTER 2
LITERATURE REVIEW

As mentioned in the last chapter, this literature review covers the basic theories, frameworks, and related findings of previous studies. Furthermore, the theoretical research questions are discussed, including our perspective on the distinction between foreign language anxiety and foreign language fear, which will be examined from several angles.

2.1 Techniques for Second Language Acquisition

In this part, the evolution of techniques for SLA in the past decades will be illustrated. This includes traditional language pedagogy such as the Audio-Lingual Method (ALM), the Grammar-Translation Method (GTM), and the Silent Way (SW). As the imperfections of these classical methods became apparent, educators sought new approaches to avoid those drawbacks. The present study reviews several relatively new L2 teaching methods: Communicative Language Teaching (CLT), Task-Based Language Teaching (TBLT), and Technology-Mediated Task-Based Language Teaching (TMTBLT). These newer pedagogies will be introduced, including their definitions, historical development, and related applied research.

2.1.1 Traditional Techniques for SLA

One of the fundamental questions in the field of second language acquisition is how to effectively learn and teach a second language. Over the past decades, traditional language pedagogy such as (1) the Audio-Lingual Method (ALM), which was advanced in the 1950s by American linguists and views language as a form of behavior learned through the formation of correct habits (Thornbury, 2018), p.21), has been influential. In other words, the aim of ALM is to support language learners in forming native-like language habits (Dendrinos, 1992). ALM advises students to learn a foreign language without using their native language to translate new vocabulary or grammar in the target language. (2) the Grammar-Translation Method (GTM), which is a method guiding the teacher to present grammar structures and then practice these structures through speaking and/or writing exercises, after which students use these structures in less controlled speaking or writing activities (Chang, 2011). (3) the Silent Way (SW), proposed by (Gattegno, 1977), which suggests that teachers remain mostly silent and encourage students to be more actively involved in language learning. Concrete techniques in SW include using various materials and activities such as colored Cuisenaire rods, word color charts, Fidel charts, and pronunciation charts. This approach is rooted in the principle that knowledge is not a transferable product, but a structure that learners actively develop through experience. In addition to the above three methods, several other influential language teaching techniques have emerged over time, both historical and modern. For example:

Direct Method: Also known as the Natural Method, it immerses learners directly in the target language with no use of the L1. Teachers teach vocabulary and grammar through demonstration and context, emphasizing speaking and listening. This approach can improve oral proficiency and listening comprehension, but it often lacks explicit grammar instruction and may disadvantage learners who need explanations.

Total Physical Response (TPR): Developed by James Asher, TPR involves students responding with physical actions to teacher commands in the target language. It leverages the coordination of language and movement to reduce learner stress and reinforce memory. TPR is very effective for beginning learners and for teaching concrete vocabulary and commands, though it is less applicable for advanced language or abstract concepts.

Community Language Learning (CLL): Created by Charles Curran, CLL treats the language class as a supportive “community”. Learners sit in a circle; they express what they want to say in their native language and the teacher (as a counselor) helps translate it into the target language, which learners then repeat. This method lowers anxiety and builds a sense of community and cooperation, but it requires a skilled facilitator and can be time-consuming, with less focus on systematic structure practice.

Suggestopedia: A method developed by Georgi Lozanov, Suggestopedia uses a relaxed atmosphere, soothing music, and positive suggestion to help students absorb language. Learners often sit in comfortable chairs, listen to dialogues with background music, and engage in dramatized texts. This approach aims to tap into subconscious learning to dramatically increase vocabulary retention. While some learners find it highly motivating and it reduces anxiety, critics note that its scientific support is weak and it may not suit all learners or higher-level language needs.

Flipped Classroom: In a flipped language classroom, instructional content (such as grammar explanations or lecture-style input) is delivered to students outside of class (e.g. via videos or readings), and class time is used for interactive practice, discussion, and production activities. This modern approach encourages active learning and allows more time for communicative tasks in class. Its advantages include greater student autonomy and more efficient use of class time, but it relies on students preparing beforehand, and those who fail to do so may fall behind.

Content and Language Integrated Learning (CLIL): CLIL is a dual-focused approach in which an academic subject (such as science or history) is taught in a foreign language. Learners acquire new knowledge while simultaneously improving their language skills. CLIL provides meaningful context and high motivation—students see the language as a tool for learning interesting content. However, it demands high proficiency and preparation from teachers (who must handle both content and language) and can challenge students if either the content or the language level is too difficult.

All of the above approaches have been studied and tested in previous investigations. Although each method can bring positive effects in certain contexts or for specific language skills, many drawbacks have also been noted. For instance, the ALM is straightforward for teachers to implement and for students to follow through repetitive drilling, and students' pronunciation can be corrected immediately by the teacher (Nita & Syafei, 2012). However, ALM is highly teacher-centered, giving students few opportunities to choose their own expressions or communicate creatively (Sunarwan, 2016). The GTM is a useful method for teaching grammatical rules and structures, as it helps learners improve their mastery of grammar (Chang, 2011). By learning explicit rules, learners can construct new sentences in speaking and writing. However, GTM tends to rely on one-way input from teacher to student rather than interactive communication, and it is difficult for the teacher to attend to students' feelings or emotional engagement (Natsir & Sanjaya, 2014). Moreover, because this method usually uses the learners' L1 as a medium to teach the L2, the development of the foreign language may be constrained by the learners' L1 proficiency (Brown, 2007; Celce-Murcia, 2001; Harmer, 2007). As for the SW, although it guides students to take the main role in the classroom and encourages the teacher to keep silent, its limitation is obvious: without concrete guidance or sufficient instruction, students might become confused and misunderstand the target language and learning objectives. In order to mitigate the limitations of traditional teacher-centered pedagogy and seek more effective methods, attention turned to approaches focused on acquiring language through natural communication. This led to the pedagogy of Communicative Language Teaching (CLT). The central theoretical theme of CLT is communicative competence, a concept advanced in the early 1970s (Habermas, 1970; Hymes, 1971). CLT emphasizes social interaction as the medium of learning, distinguishing it from most prior methods that stressed abstract grammatical or behaviorist practice (Savignon, 2005). Starting in the 1980s, the effectiveness of CLT was investigated by classroom researchers and program evaluators. Most results indicated that compared with traditional audio-lingual methods, CLT-based instruction yields higher fluency and communicative ability in learners (Spada, 2007); learners' comprehension skills can develop to approach those of native speakers (Genesee, 1987). In contrast, the weaknesses of CLT were also observed. Learners taught only through communicative exposure often did not achieve high levels of accuracy or development in certain language forms (Harley & Swain, 1984; Spada & Lightbown, 1989). To address this problem, some later studies incorporated attention to language form (explicit or implicit) within primarily meaning-focused CLT tasks. The results showed that supplementing CLT with form-focused instruction led to gains in learners' ability to use the target language accurately (Norris & Ortega, 2000). While CLT is a philosophy of learning a language through communication, Task-Based Language Teaching (TBLT) is considered the latest methodological realization of CLT (Nunan, 2003). (The above discussion introduced several examples of traditional language pedagogy, such as ALM, GTM, SW, etc., along with their origins, implementation, advantages, and disadvantages. We have seen a shift from form-focused pedagogy toward meaning-focused approaches like CLT. The following sections provide information related to TBLT.)

2.1.2 Task-Based Language Teaching (TBLT)

TBLT is referred to as an educational framework which applies both practice and theory to EFL and ESL in an instructional setting (Van den Branden et al., 2009). It was born out of the paradigm transformation from teacher-centered pedagogic methods in the 1980s. Pioneered by (Prabhu, 1987), this transformation focuses on the shift from the earlier Natural and Audiolingual Approach to the latter Communicative Approach. (Richards & Rodgers, 2014) explained the definition of TBLT: It is an approach based on the application of tasks as the core part of planning and instruction in language teaching. Before designing and applying TBLT, however, it is necessary to clarify what a 'task' is.

The definitions of 'task' are numerous, however, a common illustration is that a 'task' needs to be a target-oriented, significant activity, consisting of integral written and/or spoken use of language. In addition, a task should set a background or a context for activating the procedure of acquisition, thus promoting language development through process, product or both (Samuda et al., 2008). Specifically, a task could be things being done in everyday life, such as booking a hotel room, making an airline reservation, borrowing a book from library, etc (Long, 1985). Furthermore, when designing a language learning 'task', (Nunan, 1989) suggested that there are six indispensable components: goals, input, activities, teacher role, learner role and settings. Meanwhile, the category of pedagogic tasks were proposed by Nunan (1989), including jigsaw tasks, information-gap tasks, problem-solving tasks, decision-making tasks, and opinion exchange tasks. For example, jigsaw tasks root in the conception of cooperative learning, and it usually divides learners into four to six people in a group, in which learners could focus on their learning material and have to cooperate with peers to reach the target (Aronson, 2002).

As many researchers became interested in exploring 'tasks', there are two main reasons to motivate them, one is the production of a communicative task might stimulate learners to notice and memorize the included information in the language they used. And this could lead to a transformation in learners interlanguage method. The dominating question of TBLT is conditions, the opportunity and the approach for language learners to acquire new forms (Gilabert et al., 2016). The other reason would be from the perspective of pedagogy, which means that the exploration in 'task' could offer practical evidence to design of teaching and materials, as it identifies the task's features which might affect learners' way of understanding, production and acquisition a language, and improve their ability of communication.

In the past 20 years, TBLT has become the leading teaching approach in multiple contexts and has been widely implemented in many countries (Andon & Eckerth, 2009; Carless et al., 2012; East, 2012; Van den Branden, 2006). Referring to its effectiveness in L2 language learning, plenty of studies intended to show TBLT's positive effects in L2 performance and competence. Specifically, TBLT offers chances to apply meaningful and effective activities to boost discourse language use in language classroom (Sumarsono et al., 2020), while learners' oral accuracy and fluency in speaking skills could be positively influenced (Angelina & Garcia-Carbonell, 2019). Compared with traditional teaching methods, TBLT could encourage more production and concentrates on interaction in communication (Ellis, 2018). Moreover, TBLT could compel the learners to learn and work with their peers and interact spontaneously with each other, in the way the feel of hesitation and fear would be removed (Ganta, 2015). However, the weaknesses of TBLT was noticed: (1) TBLT encourages learners to speak or communicate more, in which the fluency would be emphasized and the accuracy might be ignored. As a result, learners could be potentially encouraged to create sentences which are not well-formed. (2) For learners who do not have related vocabulary or knowledge about the task in advance, it would be harder for them to produce any written or oral output, thus the effectiveness of TBLT itself might be negatively affected.

2.1.3 Technology-mediated Task-Based Language Teaching (TMTBLT)

Recently, with the development of integrating computer and information technologies in education, students could engage in "doing", rather than just hearing about language and culture from their teachers in classrooms or reading from their textbooks through technology-mediated creation and transformation processes. This is the reflection of the potential of the technologies which activates learners to engage in holistic tasks, and these tasks could motivate these learners to be involved in TBLT, a well-theorized approach to language teaching (Van den Branden et al., 2009). The integration between technology and TBLT has been explored frequently. Some research highlighted the crucial role technology plays theoretically and pedagogically (Chen, 2021; Eslami & and Kung, 2016; Xue, 2022). For example, in (Chen, 2021), the motivation and opportunity of practicing oral English skill for language students has been raised significantly with the assistance of technology. Apart from supported by mobile devices, another situation of the integration between technology and TBLT is learning management system (LMS), in detail, such as online learning, blended instruction, and flipped classrooms (Calderon & Sood, 2020).

TMTBLT Developed from TBLT and computer-assisted language learning (CALL) fields, in the area of applied linguistics, it has become a new and important research direction. The explosive growth of research will swiftly help us compile results to solve questions about the TBLT/TBLL intersection. From the beginning, the distinction between the "technology-mediated" and "technology-enhanced" tasks must be emphasized. Although most situations may put technology as part of second language (L2) teaching into actual practice, the technology-mediated TBLT pedagogy is grounded on full integration of technology and tasks (Gonzales-Lloret & Ortega, 2014). For instance, a technology-*mediated* task should be based on the context that it is reasonable to practice the present task with the mediation of technology, such as booking a hotel through email, discussing and making a decision through WeChat, and ordering a pizza through an online website, etc. On the other hand, technology-*enhanced* could be any tasks done through a technological medium, for instance, to do some multiple-choice questions on a cellphone. Apart from the basic requirements of a 'task' – that is from communicative activities accompanied by traditional form-focused approaches (Ellis, 2009) – authentic activities which reflect real-world interactions happening in daily life should also be considered (Long, 2016; Long et al., 1985). Most of the previous TMTBLT research applied narration tasks, information gap tasks and agreement tasks, which should be outcome-based focusing on meaning instead of linguistic form, in other word, a reason should be set to use the language beyond the activity itself (Smith & González-Lloret, 2021). Further, Gonzales-Lloret and Ortega (2014) also forward five principles for designing a TMTBLT task, including:

(1) **Primary focus on meaning:** although a learning goal is needed in the process of learning, it should be hidden or be 'implicit' for learners, therefore the learning part should be incidental.

(2) **Goal orientation**: the plan of the task has to provide a language-and-action request, for example, a communicative objective (e.g. to arrange two information-gaped participants to exchange information), or an outcome-required task (e.g. to make a decision, to produce a plan, or to play/win a game).

(3) **Learner-centeredness:** for traditional classes, teachers are usually the one who orient the learning procedure. However, in TBLT, students are the ones centered in learning. Moreover, a task should analyse and direct learners' language demands, and the implementing process should not be uniform but allow for diversity and flexibility.

(4) **Holism:** a task designed with the context of real language use, which refers to the feature of 'authenticity' and 'real-world relationship'. and

(5) **Reflective learning:** The TBLT advocates that learners learn language within the immersion of direct experience and doing things with words. Meanwhile, it should offer chances for language learners to reflect their own learning.

In Gonzales-Lloret and Ortega (2014), benefits brought by TMTBLT were listed, they argue that (1) it could decrease learners' embarrassment, fear of failure and losing face; (2) it could develop learners' tolerance and motivation to take risks, while the creation will also be stimulated when making meaning by using language; (3) it could assist learners by providing an authentic foreign language environment and meet other speakers of language even in remote areas. For instance, a daily TMTBLT scenario could be chatting with friends or co-workers via network sites, emails or the face-time platform, processing a writing activity via blogs, forums or wikis, or engaging in massively multiplayer online games (MMOGs) which build social, interactive, and immersive environments for players. More typically, since 2008, massive open online courses (MOOCs) broke the limitation of time and space, providing open and free learning resources for all individuals rather than students associated with any particular institution (Liyanagunawardena et al., 2013).

Although technology could bring convenience and benefits to language learning, (Smith & González-Lloret, 2021) state that 'not all technologies are equal', because they direct to established shapes and expectations, while some technologies might not fit in some tasks while other tasks could be accomplished. Specifically, some of the language learning apps and software are designed to convey a specific pedagogy, thus the way of using this app/software by the teacher or user is predetermined by the designer. This means users are passively required to know how to adapt in technology-mediated learning, rather than technology-mediated learning providing flexible service to users under the circumstance. On the other hand, from the objective perspective, the utilization of technology needs to be support by equipment, talent, place, bankroll, etc., any or all of them could be the obstacle for developing technology-mediated language learning.

The present study will explore the effect of 'cognitive task complexity' in the context of TMTBLT. Similarly, a few previous studies based on CALL applied Robinson's Cognitive Hypothesis Framework (this framework would be introduced in part 2.2.2) to understand the relationship between task complexity and learners' language performance. One typical series of investigations done by Adams and Nik on the basis of computer-mediated communication (CMC) environment show that a decrease of interactional modifications happened when the complexity raised (by decreasing task structure) (Alwi et al., 2012), and learners' accuracy was promoted when facing a simpler task (by increasing task structure) , but their quality and complexity of language was not changed (Alwi & Aloesnita, 2010). Furthermore, when adopting a task without prior knowledge (to make the task more complex), learners performed with higher accuracy, lexical complexity but fewer interactions (Adams & Nik Mohd Alwi, 2014). The above three studies got results running totally opposite to Robinson's CH, and scholars attribute the discrepancy to text-based CMC's unique characteristics, such as more planning time and opportunities to composing and editing text before sending messages. As for other research exploring complexity in TMTBLT, (Baralt, 2013) compared CMC and face-to-face contexts, the results showed that the situation in the face-to-face context matched Robinson's CH: more cognitive complexity leads to higher complexity, accuracy and fluency, whereas the CMC did not. In addition, learners in face-to-face groups perceived the task as more difficult than those in the CMC groups. According to above investigations, it seems that Robinson's CH could well correspond and explain complexity in classic tasks but not in TMTBLT tasks. This indicates that the CH framework might not transfer easily to TM tasks and environments. Further explorations are needed to understand and fill this gap.

For TMTBLT, there are still many unknown areas. Burgeoning topics such as ICALL (intelligent CALL) (Heift & Schulze, 2007) and attempts at helping students to be involved in more immersive and realistic settings (Peterson, 2012) might bring relatively more interesting methods and platforms for TMTBLT curriculum. On the other hand, although there is research focusing on the relationship between task performance and cognitive task complexity in traditional TBLT for L2 learning (Donate, 2018; Sasayama, 2016), there are still many undiscovered aspects of TMTBLT complexity. A basic question is: how does technology influence the task's complexity? For example, an annotation link in an e-reading material might assist readers to understand the text better and further, reduce the complexity of the reading task, however, the organizing (for designers) and using process (for readers) of the text might be more complex. Moreover, the scales of measuring the complexity of TMTBLT tasks are also needed to be considered, which will be discussed and explored in the present research.

Apart from general technology, other auxiliary methods for language acquisition have also been explored and tested by scholars. Games, which are relatively popular and easily accepted by younger groups, have also been studied, the following part will explain this.

2.1.4 Gamification in ESL

According to a review of related studies, gamification has been widely applied in many non-English-speaking countries in various skills of English language in the field of EFL and ESL from the second half of 2010 (Zhang & Hasim, 2023). Therefore, the definition of gamification was forwarded by researchers: (Werbach, 2014) defined gamification as the procedure of making non-game like activities to be more game-like. While (Landers et al., 2017) regarded gamification as the application of game-designed elements such as achievements, points and ranking lists in a non-game context to offer the experience of game-like learning.

According to (Bicen & Kocakoyun, 2018), when designing gamified learning contexts, three distinctive concepts should be involved. (1) game dynamics, which refers to the explicit and enforced competition/reward/self-expression/status under rules; (2) game mechanics, which indicates narrative context, level-system, ranking system, challenge, achievements, and others; (3) game elements/components consist of self-representation with avatars, feedback, points, trophies, badges, progress bar and virtual presents, etc. (Deterding et al., 2011). With the above ingredients, researchers believe and showed that gamification could bring positive effects in learning by offering visible incentives for learners and expected behaviors in education (Shortt et al., 2023), raising learners' engagement, creating interactive learning backgrounds (Kapp, 2012). As electrictronic devices and the Internet becomes more and more prevalent, it is common to implement gamification in portable mobile devices and digital environment (Su et al., 2021). Specifically, this also happened in ESL/EFL teaching and learning (Dehghanzadeh et al., 2021). According to the using of this method, scholars divided it into two branches: mobile game-based language learning (MGBLL) and non-mobile gamebased language learning (NMGBLL). As the present study focuses on TMTBLT, the MGBLL would be deeply discussed in the following text.

Duolingo is a widely applied popular platform of gamification in Mobile-Assisted Language Learning, it could be a typical example for exploring MGBLL. (Shortt et al., 2023) systematically reviewed 35 articles from 2012 to 2020 including the questions of pedagogy, design and application in the use of Duolingo. The results showed that the use of Duolingo leads to positive effects such as improvement in English vocabulary, listening and English communicative skills, and boost learners' academic achievement in English. They believe that meaningful feedback, context and collaboration should be valued in MGBLL to maximize the benefits of this method. As for more consequences of studies in gamification in language learning, a variety of studies has reported and confirmed the benefits: (1) learner's English learning anxiety might be reduced (Barcomb & Cardoso, 2020; Hung, 2018; Hwang et al., 2017); (2) learner's motivation, engagement and interest might be increased (Almusharraf, 2021; Bicen & Kocakoyun, 2018; Hwang et al., 2017; Reynolds & Taylor, 2020; Zohud, 2019; Zou, 2020); (3) learner's language performance might be improved (Barcomb & Cardoso, 2020; Hwang et al., 2017; Ling et al., 2019; Zohud, 2019); (4) learner's autonomy might be better fostered (Setiawan & Wiedarti, 2020; Zohud, 2019; Zou, 2020). However, some studies indicate that although gamification benefits learner's short-term performance, the final learning achievement was not affected (Calvo-Ferrer, 2017; Domínguez et al., 2013). Furthermore, Hanus and Fox (2015) declared that for learners who are already motivated to learn, their intrinsic motivation might be harmed by gamified learning activities. In addition, due to the problem of Internet connection and other objective conditions based on designing of the game itself, some students are not able to use the gamified learning APP (Ebadi et al., 2023).

According to the above, it seems that gamification in language learning is a double-edged sword, and more practices and exploration in MGBLL are needed. In the present study, gamified elements will be combined with TMTBLT to investigate the unknown potential for both contexts.

2.2 The Hypothesis of Cognitive Task Complexity

Does the complexity of a task, and the sequence of its procedures, affect language learning? Over two decades, in the area of TBLT, a debate in relation to this question has been happening. That debate is related to the procedure of sequencing and selecting tasks to improve the quality of language learning and development (Robinson, 2011). In the field of TBLT, some scholars (Candlin, 1987) have tried to investigate the best approach to organizing pedagogic tasks, which considered the cognitive demands needed to perform the tasks. According to ideas for sequencing decisions and task complexity, two important theoretical frameworks about task complexity appeared to explain the association between learners' cognitive processes and manipulations in L2 tasks. The first framework is Robinson's Cognition Hypothesis (CH) (2001a), and another framework is Skehan's Limited Attentional Capacity Model (LAC) (1998). Although most of the literature on task complexity applied these two frameworks to explain their research findings and competing hypotheses, they radically disagree on the effects that cognitive complexity has on learners' L2 performance, progress, and acquisition. Whether learners' attention could be paid concurrently to different perspectives of the language, that is, complexity and accuracy, still remains unresolved until the present day. The following sections will discuss these two frameworks in detail.

2.2.1 Skehan's Limited Attentional Capacity Model (Trade-off Theory)

Skehan's LAC model explains cognitive task complexity in terms of three main components: cognitive complexity, code complexity, and communicative stress

(Skehan, 1996). These refer, respectively, to the mental processing demands of the task, the linguistic complexity of the language used, and the performance conditions (such as time pressure, stakes, or audience) under which the task is performed. Skehan (1996) argued that higher processing demands make a task more cognitively complex, whereas greater familiarity with the task content can make the task simpler. Like Robinson's framework, LAC links cognitive processes to task characteristics, but it places a strong emphasis on the limits of human cognitive resources. The core premise of LAC is that humans have limited attentional capacity for language processing. As task complexity increases, learners do not have enough working memory resources to pay attention to everything at once (Skehan, 1998). Consequently, there is competition between different aspects of language performance—specifically, complexity, accuracy, and fluency—for the available attentional resources. This leads to a "trade-off" effect: focusing attention on one aspect (for example, attempting more complex sentence structures) will likely cause performance in another aspect (such as grammatical accuracy or fluency) to suffer. In other words, when various task demands exceed the learner's available cognitive resources, the learner must prioritize certain performance areas at the expense of others. Typically, learners prioritize meaning to achieve communicative goals, so fluency is often maintained while complexity and accuracy decline. Even when attention is directed to form, there can be a tension between complexity and accuracy, as attending to one of these often limits resources for the other. This perspective is often referred to as the trade-off hypothesis (Skehan, 1998). According to Skehan's model, familiarity with task content can mitigate some of these trade-offs: if learners know the content well, the task's effective complexity is reduced, freeing up attention for form (Skehan, 1996, 1998). The implication for task design is that increasing a task's cognitive complexity will not uniformly improve all aspects of language performance. Instructors may need to limit the complexity of tasks or provide support (e.g., planning time or task repetition) to avoid overwhelming learners. For example, providing pre-task planning time has been shown to help learners devote attention to more complex language without as much detriment to fluency. In summary, Skehan's Limited Attentional Capacity model posits that because working memory and attention are limited, increasing a task's complexity will force learners to make trade-offs in performance. If a task is very demanding, learners might attend primarily to one aspect of production (say, conveying meaning fluently) while neglecting others (such as using complex syntax accurately). This trade-off theory stands in contrast to Robinson's hypothesis, which suggests a different outcome when tasks become more complex. Specifically, Robinson's Cognition Hypothesis proposes that, under certain conditions, making a task more complex can push learners to produce more accurate and complex language (with less emphasis on fluency), rather than simply causing a breakdown in performance. In Skehan's view, however, humans have a limited capacity and "prioritization and predisposition (or both) seem to orient performance toward one (or two) of the three areas of accuracy, fluency, and complexity" (Skehan, 1998). This is the key difference between the two models: CH predicts that increasing cognitive complexity can yield improvements in some dimensions of language production, whereas LAC argues that not all dimensions can improve simultaneously due to attentional limitations.

2.2.2 Robinson's Cognition Hypothesis

Robinson's CH asserts that learners' L2 task performance can be classified and predicted by his Triadic Componential Framework (TCF), which provides a taxonomy of task characteristics. There are three main components in the TCF: cognitive task complexity, task conditions, and task difficulty (Robinson, 2005). With regard to cognitive task complexity, the TCF proposes that the cognitive demands of a task are of two types: resource-directing variables and resource-dispersing variables. In Robinson's framework, cognitive task complexity refers to the complexity of the task itself, i.e. the demands placed on the learner's cognitive abilities. Under this category, resource-directing variables are those task features that direct the learner's attention to specific aspects of language or meaning. For example, ±few elements refers to tasks involving a greater or smaller number of elements or details (a task with many elements is more complex because it requires tracking more information), and ±here-and-now refers to whether the task is about the immediate context or removed in time/space (tasks not in the here-and-now, e.g. discussing past or hypothetical events, are more complex because they require more linguistic devices like past tense or deixis). In contrast, resource-dispersing variables are task features that increase cognitive load without specifically directing attention to language form. Examples include ±prior knowledge (whether learners have background knowledge about the task content) and ±planning (whether learners are given planning time before the task). A complex task along a resource-dispersing dimension, such as no planning time, taxes the learner's processing capacity but does not inherently guide them to any particular linguistic feature. According to Robinson, along resource-directing dimensions, making a task more complex will elicit more complex and accurate (but less fluent) language performance, compared to a simpler task. This is because the additional cognitive demands direct learners to engage with language more deeply (e.g., reasoning about more elements or non-immediate contexts may push them to use more complex syntax or precise vocabulary), even though it may slow them down (fluency drops). In contrast, along resource-dispersing dimensions, making a task more complex (e.g., removing planning time or adding time pressure) will tend to reduce complexity, accuracy, and fluency of performance relative to a simple version of the task. This is because such complexities drain general attentional resources without focusing learners on useful language practice. The second component, task conditions, refers to the interactive conditions under which a task is performed. For instance, tasks can vary in participation variables like one-way vs. two-way (whether information flows one-directionally or needs mutual exchange) or convergent vs. divergent (whether learners must agree on an outcome or can have different viewpoints). These conditions affect the interaction patterns and can influence task performance. The third component, task difficulty, concerns learner factors such as affective variables (e.g., motivation, confidence, anxiety) and ability variables (e.g., proficiency, aptitude, intelligence). These learner variables can mediate how difficult a given task feels and can interact with the task's design features. Related studies about CH: In SLA research, task complexity has been manipulated in many studies to test its effects on learners' L2 performance, development, and learning (Albert, 2011; Baralt, 2013; Donate, 2018; Dörnyei & Kormos, 2000; Robinson, 2001a; Sasayama, 2016; Sasayama & Izumi, 2012; Xu et al., 2023). One major reason cognitive task complexity is widely studied is that it allows researchers to systematically control task demands in ways that might facilitate L2 processing (Torres, 2013). For example, (Xu et al., 2023) conducted two studies in this area: one was a validation study to ensure that their task-complexity manipulations had the intended cognitive effects, and the other was the main study to assess L2 learners' writing performance under different task complexities. In the validation study, they used a dual-task methodology, expert judgments, and post-task questionnaires; the results confirmed that the "complex" task version was indeed more cognitively demanding than the "simple" task. In the main study, 65 Chinese L2 learners completed two argumentative writing tasks (one simple, one complex), and their performance was measured in terms of lexical complexity, syntactic complexity, accuracy, fluency, and functional adequacy. The results showed no significant differences in accuracy, fluency, or syntactic complexity between the simple and complex tasks, while there was an increasing tendency in functional adequacy and lexical complexity for the complex task. Sasayama and Izumi (2012) investigated how task complexity influences oral production by manipulating cognitive demands along both of Robinson's dimensions. They varied task complexity via ±elements (a resource-directing factor) and the presence/absence of pre-task planning (a resource-dispersing factor). Twenty-three high school ESL learners were asked to complete narrative tasks that differed in the number of characters in picture prompts (more characters = more elements). The results showed that the more complex task (with more elements and no planning) led to higher syntactic complexity in learners' speech, but accuracy and fluency were negatively affected. Moreover, providing pre-task planning helped learners increase syntactic complexity, while lack of planning negatively affected fluency. Overall, Sasayama and Izumi's study concluded that elements of both Robinson's CH and Skehan's LAC were partially supported by the data (some measures improved with complexity, others declined), highlighting the need for a nuanced view. (Michel, 2011) examined how task complexity affected interaction and various CAF measures in spoken performance. Sixty-four ESL learners (L1 Dutch) were assigned to either a conversation (dialogic) or monologic condition to complete a decision-making task. The tasks involved making decisions (e.g., choosing activities for a dating couple or a study pair) and were manipulated by ±elements to adjust complexity. The findings revealed that the complex versions of the tasks led to greater lexical diversity in both monologic and dialogic conditions, while no clear effects were found on other measures of accuracy or fluency. Three-way Mapping Constraint: Robinson (2010) has argued that any taxonomy of task characteristics must satisfy a "three-way mapping constraint". This refers to aligning task design along three interfaces:

Target task analysis: ensuring that pedagogic tasks correspond to real-world target tasks in the abilities they demand (i.e., the task is chosen because its requirements map onto the language and skills needed in real-life performance).

Learning process: ensuring that tasks are selected and sequenced in a way that promotes learning and performance based on SLA theory (i.e., guided by what we know about how input, interaction, and performance affect learning).

Operational consistency: ensuring the task design is feasible and coherent for both teachers and learners, meaning that the task components and categories can be applied consistently so that task designs are comparable across different contexts.

Based on these principles, many researchers aimed to grade and sequence tasks, seeking pedagogically sound, research-supported, and theoretically driven criteria for task sequencing. Optimally sequencing tasks is vital for both immediate performance and long-term acquisition. Notably, most existing studies measure learners' task performance rather than underlying language competence — a gap which will be discussed in Section 2.4. Earlier, in the 1980s, there were efforts to improve theories of task complexity and sequencing. (Candlin, 1987) suggested six factors to aid in sequencing decisions: cognitive load, communicative stress, code complexity, content continuity, process continuity, and particularity & generalizability.

Similarly, (Brindley, 1987) proposed a three-category framework to distinguish task complexity: text factors (e.g., length or complexity of input texts), learner factors (e.g., learner motivation, prior experience, cultural background), and task factors (e.g., inherent cognitive complexity, number of steps, available time). These early frameworks by Candlin and Brindley provided a foundation for Robinson's later work, as they share many common ideas. In line with these, Nunan (1989) put forward a set of criteria based on Brindley's framework, with more detailed subcomponents such as the grammatical complexity of input, the amount of low-frequency vocabulary, and procedural factors like the type of operations learners must perform on the input (Nunan, 2004), p.122). These early frameworks offered ways to deliberately manipulate task complexity (e.g., adjusting cognitive load). However, it became clear that one must distinguish the manipulable factors (those task features we can design and control) from inherent or uncontrollable factors (like a given learner's aptitude or prior knowledge). This conceptual progress paved the way for Robinson's SSARC model, which provides a systematic approach to sequencing tasks.

2.2.3 Robinson's SSARC Model

Following demands mentioned in the previous section, Robinson's SSARC (Simplify, Stabilize/ Automatize/ Restructure, Complexity) Model is the most recent and convincing proposal, which relates to his Cognitive Hypothesis. Two claims were advanced: (1) sequencing must be solely based on cognitive factors; (2) tasks must be sequenced with the order from simple to complex (Robinson, 2010). Briefly, The SSARC Model provides a guidance for sequencing and designing tasks, basically follow the order from the lower to higher cognitive demand and gradually approach the demand of realistic target tasks.

In detail, the model suggests to sequence tasks in accordance with the following three steps (Robinson, 2022):

Step 1: ('SS' for Simply and Stabilize) task versions are simple on both resource-dispersing and resource-directing dimensions of task demands specified in the TCF. (e.g. - few elements, + prior knowledge)

Step 2: ('A' for Automatize) task versions are simple on resource-directing dimensions of task demands, but complex on resource-dispersing dimensions specified in the TCF. (e.g. - few elements, - prior knowledge)

Step 3: ('RC' for Restructure and Complexify) task versions are complex on both resource-directing dimensions of task demands, and resource-dispersing dimensions specified in the TCF. (e.g. + few elements, - prior knowledge)

Some studies try to test and verify the suppose of SSARC model. Most of them compared the supposed sequence to other sequences (such as from complex to simple or random).(Baralt et al., 2014) compared the sequence of the simple-complex to other sequences (CSC, CCS and SCS) with a narrative task in traditional face-to-face and computerized online settings. While (Levkina et al., 2014) tested the SSARC model by comparing three sequences of tasks (SC, CS, and random). (Malicka, 2020) focused on how simple-complex sequencing benefits learners' L2 production, the degree was defined like this: simple is 'S', complex is 'C' and most complex is '+C'. The following sequences were tested: 1) S+CC; 2) CS+C; 3) C+CS; 4) +CCS and 5) +CSC. The research results show that the cognitive task complexity led to the prominent qualitative changes in language performance, as opposed to sequencing, which is similar with what (Baralt et al., 2014) found: compared with manipulating the sequence, manipulating the cognitive task complexity would bring more obvious effect on learners' performance.

Because task sequence can produce varied outcomes, the present study explicitly manipulates three distinct sequences—SMC, CMS, and MCS—to investigate how different orders of tasks shape L2 performance.

The above contents cover two manipulable elements in tasks: cognitive complexity and sequence of tasks. The two main hypotheses promoted by Robinson and Skehan indicate different results among the dimensions of complexity, accuracy and fluency. The task sequence is also exploitable, thus further studies could concentrate more on SSARC model. In the next section, elements refer to the emotion would be introduced, including foreign language anxiety (FLA), foreign language fear (FLF) and foreign language enjoyment (FLE).

2.3 Foreign Language Emotion

Language learning is easily affected by learners' internal and external factors, and many scholars are curious about sociopsychological factors in L2 learning. From early studies, people explored how negative emotion effects learners' outcome performance, especially foreign language anxiety (Horwitz et al., 1986). However, later people started to focus more on positive psychology (PP), therefore emotions like enjoyment gained more attention (Dewaele & MacIntyre, 2014). It is likely that both anxiety and enjoyment affect language learning, although I will argue that most of the previous studies mix anxiety and fear together. This study proposes that they are separate conceptions based on their definition, context, cause and effect, and this point will be argued in the following parts. Thus, the present study will investigate the effects brought by both fear and enjoyment to learners' competence, thus in the following paragraphs, the information about these emotions would be presented.

2.3.1 The Concept of Foreign Language Anxiety (FLA)

The FLA is considered by many to significantly impact learners' performance and achievements in their L2 learning processes (Baralt & Gurzynski-Weiss, 2011; Horwitz, 2001). The concept of FLA was first proposed by (Horwitz et al., 1986). She introduced language anxiety as arising from the inherent difficulties in learning a second language. The learning of a foreign language is often accompanied by anxiety, the degree of which varies depending on how well the student is doing throughout their education. It might cause a disparity in the results of learning for students of a certain L2 when compared to students of other L2s. Anxiety may develop as a result of the learner's fear of receiving a negative evaluation, their anxiety about performing poorly on a test, or their concern about their ability to communicate effectively ((Horwitz et al., 1986), p.127):

> *Because foreign language anxiety concerns performance evaluation within an academic and social context, it is useful to draw parallels between it and three related performance anxieties: 1) communication apprehension; 2) test anxiety; and 3) fear of negative evaluation. (Horwitz, et al, 1986, p.127)*

Horwitz (1986) also produced the definition of Language Classroom Anxiety Scale, which is in the foreign language classroom, the most widely used scale, is considered the standard method for testing language anxiety. The scale consists of 33 questions and statements (e.g. 'I never feel quite sure of myself when I am speaking in my foreign language class', 'I don't worry about making mistakes in language class' and 'I tremble when I know that I'm going to be called on in language class') with obvious part-whole corrections, with terms of social comparisons, negative performance expectancies, social comparisons and avoidance behaviors. As noted above, the FLA identified three major anxiety-provoking L2 situations: 1) communication apprehension (CA), 2) fear of negative evaluation (FNE), and 3) test anxiety (TA).

More previous research in this area have been interested in two perspectives:

(a) to find out sources that contribute to explain it, and

(b) to explore the association between L2 achievement and FLA.

Generally, the most prominent strand of research focused on the relationship between FLA and L2 achievement in the context of foreign language classrooms by using anxiety scales and questionnaires (Cheng et al., 1999), and a negative correlation between FLA and L2 achievement has been detected from the results of most of these studies (Aida, 1994; Horwitz, 2001). Only few studies showed that there is no relationship (MacIntyre & Gardner, 1989) or even ones (Spielmann & Radnofsky, 2001).

Generally, studies for FLA followed Horwitz's framework, they mix the emotion of anxiety and fear together, this might be caused by Horwitz's category for foreign language classroom anxiety (communication apprehension, fear of negative evaluation, and test anxiety). Whereas, the present study argues that the distinguish between foreign language anxiety and foreign language fear based on their definitions, causes and phenomenon, must be made explicit. Further details would be illustrated in the next section.

2.3.2 The Concept of Foreign Language Fear (FLF)

Foreign Language Fear, as a separate emotion to anxiety, has not been studied as an independent object in SLA. Usually, scholars would bind it with anxiety and regard it as a subcategory of Foreign Language Anxiety. For instance, 'fear of negative evaluation in foreign language anxiety' (Šafranj & Zivlak, 2019; Tzoannopoulou, 2016; Wardhani, 2019). On the other hand, scholars tend to link fear to specific contexts, for example, 'fear of negative evaluation' and 'fear of speaking' (Cutrone, 2009). This naturally leads to the questions: Is there a relevant difference between fear and anxiety, and, if so, what is the distinction and how is it important to SLA studies? These questions would be answered and explained in the following sections, and the difference between fear and anxiety will be shown based on their standard psychological definitions, how they are categorized as achievement emotions, their precipitating events, consequences and caused behaviour.

2.3.2.1 The Definition of Anxiety and Fear

According to the American Psychological Association (APA) Dictionary of Psychology, anxiety is a kind of emotion caused by something that might happen in the future, and it will lead to a series of prototypical physiological reactions:

"Anxiety is an emotion characterized by apprehension and somatic symptoms of tension in which an individual anticipates impending danger, catastrophe, or misfortune. The body often mobilizes itself to meet the perceived threat: Muscles become tense, breathing is faster, and the heart beats more rapidly. Anxiety is considered a future-oriented, long-acting response broadly focused on a diffuse threat" (APA Dictionary of Psychology: Anxiety, n.d.)

While 'fear' is defined as:

"a basic, intense emotion aroused by the detection of imminent threat, involving an immediate alarm reaction that mobilizes the organism by triggering a set

of physiological changes. These include rapid heartbeat, redirection of blood flow away from the periphery toward the gut, tensing of the muscles, and a general mobilization of the organism to take action. Fear differs from anxiety in that the former is considered an appropriate short-term response to a present, clearly identifiable threat" (APA Dictionary of Psychology: Fear, n.d.)

Anxiety and fear are distinct from both conceptual and physiological perspectives. Anxiety usually focuses on a diffuse threat, and it is considered a future-oriented and long-acting response. In contrast, fear is regarded as an appropriate short-term response to a present, clearly identifiable threat. Particularly, Fear would happen when people are avoiding or escaping an aversive stimulus, on the other hand, anxiety is induced in situations when entering a potentially dangerous situation.

Based on the above comparison, it is necessary to re-examine the negative emotion called 'anxiety' that arises among foreign language learners when they are required to deal with language tasks or activities. Typically, for example, for some ESL learners, on the one hand, they could handle a daily English conversation with their peers or teachers relatively well, however, on the other hand, when they are put into more formal context like delivering a speech, their performance might not reach the same level as the previous one. Specifically, they could perform well in the group discussion with their peers, with creative ideas and good performance, but tend to avoid or be afraid of presenting and sharing their discussion results in front of the whole class. Moreover, some learners could get satisfactory scores in daily quizzes or informal tests, but they would perform poorly in the final test or important exam. The above anecdotal descriptive cases could be common in and out of foreign language classes across the globe. And this temporary, occasional emotion appears to be closer to the conception of 'fear' than 'anxiety'. Because these occasions match the description of 'fear' above: an appropriate short-term response to a present unpredictable threat, rather than a future-oriented, long-acting response broadly focused on a diffuse threat (anxiety).

The above paragraphs distinguish between fear and anxiety using their standard definitions in the APA Dictionary of Psychology. To further clarify how these distinct emotions function in learning contexts, the Control-Value Theory (CVT) of achievement emotions (Pekrun et al., 2007) will be introduced next. CVT highlights two critical appraisal dimensions that give rise to emotions in academic settings: perceived control over learning activities and outcomes (e.g., “Can I handle this task successfully?”) and perceived value of these activities and outcomes (e.g., “How important or useful is success here?”). When analyzing the taxonomy of achievement emotions in light of CVT, additional nuances appear regarding the differences between fear and anxiety.

In Control-Value Theory, the concept of “achievement emotion” refers to “emotions directly tied to achievement activities or achievement outcomes” (Pekrun et al., 2007), p. 316). Here, the achievement refers to the learner’s final production—often conceptualized as “performance” or “competence”. According to CVT, achievement emotions vary along multiple dimensions, but Pekrun et al. (2007) emphasize three in particular—sometimes called the Three-Dimensional Taxonomy of Achievement Emotions:

Object focus: Distinguishes activity-related emotions (those occurring during the learning or performance process, e.g. enjoyment, frustration, boredom) from outcome-related emotions (those linked to after the activity, e.g. hope, anxiety, shame, or relief).

Valence: Differentiates positive emotions (e.g., enjoyment, relief) from negative emotions (e.g., anxiety, fear, shame).

Activation: Separates activating emotions (e.g., enjoyment, anxiety, fear, anger) from deactivating emotions (e.g., relaxation, contentment, boredom).

Some concrete examples help illustrate how these dimensions interact. For object focus, if an emotion arises while the learner is in the midst of a task, it is an activity-related emotion such as enjoyment (positive) or frustration (negative). But if it is tied specifically to the result or potential result (e.g., success, failure), it is outcome-related, such as pride, anxiety, or fear about the consequences. For valence, an unpleasant emotion (e.g., fear, anxiety) is negative; a pleasurable emotion (e.g., enjoyment, contentment) is positive. Finally, for activation level, an activating emotion typically features heightened physiological arousal and motivates the learner to act (e.g., fear can trigger the urge to escape or avoid further engagement, while anxiety may heighten vigilance); a deactivating emotion (e.g., boredom, relaxation) tends to reduce one's immediate drive or energy to respond.

According to this Three-Dimensional Taxonomy (Pekrun et al., 2007), it appears that anxiety is usually an outcome-related, negative, and activating emotion, as it arises from worrying about future feedback or possible negative results of a learning activity. It evokes increased physiological and psychological arousal (e.g., a pounding heart and racing thoughts). Fear, in contrast, is often an activity-related, negative, and activating emotion: it emerges from a present threat in the ongoing learning event—such as being called on unexpectedly—and tends to subside once the event has passed. Thus, fear focuses more on the immediate situation (e.g., "I must escape this threat right now"), whereas anxiety centers on uncertainty or potential negative results (e.g., "I might fail an exam tomorrow, and that will be awful").

By linking these ideas to "control" and "value" perceptions, CVT clarifies why fear tends to happen during a threatening classroom moment (when the learner feels low control over a highly valued performance demand), whereas anxiety often occurs prior to or following an event (when uncertainty about control or outcomes remains high). In other words, if learners perceive an immediate negative outcome they cannot escape, they are likely to feel fear; if they perceive a possible future negative outcome, often with some degree of uncertainty, they are likely to experience anxiety. This expanded theoretical framework supports the argument that fear and anxiety—although frequently merged in language research—are more accurately treated as separate but related constructs.

Based both on APA's definition and Pekrun's Three-Dimensional Taxonomy of Achievement Emotions, fear and anxiety are two separate emotions. This brings potential impact to what prevenient studies follow as their theoretical support. Although anxiety and fear have many aspects in common, it is necessary to emphasize the rigor of defining an emotion in psychology. It is hoped that this will be of use to scholars exploring these negative emotions in order to eliminate or control them by knowing them in deeper and clearer ways. As mentioned, an influential related framework is Horwitz's Foreign Language Classroom Anxiety (1986), and it categorizes FLCA as communication apprehension, test anxiety and fear of negative evaluation. The present study would not overthrow the whole framework, but would ask if 'anxiety' is the real and accurate indication for the kind of emotion FLCA refers to.

First, for communication apprehension, it emphasizes the difficulties ESL learners face when they are required to speak their L2. However, this subcategory is relatively general. Further, communication apprehension occurs differently depending on different situations: 1) Once ESL learners speak English, they might feel nervous or threatened whenever speaking English; 2) ESL learners might, in most circumstances, speak English normally, but once they are required to face implicit or explicit evaluations (e.g. deliver a public speech or navigate an oral English test), they would not perform as good as in normal situations. Second, for test anxiety, the conception is also ambiguous. 1) when facing a test, learners are essentially facing evaluation, the test could be a math test, a language test or any other test; 2) The negative emotion would usually exist until the test finishes, hence the trigger of the emotion could be the test, not the foreign language itself; 3) when facing a daily quiz, compared with the more important final exam, learners might experience a lower degree of negative emotion. Third, for fear of negative evaluation, Horwitz uses 'FEAR' to indicate learners' negative emotion of being evaluated poorly based on their language performance. Similar with the first two categories, this is also directed at the negative emotion triggered by being evaluated by others (e.g. English test judged by scores; public speech judged by audiences like classmates and teachers). The fear here could match the definition of FEAR in APA's Dictionary of Psychology, which, as I have shown above, is distinguished from anxiety. As it is elicited with a specified imminent threat caused by a specialized reason (an unsatisfactory evaluation, for example) and the negative emotion would arise and exist during the processing of being evaluated until it finished.

According to the above illustration and analysis, the FLCA framework could be criticized based on, on the one hand, its ambiguity between fear and anxiety, and, on the other hand, the overly-general subcategories.

Further support and evidence for the present opinion could be provided by the differences of learners' concrete behaviour when anxious or fearful. In neurobehavioral science, scholars also emphasize the usefulness of the distinction between anxiety and fear, meanwhile, the distinction between these two items was supported by the different causes and effects (behaviour) performed by experimental subjects.

Animals' natural responses to fear are indicated as species-specific defense reactions aimed to protect them from threats (Bolles, 1970; Bolles & Fanselow, 1980; Fanselow, 1989; Fanselow, 1994; S. Fanselow, 1984). Therefore, underlying fear, the neural circuitry is a kind of defensive behavioral circuitry (Fanselow, 1994). According to Davis et al, a distinction between sustained threat and acute threat was proposed and it is linked to the distinction between anxiety and fear (Lee & Davis, 1997; Walker & Davis, 1997). This seems to match the definitions of anxiety and fear (APA Dictionary of Psychology) mentioned in previous sections as it defined them as a reaction caused by a short-term stimulus (fear) and a long-term stimulus (anxiety) separately.

Over the process of evolution, all living creatures developed a repertoire of adaptive reactions to evade kinds of threats (Bolles, 1970). (Timberlake & Lucas, 2019) advanced the Predatory Imminence Theory, which divides creatures' defensive behaviors into three stages: pre-encounter, post-encounter, and circa-strike (Three modes of defense). The following paragraphs will use a mouse as an example (CASE 1) to explain the three stages.

1) The pre-encounter reactions. A safe mouse nest is left in a potentially dangerous environment (a new, dark, open area). A mouse in the nest wants to get some food but there is no food in the nest; it has to seek food outside the nest in its wider habitat. In order to minimize the risk, the mouse will reorganize their meals

patterns (Fanselow et al., 1988). And the mouse leaves its nest by using a stretched approach behavior. (Pinel et al., 1994) speculated that this behavior provides a chance for risk assessment, vigilance enhancement and information gathering. Further, this helps ensure that danger is not encountered. In this stage, the new, dark, open area is not dangerous in itself, and will not harm the mouse, but it increases potential risks such as an attack by predators. In this stage, pre-encounter defensive behaviors rely on the prefrontal cortex (a part of the brain), and this is recognized as 'anxiety' (Fanselow, 1989).

2) Post-encounter defense. The story continues. The mouse goes out of its nest, but unfortunately it encounters a predator. Its behaviour changes dramatically when actual danger is encountered (Perusini & Fanselow, 2015). Now the mouse is freezing to avoid being noticed by the predator. As mice are a small, relatively weak and slow rodent, freezing is an effective countermeasure, as a moving prey is more likely to be detected and attacked than a non-moving prey (Perusini & Fanselow, 2015). In this stage, post-encounter defense behaviors rely on the sub-cortical forebrain, and this is recognized as 'fear' (Fanselow, 1989).

3) circa-strike. Unfortunately, the mouse is noticed by the predator, and the predator tries to catch and eat the mouse. Surprisingly, the mouse also tries to fight and struggle, and it bites the predator. It was found that rats present specific biting patterns directed at cats, especially for male rats when the contact is accompanied by pain (Blanchard et al., 1980). This is not the main emotion and behavioural link in the present study, thus the pre-encounter and the post-encounter behaviour will be focused on.

Based on the above illustration, from the perspective of neurobehavioral science, living creatures would present different behaviour as different parts of their brain are activated when they are experiencing different threatening stages. This, again, provides strong evidence for the distinction between anxiety and fear. These will happen naturally in almost all animals, including humans. If we transform the example of the mouse into an example of a human, it could be as follows (CASE 2):

Background: a person who would always be frightened by spiders anticipates a potential encounter with a spider at home, perhaps they saw the spider quickly rushes under the table earlier in the day.

1) The person breathes deeply and wants to let the spider out of the room, so the person takes a broom and walks slowly and quietly to the table, carefully looks under the table.

2) Suddenly, the spider jumps out onto their left hand, then they freeze for a very short while.

3) After that, the person swings the left hand quickly and screams at the same time. The spider falls down on the ground and they use the broom to beat the spider recklessly and sweep it out of the room.

In CASE 2, the human's behaviour matches Fanselow's Three Modes of Defense. I.e. 'walk slowly and quietly' and 'carefully look' reflect the pre-encounter anxiety reactions (potential threat), 'freezes for a very short while' reflects the post-encounter fear response defense (actual threat), 'swings the hand quickly' and 'beat the spider recklessly' reflect the circa-strike (fight with the threat).

Further, the story can be adapted to the context of language learning (CASE 3).

Background: A group of Chinese freshmen sitting in a classroom waiting for their first English class to begin, and the English teacher walks into the classroom as the class will begin soon. Unexpectedly, the teacher announces: 'Please prepare for five minutes and later I will ask some of you to introduce yourself in English in front of the class'.

1) *Students start to 'panic' while knowing about this task. Some of them start to draw a draft to prepare their oral introduction, some of them try to discuss with their peers to make a strategy, and some of them even start to sigh and breathe deeply.*

2) *After around 5 minutes, as the teacher announces 'time up! Let me pick someone.' Some students lower their head and stop all of their actions, freezing on their chair. Some students avoid making eye contact with the teacher.*

3) *The teacher randomly picks Student A to do the oral introduction. Almost all of the other students take a sigh of relief when they know they are not picked. Student A moves to the platform with hesitation. Student A makes a short English speech with slight stammer, low voice and again, avoid having eye contact with peers and the teacher.*

4) *The teacher provides appreciation and praises to student A after the speaking finished. Student A heaves a sigh of relief and goes back to the seat quickly.*

The Three Modes of Defense could also be seen through CASE 3. Students are facing the predator named 'an unexpected foreign language public speech'. They present different behaviour when facing different stages of threat:

Pre-encounter reactions: drafts drawing, discussion, deep breath, head lowering, eye-contact avoidance, freezing. These behaviours are aim to assess and lower the risk brought by the potential danger. This is anxious behaviour.

Post-encounter defense: slight stammer, low voice, eye-contact avoidance. These behaviours are caused by a high degree of tension brought by the actual threat. This is fearful behaviour.

According to the above analysis, starting from animals' instinct to learners' foreign language anxiety/fear, the brain's activation-emotion-behaviour link can be seen. Therefore, we could advance the thesis that Foreign Language Anxiety arises pre-encounter with the potential risk before the threat truly happens; While the Foreign Language Fear arises with the actual risk during the time that the threat truly happens.

The present study intends to argue for Foreign Language Fear (FLF) as an independent conception. According to the illustrations above and referring to the definition of foreign language anxiety, the definition of the FLF is as follows:

FLF: a negative emotion which happens during the period of foreign language activities until they finish, and this emotion is usually triggered by *actual* evaluation by external sources (e.g. peers, scores or teachers). FLF activates in learners a set of physiological changes (e.g. fast pounding heart, sweating, muscles tensing), which stimulate learners to escape from, or put an end to, the ongoing activity as soon as possible.

According to other scholars, they label 'fear of negative evaluation in foreign language anxiety', 'fear of speaking' and 'communication apprehension' as a branch or indication of foreign language anxiety, however, all of these count as FLF based on the above illustration.

In contrast, FLA is defined by other scholars as follows:

'distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process' (Horwitz et al., 1986, p. 128).

or

'a situationspecific construct ((MacIntyre, 1999); (Teimouri et al., 2019), which suggests that people who do not normally feel anxious may be struck by anxiety in language classrooms.' (Yan & Liang, 2022), p. 2).

Based on my theory above, however, FLA should be defined as:

FLA: a negative emotion which happens *before* or/and *after* foreign language activities, and this emotion is usually triggered by *potential* or/and *unpredictable* evaluation by external sources (e.g. peers, scores or teachers). FLA activates in learners a set of physiological changes (e.g. fast pounding heart, sweating, muscles tensing), which stimulate learners to feel uncomfortable and worry about the upcoming challenge.

Many previous studies did not notice and discuss the distinction between fear and anxiety and it might be contended that the distinction between fear and anxiety is of little or no relevance to SLA studies. People may ask, 'why do we need to make a distinction between fear and anxiety?'. The main intention when SLA scholars study emotion is to explore the causes of any positive/negative emotions and the effects of them on the effectiveness of language learning. If, then, we do not understand a kind of emotion and distinguish it from other emotions, it could be impossible for us to overcome (mostly for negative emotion) or boost (mostly for positive emotion) the emotions under consideration. The above sections exposit the difference between fear and anxiety, especially against the background of foreign language learning. The difference could be obviously seen. Therefore, it is meaningful and necessary to separate them and adopt different strategies to eliminate them.

In conclusion, the fear discussed in this part explicates the situation that learners are willing to learn the specific foreign language, but they might feel frightened when facing answering questions, linguistic tests or public speeches, and this could be reduced to 'fear-of language learning'. However, some learners could be unwilling to learn, which is relatively close to 'hate' learning a foreign language, this might be caused by non-language learning reasons, such as 'I do not like my English teacher, I could not get along well with my teacher' (bad teacher-student relationship), 'I'm not willing to be cultivated and affected by foreign cultures, I do not like foreigners' (ethnocentrism) and 'I do not have to learn any foreign language, I think it is useless and bring no help for my future individual development' (lack of intrinsic motivation), etc. These could be reduced to 'fear-in language learning', which would be discussed briefly in the next section.

2.3.2.2 Fear-of and Fear-in Foreign Language Learning

At the end of the last section, examples have been provided to build a basic view of the difference between what I will call *'fear-in'* and *'fear-of'* language learning. Briefly, for 'fear-in', students hate/are not willing to learn a foreign language because of non-language related reasons, while for fear-of, students are afraid of using a foreign language because of language related reasons. Still, the rationale of emphasizing this difference is to offer a relatively clear guidance for language teachers and learners to know the actual reasons of being threatened by learning a foreign language. Thus, people could adopt corresponding strategies.

Moreover, it is possible that the 'fear-in' language learning could transform to the 'fear-of' language learning. Here is a typical example (CASE 4):

Student A' s first language is not English, and started to learn English from age 6 in primary school. Student A's first teacher taught usually with a higher speaking volume and was even very rude to the kids when they could not answer questions or performed unsatisfactorily in tasks/tests. This increased student A' s pressure every time involved in the English learning context, and, consequently, Student A became afraid of learning English and was not willing to attend the English class unconsciously. This situation continued, even though the teacher changed into a kind and professional one. Student A would naturally be stammering, speaking quietly, and avoiding eye-contact when doing oral performance and answering questions in English.

For explanation of the above case, here, the famous experiment conducted by Ivan Pavlov would be introduced. Add basic outline of IP experiment.

Pavlov showed that dogs could learn to associate the ringing of a bell with food through repeated experiments. Initially, the bell's sound was meaningless to the dogs, but after it was consistently paired with the sight and smell of food, they began to salivate merely at the sound, anticipating food even when none was provided.

1) the sound of ring is a neutral stimulus;

2) the food is an unconditioned stimulus for salivation;

3) the salivation is an unconditioned response for the food;

According to Pavlovian conditioning (classical conditioning), an unconditioned stimulus (US) refers to a trigger/stimulus leads to an automatic response. And the unconditional response (UR) refers to an automatic response that occurs without thought when an unconditioned stimulus is presented. For example, we will tremble involuntarily (UR) when we feel cold (US). Moreover, a neutral stimulus (NS) is a stimulus that doesn't initially lead to a response on its own, and a NS could become a conditional stimulus (CS). And, conditioned response (CR) is a learned response that is created where no response existed before. For instance, a person always see dogs (NS) on the street, but the person has no response, until one day the person is bit by a dog.

And after that happen, the person becomes afraid (CR) of dogs (CS) every time walking on this street.

Back to Pavlovian experiment, as the sound of a ringing bell (neutral stimulus) and the food (unconditioned stimulus) are always presented to the dog simultaneously, the salivation (unconditioned response) in dogs emerged with the ringing alone after a period of training. And this would not happen naturally before the experiment.

Now, the CASE 4 could be interpreted as follow:

1) the English learning is a neutral stimulus;

2) the rudeness is an unconditioned stimulus for the emotion of fear;

3) the emotion of fear is an unconditioned response for the rudeness.

As the English learning (neutral stimulus) and rudeness from the teacher (unconditioned stimulus) are always presented to student A simultaneously, the emotion of fear (unconditioned response) in student A emerged with the happen of English learning gradually. And this would not happen naturally before student A was taught by the initial teacher. This is the process of how the ‘fear-in’ becomes ‘fear-of’ in foreign language learning through Pavlovian conditioning.

When thinking about the interaction between two elements, it is natural to think and wonder if relationships work the opposite way, therefore, here comes a question: Will the ‘fear-of’ transform to ‘fear-in’ in foreign language learning? The answer is NO, the direction of transformation is single, which means that the ‘fear-of’ could not become ‘fear-in’. Basically, in the context of learning a foreign language, language itself cannot naturally become an unconditional stimulus to humans’ emotion of fear. However, instinctively, humans would be afraid of being blamed, evaluated, or laughed at. In brief, foreign language fear could occur, and learners do not have fear-of the language itself, but have fear-in the language classroom of being blamed, evaluated, or laughed at. And the fear-in non-language elements could transfers into fear-of language elements over time through a process of Pavlovian conditioning. But the fear-of language could not happen unconditionally.

Based on the conclusion above, the most effective method to eliminate learners' FLF seems to cut off it initially, which means to control the 'fear-in' emotion to a lower degree. Such as lower the existence of evaluation, hiring patient and professional teachers, and educating peers to be polite about others' performance, etc. In this way, there is no or low chance for the 'fear-in' becoming 'fear-of'.

In this part, two main negative emotion were introduced: *foreign language anxiety* and *foreign language fear*. The viewpoint of the distinction between FLA and FLF was advanced and evidence were provided from the perspectives of the emotion conceptions from APA's Dictionary of Psychology, the Three-Dimensional Taxonomy of Achievement Emotions and the Predatory Imminence Theory. Moreover, the importance and practicability of the distinction were offered. In contrast, scholars concentrate more on the conceptions in recent years, causes and effects of positive emotions in second language acquisition, therefore, the foreign language enjoyment would be introduced in the next part.

2.3.3 The Concept of Foreign Language Enjoyment (FLE)

With the booming of positive education (PE) in recent years, Positive Psychology (PP), which could contribute to guide people to success (Csikszentmihalyi & Nakamura, 2011), gradually started to be explored in SLA. In the beginning, positive concepts were introduced into educational psychology by (Csikszentmihalyi & Nakamura, 2011), while they claimed the importance to make a balance in the learning context with considering hope, mindfulness, well-being, communicative skills, etc. Later, FLE was defined: in language learning, learners' attempt to face challenges and extend the proficiency and knowledge in the class (MacIntyre, 2016). Broadly speaking, previous studies relate to FLE are mainly focused on four categories:

1) Validity of the FLE Measurements. The building block of FLE was set by (Dewaele & MacIntyre, 2014), and the study investigated the correlation between FLA and FLE. A researcher-made questionnaire was released online to 1746 language learners from different countries and regions learning English, Chinese, Spanish, German and Dutch as foreign language. The measure scale contains 29 items, in which

21 items aim to observe positive emotions regarding the teacher, peers, and the learning experience, while the rest of 8 items was took from the FLCA. Afterwards, (Dewaele & MacIntyre, 2016) modified and improved the FLE scale into a 14-item questionnaire, which was frequently applied by later related FLE research. Moreover, (Jin & Zhang, 2021) reported a revised measurement named English Classroom Enjoyment Scale (ECES), including Enjoyment of Teacher Support, Enjoyment of Student Support and Enjoyment of Foreign Language Learning. They assert that it is a more reliable and efficient scale.

2) The Connection Between Demographic Variables and FLE. For instance, it was claimed that younger learners (such as high-school students) showed lower degree of FLE compared with older ones (such as university students) (Dewaele & Alfawzan, 2018; Dewaele & Dewaele, 2018; Dewaele et al., 2018). For gender, the results were inconsistent, some studies showed that female language learners have demonstrated higher FLE (Dewaele & MacIntyre, 2016), while others said the difference is insignificant (Alenezi, 2020; Mierzwa, 2018).

3) The Connection Between FLA and FLE. Some previous studies showed a negative relationship between these two conceptions (Dewaele & Alfawzan, 2018; Resnik & Dewaele, 2020). However, it is not reasonable to put FLE and FLA on opposite sides, as the feeling of enjoyment and anxiety could exist simultaneously. For FLA, willingness to communicate (WTC) is one of the research variables, on this point, recently scholars have begun to argue that language teachers are supposed to create a positive classroom atmosphere and to promote WTC, in turn the language learning enjoyment could be expand (Elahi Shirvan & Talebzadeh, 2018).

4) FLE as a Complex and Dynamic Structure. Some studies indicate that FLE is a dynamic concept, and scholars even measure FLE by a dynamic approach, participants reported that all of their FLE were growing over time (Dewaele & Dewaele, 2017), similarly, learners' FLE could also fluctuate from topic to topic (Elahi Shirvan & Talebzadeh, 2018). Additionally, a teacher's using of target language and learners' positive perception of the teacher could bring promotion of FLE in a language learning

classroom (Dewaele & Dewaele, 2017). Based on the the above, FLE is a relatively new and not fully-developed concept. Some directions which could be considered and discussed about FLE are: the association between FLE and FLA, the utilization of FLE for greater language performance and acquisition, the method of improving FLE during and after language learning, etc,. Thus, in my research, except for exploring what effects will TMTBLT bring to L2 oral performance, the association between FLE and TMTBLT will also be involved.

2.4 Language Performance and Language Competence

Early in 1960s, Chomsky advanced the description for terms of competence and performance. The latest scholarship relates to these two concepts focusing on 1) the nature of language competence; 2) the relation between competence and performance, and the actual roles they played in linguistic description and analysis, respectively. Generally, competence is regarded as the 'speaker-hearer's knowledge of the language', while the performance is 'the actual use of language in concrete situations' (Brown et al., 1996).

The reason of mentioning the distinguish between performance and competence here aims to clear that in this study, learners' speaking performance, rather than competence, will be measured, as the objective of the research is to see the instant reflection (performance) under the established learning context (TMTBLT & simple-complex task condition). Before the main test, a pilot test will be done to ensure the validity and effectiveness of the test.

In conclusion, my study will be built on the three-way interface, that is ***emotion-task complexity-L2 performance***, in which 'emotion' refers to FLE and FLF, task complexity refers to tasks' cognitive demand for students, and L2 performance will be measured through participants' language complexity, accuracy and fluency. As a learning context, TMTBLT will also be explored in order to offer support for its superiority for assisting learners to greater performance and acquire language knowledge effectively.

The theoretical aims of the thesis were undertaken in Chapter 2 above. The experimental aims were outlined in Chapter 1 and in the next chapter, for methodology, the methods and proceedures for exploring the empirical research questions will be explored independently.

CHAPTER 3
METHODOLOGY

As Chapter 2 systematically provided all the dependent theories as theoretical supports for the present thesis, further, in Chapter 3, the practical process and corresponding rationale will be included. In what follows, the method employed in the present thesis will be explained in detail. This includes participants, instruments and materials, main procedures, coding and scoring, and finally, data analyses. Generally, before the main experiments, there will be a pilot study to ensure the practicality and effectiveness of the later.

3.1 The Main Experiment

In the present study, a pilot study was conducted prior to the main one. This section includes the relevant details of the main study, covering participants, tasks, electronic and non-electronic equipment, the steps involved in conducting the study, and the methods used for analyzing the collected data.

3.1.1 Participants

Sixty L1 Chinese learners of English were selected for the study on a voluntary basis. The participants were homogeneous in terms of their nationality, age, and L1, and they were undergraduate students from a university in China. An email was sent to the students' inboxes, requiring them to complete an attached form (to collect their personal background information) to ensure that they were voluntarily willing and suitable to participate in the experiments. After receiving the students' responses, 60 participants were randomly chosen to take part in the study. These participants were then divided into three groups, with 20 students in each group.

Additionally, all participants were required to have scored 5 points in the IELTS speaking component to ensure they could handle and manage the tasks provided in the experiments. This standard was adopted because an IELTS speaking band of 5 typically indicates (according to official IELTS descriptors) that learners:

1.Maintain some flow of speech using basic vocabulary and grammar, though more complex communication might cause fluency issues.

2.Attempt paraphrasing with mixed success.

3.Use simple structures with reasonable accuracy but may struggle with more advanced forms.

4.Show a generally understandable accent with some lapses.

We thus set a band 5 threshold to ensure participants had sufficient English proficiency to understand instructions and manage basic dialogue in the TMTBLT tasks, while still being relatively intermediate-level speakers, so that any performance changes were not obscured by extremely high or low proficiency.

However, the 60 participants were mixed across the three groups, meaning there was no control group based on English language proficiency. Since the study aimed to explore the effects of different levels of cognitive task complexity on learners' performance and emotions rather than test the effectiveness of TMTBLT, there was no specific target language in the experiments. Participants were required to complete three tasks of varying levels of cognitive complexity in different orders, with the detailed procedures and rationale for this approach introduced and explained in Section 3.1.2.1.

3.1.2 Materials

This section provides a description of 1) the three TMTBLT tasks; 2) the FLF questionnaire and the FLE questionnaire; 3) the camera monitors; 4) the performance measuring scale.

3.1.2.1 The Three TMTBLT Tasks

In the main study, experiments were conducted with three groups separately. Each group completed three distinct tasks at different levels of complexity in varying orders. This approach ensured the balance and accuracy of the results, as it was anticipated that the sequence of tasks with differing levels of complexity might influence learners' emotions and performance. Robinson's SSARC model also highlights the importance of sequencing tasks and demonstrates the potential effects of task order on performance. By using groups with varied task sequences, the study was able to provide more comprehensive insights. These details are further elaborated in the following sections.

Technology-Mediated vs. Technology-Enhanced

As discussed in Chapter 2, Technology-Mediated tasks differ from merely Technology-Enhanced tasks in that technology is integrally woven into the communicative context rather than merely being an online platform for quizzes or tests. In this study, the “police reporting” interactions occur entirely through an interactive computer program simulating a real-life emergency call scenario. The tasks require participants to speak in English to the system, which “responds” with questions from a police operator or officer.

This is more than just a “technology-enhanced quiz”, because the entire structure of the task (observing a CCTV video, calling the police, giving a statement) relies on the digital medium for authentic language use.

The tasks simulate a real-life event—reporting a crime by phone—thus blending technology with genuine language communication as opposed to a generic online test. Learners are not simply ticking boxes or completing text-based quizzes; they are actively producing oral language in a scenario that mirrors daily life (e.g., phoning law enforcement).

The main study involved three tasks: a simple task (S), a middle task (M), and a complex task (C). As one of the research aims was to explore the relationship between cognitive task complexity and performance, task complexity served as the independent variable, while performance was the dependent variable. Cognitive complexity was manipulated in accordance with CH (cognitive hypothesis) from the dimension of resource-directing variables, using +/- elements in the corresponding tasks. There were two reasons for establishing three levels of complexity:

(1) From an objective perspective, this design avoided a potential issue: as task complexity increased gradually, performance—such as lexical complexity—might initially improve, reach a peak at an optimal level of complexity, and then begin to decline. If only two levels of complexity (low and high) were included,

without a middle level, it is possible that both levels could coincidentally correspond to the same or similar level of performance. As a result, the complete trend of performance changes under increasing task complexity might have been obscured. Therefore, a middle level of complexity was included to capture this potential variation and provide a more accurate depiction of the relationship between task complexity/task sequence and performance.

(2) Another reason for including three levels of complexity in varying sequences was to address potential subjective effects on participants. If participants were to complete the tasks consecutively in the same sequence (e.g., S-M-C) and within the same task style, the influence of task complexity on both linguistic performance and emotional state might diminish over time. In other words, participants could gradually adapt to the tasks, which might affect the accuracy of the results and lead to reduced emotional reactions.

To mitigate this issue, the 60 participants were divided into three groups of 20, with each group completing the three levels of task complexity in a different order. For example:

Group 1, S-M-C;
Group 2, C-M-S;
Group 3, M-C-S.

The following diagram shows the information about the main study:

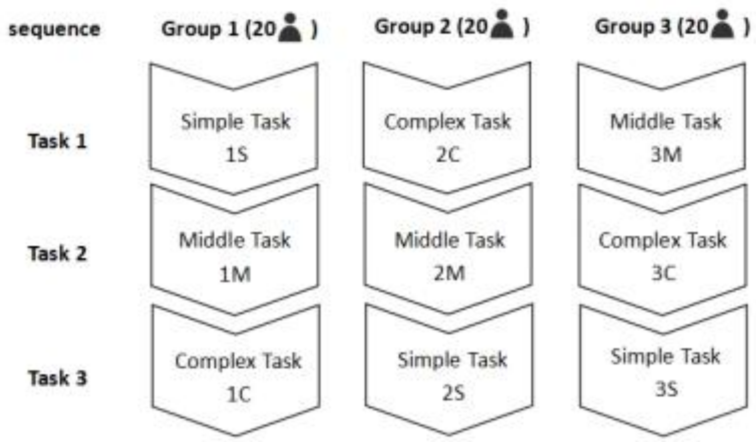

Figure 1 The Diagram for the Design of the Study

According to this approach, the results (performance and emotion) could be compared both vertically (e.g., between 1S, 1M, and 1C) to determine whether the order of task completion had any noticeable effect on performance and emotional reactions, and horizontally (e.g., between 1S, 2S, and 3S) to assess the impact of task complexity on performance and emotional responses. Conducting the experiment in this manner provided insights into whether there was a statistically significant relationship between task order and performance. Additionally, it allowed for the detection of any potential biases in the data that might have arisen from emotional or intellectual desensitization to the tasks due to their order or repetition. Other permutations (e.g., S ⟶ C ⟶ M) might yield further nuance, but the current study's scope was primarily to check if changing the order had an observable effect and to avoid potential fatigue, desensitization, or trivial repetition. These three sequences offer enough variation to identify whether sequence matters without overwhelming participants or complicating analyses excessively.

The Task

The following description is the theme and framework of the TMTBLT task that was applied in the experiment.

The theme of the task was reporting a criminal case, where participants played the role of witnesses (speakers), and the computer simulated the roles of the police operator and police officer (listeners). Before the interactive conversation section, a CCTV video was shown to the participants on a screen. The video depicted a criminal case and included the following elements: characters (who), locations (where), time (when), and events (what).

The video featured various characters, such as offenders, victims, and bystanders, each distinguishable by their gender, hairstyle, and clothing. Additionally, details of the surroundings, objects, road signs (indicating the location's details and positioning), and a clock (displaying the current time) were also included. For example, the criminal case shown in the video might involve a robbery. Participants observed the case's progression and were then prompted by the computer system to call the police and report the incident.

Why Criminal-Case Reporting Tasks?

Authentic, Real-Life Theme: Phoning the police to report a crime is a plausible everyday scenario—particularly relevant to adult learners. This real-life connection promotes the TBLT principle of meaningful language use.

Engaging and Motivating: The "crime" context can be more stimulating than routine topics, prompting participants to pay attention and produce language under mild tension.

Suitability for TMTBLT: The technology-based simulation (CCTV videos + virtual police calls) effectively integrates real-world tasks with digital tools, reinforcing the idea that language is used to accomplish tasks in an authentic environment.

The computer, acting as the police operator, followed a predefined pattern of questions to gather information from the participants. These questions covered essential details about the incident.

1) This is the Police Station, how can I help you?

2) Okay. May I have your name please?

3) Where did the case happen?

4) May I have your phone number please?

5) Okay. Please calm down, we will arrive there within 10 minutes.

The participants were required to answer these questions based on the information provided in the video shown beforehand. After this initial conversation, another interaction took place between the participant and the police officer who arrived at the scene. The questions asked by the police officer followed a predefined pattern, as outlined below.

1) Hi, did you call the police just now?

2) OK. Thanks for your report. What happened?

3) Can you describe the appearance of the offender?

4) Did you see where the offender escaped to?

5) Is there anything else you want to tell us?

6) Thanks for your cooperation, we will inform you if we catch the offender.

The cognitive complexity in the study was manipulated through the addition or reduction of elements (+/- elements) across three tasks of varying difficulty: simple (S), middle (M), and complex (C). This manipulation was achieved by altering the amount and nature of details in the scenarios, particularly along the dimensions of Who, When, Where, and What. Each task presented participants with a criminal case to narrate, with increasing complexity as the tasks progressed.

In the simple task (S), the scenario was straightforward, involving minimal elements. For example, a slightly overweight woman with dark skin, dressed in a pink top and black pants, stole a new television in its box from an unattended grocery store during the day. The task required participants to describe only a limited number of characters and straightforward actions.

The middle task (M) introduced additional complexity through more intricate details and sequential actions. For instance, the scenario involved a middle-aged man in an orange jacket and dark pants. Late at night, he drove an engineering vehicle to a bicycle shop at a crossroads, shattered the storefront window, tied two new bicycles to his vehicle with a rope, and drove off. This task required participants to manage more information, including timing, location, and a sequence of actions.

The complex task (C) presented the highest level of challenge, involving multiple characters and a series of events. The scenario involved a motorcycle being knocked over by a speeding white car. Three young men, dressed differently, exited the vehicle, approached the fallen motorcyclist, and took valuable items like a wallet and phone after the motorcyclist fled. They then took the motorcycle and left the scene. These complex tasks required participants to describe intricate interactions, multiple actions, and detailed events.

Concrete Details and +/- Elements

Three different tasks were created at simple (S), middle (M), and complex (C) levels, primarily manipulating "+/- elements" along four key dimensions: Who, When, Where, and What. This approach follows Robinson's (2001a, 2005) resource-directing dimension, where complexity is raised by adding or removing details (elements) that the learner must attend to when reporting. Below is a concise summary adapted from Appendix 3, showing how complexity grows:

Simple Task:

Who: One thief (a middle-aged woman with dark skin, pink sleeveless top, black over-knee pants, pink face mask, carrying a bag).

Where: An unattended store with general goods and appliances.

When: Daytime.

What: She stole a TV in its box. Pedestrians passed by but did not notice.

(Fewer details, fewer characters, straightforward location/time/event.)

Middle Task:

Who: One tall, slim man (black-and-orange jacket, black trousers, knitted hat).

Where: A motorcycle/bicycle shop.

When: Late at night.

What: He used a construction vehicle to smash the display window, tied two green bicycles to the mechanical arm with a rope, and drove away. The street was deserted.

(More details, more sequential actions, more elements to track.)

Complex Task:

Who: Multiple people – a motorcycle driver (black helmet, green jacket) plus three men from a white car (different T-shirts and baseball caps).

Where: A residential road with greenery on both sides.

When: Daytime.

What: The white car hit the motorcycle. The motorcycle driver fled, leaving a wallet and phone. The three men took valuables and pushed the fallen motorcycle away.

(Multiple characters, multiple interactions, more complicated timeline.)

Each participant watched a short CCTV-style video for the assigned scenario, then was asked to “report the incident” to a simulated police operator and later to a police officer. This required describing the who/where/when/what details as accurately as possible. Increasing the number of details, characters, and sequences of events systematically raised the cognitive complexity of these tasks.

This structured variation in task complexity enabled a comprehensive examination of the relationship between cognitive task complexity and participants’ performance and emotional responses.

To facilitate understanding, the following illustrations include a photo of one participant during the experiment and screenshots of the task as displayed on the computer.

One participant testing the procedure before the task began (left) and

Screenshot of the simple task video (right)

Instructions from the task procedure (left) and The interface for conversations

with the police operator (right).

After completing the two conversations, the computer system recorded the participants' responses via audio and transcribed the recordings into text for subsequent analysis. In the current study, participants did not receive feedback after finishing the tasks. However, if similar tasks were implemented in real language classes, providing feedback would be recommended to help learners understand how to improve their performance.

For example, the App system could analyze the participants' audio recordings, comparing their responses to the correct information provided in advance by the system. If the participants achieved over 80% accuracy in their spoken output, the system could notify them that the offender was successfully caught, signaling success. Conversely, if their accuracy was below 80%, the system could indicate that the offender escaped, marking failure. Additionally, the system could present a detailed list of incorrect responses alongside the correct answers for participants' reference.

To further enhance engagement and motivation, gamified elements could be incorporated into the App, offering positive incentives for learners. These elements, such as achievements, points, and ranking lists, have been shown to boost learner engagement (Landers et al., 2017). Specifically, the following gamified features could be applied:

Ranking Lists: Participants' accuracy scores could be integrated into a ranking system, allowing them to compete with peers under the theme, "Who is the best assistant for the police officer?" The ranking list would be updated weekly to maintain engagement and flexibility.

Points: Participants could track their progress over time, comparing their own performance across extended practice sessions. Points could be awarded for each practice session or improvement milestone, which could then be used to unlock additional in-App tools or features as bonuses.

Achievements: Strong performance could be directly linked to academic rewards.

Such gamified features would make the tasks more interactive and enjoyable, encouraging sustained engagement and fostering continuous improvement in learners' oral expression skills.

In conclusion, the concrete steps for conducting the main study are as follows:

Sixty participants were gathered in classrooms equipped with computers preloaded with the task programs. They were divided into three groups, with

20 participants in each. Group 1 completed the tasks in the order of ‘S-M-C’, Group 2 in the order of ‘C-M-S’, and Group 3 in the order of ‘M-C-S’.

After completing each task, participants were given questionnaires on Foreign Language Fear (FLF) and Foreign Language Enjoyment (FLE) to collect their emotional responses to the current task. Simultaneously, the camera in each room recorded the participants’ behaviors and facial expressions, allowing for an observation of their emotional changes during the tasks.

All the above equipment and data collection methods ensured comprehensive data for subsequent analysis, providing both subjective and objective insights into the participants’ emotional and linguistic performance.

Before the main study, a pilot study was conducted to test the practicability of the tasks and to ensure that the levels of cognitive complexity aligned with expectations. Ten participants were enrolled in the pilot study. They were asked to complete the three tasks, each representing a different level of complexity, in a random order. After completing each task, the FLE and FLF questionnaires were administered to collect data on their emotional responses. Upon finishing all tasks, the participants were required to rank the complexity of the three tasks subjectively. This feedback helped refine the task complexity design for the main experiment.

Additionally, the ten participants in the pilot study were interviewed with a set of brief questions to gather insights into their experiences with the tasks. This feedback helped determine whether the designed TMTBLT tasks and the experimental procedures were clear and understandable for the participants.

In the present study, the primary aim was to analyze the participants’ performance rather than provide feedback on their performance. The detailed methods for measuring and analyzing this production are introduced in Section 3.1.2.3.

3.1.2.2 The FLE and The FLF Scales

After completing the tasks, participants were given two questionnaires to assess their emotional responses, focusing on both enjoyment and fear.

The FLE (Foreign Language Enjoyment) scale was adapted from the questionnaire developed by (Dewaele & MacIntyre, 2016). While the original questionnaire consisted of 21 items on a 5-point Likert scale (strongly disagree, disagree, undecided, agree, strongly agree), a modified version tailored to the current study was used. The new FLE questionnaire contained 6 items designed to assess positive feelings related to the context (TMTBLT), the format (game-like elements), and the theme (task content). It retained the same as the original.

On the other hand, as discussed in Chapter 2, the concept of FLF (Foreign Language Fear) was clarified, and the distinctions between FLF and FLA were explained. For this study, the FLF questionnaire was inspired by Horwitz's FLCAS (Foreign Language Classroom Anxiety Scale) from 1986. However, it was adjusted to fit the study's focus: some items were removed, and new ones were added. The FLF questionnaire also included 6 items, addressing negative feelings arising from the context (TMTBLT), the required language skill (speaking), and the task theme. Like the FLE scale, it employed a 5-point Likert scale (strongly disagree to strongly agree).

Changes Made When Adapting FLF and FLE Questionnaires

Reduced item count: From Horwitz's 33 items and Dewaele & MacIntyre's 21 items to 6 core statements each, closely aligned with the short TMTBLT tasks.

Terminology: Substituted "negative emotion" with "discomfort" in the FLF items, focusing on short-term threat/fear during the tasks rather than long-term classroom anxiety.

Contextual modifications: Removed references to formal classrooms or teacher-led exams, emphasizing the "speaking to a police operator/officer" scenario.

Likert scale: Retained "strongly disagree" to "strongly agree", consistent with the original, but reworded statements to match the immediate experience of reporting a case in English.

This tailoring ensures the scales measure the immediate enjoyment or discomfort linked to the TMTBLT tasks, rather than broad, semester-long sentiments.

Open-Ended Questions

In addition to the 6 Likert items, each questionnaire included one open-ended prompt, such as:

"Please describe any other positive feelings you had during the task" (for FLE), or

"Please describe any other feelings of fear or discomfort you experienced" (for FLF).

Participants could optionally write a few lines explaining why they felt a certain way. This qualitative data provided insights that pure Likert items might miss—e.g., specific triggers or personal anecdotes about enjoyment/fear.

These two scales were administered after each task (S, M, and C) to capture participants' emotional responses in real-time before the next scenario.

Both questionnaires are provided in the Appendix.

3.1.2.3 The Performance Measuring Scale

As introduced in the previous section, and for comparability with prior TBLT literature (Donate, 2018; Kormos, 2011; Sasayama, 2016), participants' task performance in the current study was assessed using four dimensions: complexity (including both syntactic and lexical aspects), accuracy, and fluency (CAF). The CAF framework is a widely recognized and reliable method for evaluating learners' language performance, forming a critical area of research in ISLA (Instructed Second Language Acquisition).

(Skehan, 1989) expanded earlier methods of measurement, which primarily focused on accuracy and fluency, by incorporating complexity. This innovation established the triad of complexity, accuracy, and fluency (CAF) as the three fundamental dimensions for characterizing L2 performance (Housen & Kuiken, 2009). The following paragraphs detail the specific sub-dimensions within each of these categories.

Linguistic Complexity, as a multidimensional construct, is defined as "the number of discrete components that a language feature or a language system consists of, and the number of connections between the different components" ((Bulté &

Housen, 2012), p. 24). This conceptualization establishes a foundational distinction between grammatical complexity and lexical complexity. Scholars have further attempted to disentangle the subcategories within these two dimensions and identify appropriate measures (Bulté & Housen, 2012; Jarvis, 2013; Jarvis & Daller, 2013; Malvern & Richards, 1997; Vermeer, 2000).

Distinction Between CAF and CALF

While many studies on L2 performance have not explicitly distinguished between CAF (Complexity, Accuracy, Fluency) and CALF (Complexity, Accuracy, Lexical Complexity, Fluency), the two frameworks largely overlap. Complexity, for instance, encompasses grammatical/syntactic complexity and lexical complexity, making the inclusion of "lexical complexity" in CALF a subcomponent rather than a distinct dimension. Therefore, the underlying structure of these two methods remains consistent, emphasizing the same fundamental constructs: complexity, accuracy, and fluency.

Based on the above introduction, numerous sub-dimensions can be used to assess L2 learners' language production. However, their applicability varies depending on the nature of the task (e.g., written or oral performance). Consequently, the present study selects the most suitable subcategories as concrete measurement items.

The following table (Figure.2) presents the CAF measures used in this study to evaluate participants' oral task performance. It includes all the categories that will be analyzed based on participants' spoken output. Notably, this study does not prescribe a target language proficiency level and does not evaluate participants' performance as beginner, intermediate, or advanced. Instead, the analysis focuses on comparisons between the three experimental groups to explore relative differences.

This approach ensures that the selected CAF measures align closely with the study's objectives while maintaining the flexibility to capture a wide range of performance outcomes.

Participants' spoken output was evaluated using Complexity, Accuracy, and Fluency (CAF) plus lexical complexity, i.e., CALF. Importantly, there is no absolute benchmark or external linguistic standard in this study. We do not classify performance as "good" or "bad" against native norms; instead, we use CALF metrics comparatively to see differences in performance across the three tasks (S, M, C) and across different groups. The study's focus is on relative variations, not on an absolute proficiency scale.

Complexity

Syntactic Complexity: Mean length of AS-unit (MLAS) and mean length of clause (MLC).

Lexical Complexity: Type-token ratio (TTR).

(Lexical diversity, sophistication, and density can be explored, but TTR remains the principal measure for easy comparability.)

Accuracy

Error-Free Clause Ratio (EFCR), capturing the overall percentage of error-free clauses in a given sample.

Fluency

Productivity: Total number of words (TNW) or syllables (TNS) excluding false starts, repairs, and reformulations.

Speech Rate: Words per minute (WPM) or syllables per second (SPS).

(Handling Pauses, Repairs, and Self-Corrections)

As shown in Figure 2, silent pauses, false starts, and self-corrections were not included in the final word or syllable counts. Specifically:

Silent pauses over 0.3 seconds were tagged as breaks but not counted as words/syllables.

False starts (repeating the first few words of a sentence without completing it) were trimmed.

Self-corrections (e.g., “He… he was wearing a black… no, it was orange jacket”) were also excluded from the final token count, though the corrected version was counted if it was completed.

This ensures that the measured “fluency” reflects actual delivered speech minus obvious disfluencies. Further details on tagging were handled by software that transcribed the recordings and flagged disfluencies. Annotations were reviewed manually for accuracy.

No Fixed Linguistic “Standard”

In line with the study’s comparative goal, these CALF measures are used only to compare performance across the different tasks or groups. They do not denote an absolute standard of “advanced” or “beginner.” Hence, a participant’s performance can only be judged relative to their production on another task or to the group average, not against an external norm.

Category	Subcategory	Measure
Complexity	Syntactic complexity	Mean length of AS-unit (MLAS)
		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)
		Total number of syllables (excluding repetitions, repairs, reformulations and false-starts) (TNS)
	Speech rate	Number of words per minutes (WPM)
		Number of syllables per second (SPS)

Figure 2 The CALF Measurement for the Present Study

With the assistance of CAF measurement, the speaking performance of the pilot study will be shown, and the related data could provide a guidance for improving the experiment in the main study.

3.1.3 Procedure

Summarizing the entire study procedure:

1.Pilot Study:

10 participants each completed the 3 tasks (S, M, C) in random orders, then filled out the FLF/FLE scales.

Role of External Experts:

- A small panel of TBLT specialists and language assessment experts reviewed the pilot's audio/video data, the tasks themselves, and participant feedback.
- They confirmed that the three tasks indeed represented simple, middle, and complex levels as intended.
- They also checked if participants' responses indicated that complexity differences were noticeable.

Expert & Participant Feedback led to further adjustments:

- All found the S, M, C tasks suitably matched the intended levels of difficulty (no major scenario changes).
- Added an on-screen text-based "Important Notes" before tasks begin, reminding participants about timing, clarity, and that the session was recorded.
- Converted questionnaires to QR code format so that participants could quickly scan and fill them immediately after each task, maintaining continuity and saving time.
- Replaced "negative emotion" wording in FLF items with "discomfort."

Consequently, the final tasks and instruments were judged to be coherent, feasible, and aligned with the study objectives.

2. Main Study:

- 60 participants were split into 3 groups (n=20 each).
- Group 1 did tasks in order S ⟶ M ⟶ C.
- Group 2 did tasks in order C ⟶ M ⟶ S.

- Group 3 did tasks in order M ⟶ C ⟶ S.

After each task, participants completed the FLF and FLE questionnaires (including open-ended responses). They were also videotaped to capture behavioral cues. Their spoken interactions with the computer-based "police" were recorded, transcribed, and prepared for CALF analysis.

3. Post-Processing:

- Recorded audio was checked for clarity.
- Transcripts were prepared, removing silent pauses, false starts, and repeated words before final word/syllable counts.
- CALF metrics were computed for each participant's performance on each task.

3.1.4 Data Analysis

After scoring and coding all responses, statistical analyses were performed to address the research questions (whether cognitive task complexity and its sequence affect performance and emotion).

Analysis of Variance (ANOVA) was employed because:

We have more than two groups or conditions to compare. In many cases, we compare the three levels of task complexity (S, M, C) or the three group sequences (S-M-C, C-M-S, M-C-S). ANOVA efficiently tests the differences in mean scores (e.g., lexical TTR, error-free clause ratio, fear/enjoyment ratings) across three or more groups simultaneously.

ANOVA controls the Type I error rate, avoiding multiple t-tests that inflate the chance of false positives.

A one-way ANOVA was used in simpler comparisons (e.g., performance differences across tasks S vs. M vs. C), while a two-way ANOVA could test both "task complexity" and "task sequence order" effects. After the ANOVA, post hoc tests (e.g., Tukey) pinpoint which specific groups differed if the overall F-test was significant.

Although the final results in Chapter 4 emphasize p-values (e.g., $p < .05$), we also interpret effect sizes to gauge the magnitude of the difference. Descriptive statistics (means and standard deviations) offer further clarity on group differences.

This chapter outlined the participants, instruments, tasks, and procedures used in the main study. We required an IELTS speaking band of 5 to ensure all participants had sufficient English proficiency to handle interactive speaking tasks. Task complexity was carefully manipulated by adding or removing scenario elements across three tasks (S, M, C). Participants completed these tasks in three different orders, which allowed analysis of both complexity and sequence effects on performance and emotion. CALF measures (complexity, accuracy, and fluency, including lexical complexity) were used to compare participants' oral output across tasks and groups, but not against an external linguistic norm. Learners' emotions were captured via FLF and FLE scales—both adapted with fewer items and revised phrasing ("discomfort" rather than "negative emotion"), plus open-ended questions. External TBLT experts helped validate the final design during the pilot, confirming that the tasks and questionnaires matched the intended difficulty levels and content. Finally, ANOVA was employed as the principal statistical method to compare multiple group means while controlling for error rates.

Chapter 4 will present the quantitative and qualitative results of these analyses, including p-values, descriptive statistics, effect sizes, and thematic insights from open-ended responses, illustrating how task complexity, sequence, and emotional factors interrelate in TMTBLT contexts.

CHAPTER 4
RESULTS AND DATA ANALYSIS

Following comprehensive planning, the study's one-day experimental session involved each participant arriving to complete three speaking tasks, during which speech data was recorded and questionnaires were administered. By the end of the session, 360 questionnaires were successfully gathered, comprising 180 assessing Foreign Language Fear (FLF) and 180 focusing on positive emotions (FLE). Each participant filled out a total of six questionnaires—two for each task—so that both fear-related and enjoyment-related responses were captured. This design ensured a balanced yet streamlined data-collection process, integrating objective speech analysis with subjective emotional feedback.

For the speaking recordings, audio data were collected from all 60 participants. The recordings were analyzed based on metrics such as type-token ratio to measure lexical diversity, mean length of AS-units and clauses to assess syntactic and grammatical complexity, and overall verbosity indicated by the total number of words and syllables. Additionally, speech fluency was examined through words per minute and syllables per second, while linguistic accuracy was assessed using the ratio of error-free clauses. These measures provide a comprehensive framework for analyzing participants' oral performance with precision and rigor.

In the following sections, the data will be presented in tabular format, accompanied by a comparative analysis of each group's performance across the simple-middle, middle-complex, and simple-complex levels of cognitive task complexity (CTC). To investigate the potential relationship between CTC levels and L2 performance, statistical measures such as the p-value will be calculated, with this relationship referred to as the ***task-level effect***.

Furthermore, comparisons will be conducted across tasks S1-S2-S3, M1-M2-M3, and C1-C2-C3, representing the same CTC levels performed by Group 1, Group 2, and Group 3, respectively. Here, the p-value will also be calculated to examine the potential relationship between task sequence and L2 performance, referred to as the

task-order effect. These analyses aim to provide a nuanced understanding of how task complexity and sequencing influence participants' language performance.

Similarly, the questionnaires will undergo detailed analysis. For the FLF emotion questionnaire, comparisons will be conducted across tasks S1-M1-C1, S2-M2-C2, and S3-M3-C3 to assess the task-level effect and explore the potential relationship between cognitive task complexity (CTC) levels and FLF emotions. Additionally, comparisons among S1-S2-S3, M1-M2-M3, and C1-C2-C3 will be made to examine the task-order effect. The fluctuation of participants' FLF emotions before, during, and after each task will also be analyzed, with a line graph presented to visually represent the dynamic changes in FLF emotional states for each task. This visual depiction will provide insights into the differences between Foreign Language Anxiety (FLA) and Foreign Language Fear (FLF), as previously discussed.

For the positive emotion (enjoyment) questionnaire, both the task-level and task-order effects will also be evaluated, allowing for a comprehensive understanding of how CTC levels and task sequencing influence participants' emotional responses.

4.1 Dynamic Changes in Speech Data-Along CTC levels

This section presents the results of the CALF dimensions for this experiment, alongside the corresponding p-values, emphasizing the dynamic changes observed in response to varying cognitive task complexities (CTC). These findings provide a comprehensive understanding of how CTC levels influence L2 performance. The calculated p-values and visualized charts are explained concisely, highlighting their statistical significance and revealing potential relationships between task complexity and language performance, while laying the groundwork for further detailed analysis. Note that this section focuses solely on the relationship between CALF and CTC, whereas Section 4.2 ("Dynamic Changes in Speech Data Along Task Sequence") will later discuss the impact of task sequence on performance.

4.1.1 Lexical Complexity-Type Token Ratio (TTR)

Type-Token Ratio (TTR) serves as a key indicator of lexical variety within a text or spoken language sample. It is calculated by dividing the number of distinct words (types) by the total number of words (tokens) in the sample (Richards & Schmidt, 2013). A higher TTR reflects greater lexical variety, signaling more diverse vocabulary usage, while a lower TTR suggests increased word repetition. TTR is widely used in linguistic research to assess vocabulary richness and the degree of linguistic variation in both written and spoken outputs.

The following table presents the TTR results for all three groups across the three levels of Cognitive Task Complexity (CTC), providing insights into the impact of task complexity on lexical variety.

TTR FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	0.768	0.670	0.509	P1	0.5	0.597	0.667	P1	0.619	0.679	0.708
P2	0.684	0.750	0.544	P2	0.625	0.455	0.638	P2	0.546	0.662	0.513
P3	0.697	0.700	0.712	P3	0.616	0.624	0.623	P3	0.667	0.677	0.607
P4	0.676	0.815	0.518	P4	0.81	0.819	0.842	P4	0.59	0.635	0.714
P5	0.771	0.6	0.627	P5	0.635	0.64	0.681	P5	0.514	0.53	0.584
P6	0.588	0.63	0.550	P6	0.5	0.55	0.741	P6	0.558	0.619	0.645
P7	0.465	0.71	0.603	P7	0.64	0.76	0.695	P7	0.511	0.567	0.587
P8	0.875	0.48	0.627	P8	0.55	0.77	0.580	P8	0.613	0.511	0.6
P9	0.613	0.578	0.465	P9	0.565	0.57	0.524	P9	0.652	0.694	0.826
P10	0.804	0.635	0.488	P10	0.612	0.63	0.516	P10	0.45	0.438	0.591
P11	0.840	0.76	0.59	P11	0.563	0.57	0.638	P11	0.493	0.447	0.459
P12	0.692	0.57	0.463	P12	0.661	0.67	0.698	P12	0.505	0.589	0.556
P13	0.895	0.66	0.565	P13	0.496	0.6	0.687	P13	0.381	0.565	0.478
P14	0.743	0.58	0.613	P14	0.632	0.69	0.615	P14	0.726	0.559	0.435
P15	0.652	0.871	0.488	P15	0.479	0.532	0.602	P15	0.518	0.701	0.578
P16	0.759	0.887	0.623	P16	0.781	0.518	0.827	P16	0.703	0.744	0.574
P17	0.795	0.512	0.589	P17	0.516	0.903	0.480	P17	0.536	0.727	0.621
P18	0.699	0.577	0.570	P18	0.572	0.607	0.805	P18	0.668	0.437	0.643
P19	0.608	0.530	0.577	P19	0.699	0.742	0.442	P19	0.563	0.479	0.576
P20	0.821	0.539	0.527	P20	0.556	0.532	0.763	P20	0.366	0.419	0.565
AVE	0.722	0.653	0.562	AVE	0.600	0.639	0.653	AVE	0.559	0.584	0.593

Figure 3 TTR for All Three Groups

Additionally, the following line graph visually depicts the trend in Type-Token Ratio (TTR) for the three groups across the simple, middle, and complex tasks, based on the average TTR values for each group. This visualization highlights how lexical variety changes in response to different levels of cognitive task complexity.

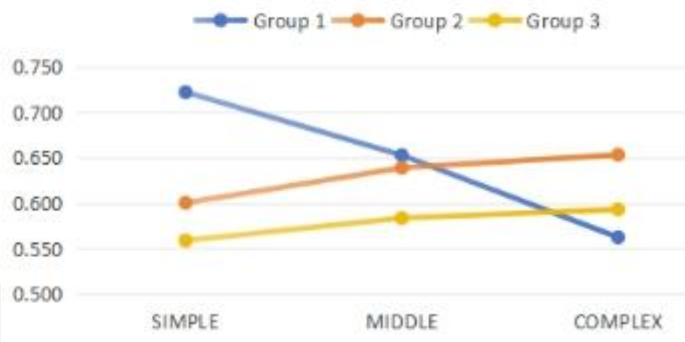

Figure 4 The TTR Trend for All three Groups

4.1.1.1 TTR for Group 1_SMC

The Type-Token Ratio (TTR) trend for Group 1_SMC across the three levels of cognitive task complexity reveals a noticeable decline in lexical variety as tasks become more demanding. By referring to both the line graph and the numerical results, we can identify specific segments where TTR decreases more sharply.

Simple ⟶ Middle

The TTR shows a moderate drop when comparing the Simple to Middle tasks, though its p-value is .0737, which is above the conventional .05 threshold. In statistical terms, this means the difference does not quite reach standard significance. Nonetheless, we can still observe a slight negative trend in lexical diversity. One possible explanation is that as participants move from a very basic task to a moderately more demanding one, they begin to allocate more attentional resources to maintaining accuracy or structure, and thus their lexical variety may start to dip—but not drastically enough to be deemed statistically significant.

Middle ⟶ Complex

When moving from the Middle to the Complex tasks, the line graph demonstrates a relatively more pronounced decrease in TTR, and the associated p-value is .0072, falling well below the .05 standard. This indicates a statistically significant drop in lexical variety. Such a pattern suggests that as complexity continues to rise beyond the moderate level, participants find it increasingly challenging to sustain the same breadth of vocabulary. In other words, higher cognitive demands prompt them to reduce word variety, likely due to the need to manage more complex syntax, maintain fluency, or attend to accuracy.

Simple ⟶ Complex

A direct comparison between the Simple and Complex tasks corroborates this downward trend. The p-value of .0002 is highly significant, confirming a marked decline in TTR between these two extremes of task complexity. Notably, the difference here is even more striking than the Simple–Middle contrast, implying that the lexical variety is notably sacrificed once participants reach the highest level of cognitive load.

Overall, the above results paint a consistent picture: increasing complexity corresponds to a progressive (and statistically evident) loss in lexical variety for Group 1, with the largest gap emerging between the simplest and most demanding tasks. Such findings align with the notion that when tasks become more mentally taxing, participants tend to concentrate on constructing grammatically acceptable or sufficiently fluent oral output—potentially at the expense of exploring a broader vocabulary range.

4.1.1.2 TTR for Group 2_CMS

Turning to Group 2_CMS, we notice a different or at least less straightforward pattern in how TTR evolves with task complexity. Unlike Group 1, which displayed a consistent decline, Group 2 exhibits some small fluctuations around the various comparisons, but most do not reach conventional significance levels.

Simple ⟶ Middle

From Simple to Middle tasks, the line graph indicates a slight increase in TTR, but with a p-value of .2679, this difference is not statistically significant. Although a minimal upward trend might exist visually, the data do not provide firm evidence that participants truly expanded their lexical variety in a reliable way. It might be that the moderate level of complexity here resonates favorably with some participants, but the effect is neither consistent nor strong enough to show significance across the group.

Middle ⟶ Complex

Moving from Middle to Complex tasks, the TTR exhibits a stable or marginal change, supported by p = .6815—again well above .05. Such a result indicates that the lexical variety remains roughly the same and is not notably impacted by the shift from moderate to high complexity in this group. One possible rationale is that participants in Group 2, already starting with a certain approach to vocabulary use, did not adjust their lexical strategies significantly when further complexity was introduced.

Simple ⟶ Complex

When directly comparing the Simple and Complex tasks for Group 2, the data suggests a moderate upward difference in TTR, but p = .1062 shows no statistical significance. Although the result is somewhat close to .1, it remains above .05, and we typically interpret that as “no clear evidence of a real difference”. Thus, while the line graph might depict a small upward shift, the variation is still classified as not reliably different.

All in all, Group 2 does not exhibit any significant TTR changes as task complexity increases. Most p-values remain above .05, pointing to the conclusion that lexical variety in Group 2 is relatively unaffected by the changes in cognitive demands. They may either adopt a stable vocabulary usage pattern regardless of complexity, or the group’s variability is too large to detect a consistent effect.

4.1.1.3 TTR for Group 3_MCSI

Finally, examining Group 3_MCS yields yet another perspective on TTR in response to task complexity. The line graph suggests a subtle upward slope, but a closer look at p-values clarifies the degree to which these changes are conclusive.

Simple ⟶ Middle

The comparison between the Simple and Middle tasks shows a TTR difference with p = .4617, well above .05. Although participants may display a minor rise in lexical variety, the statistical outcome deems it not significant. This implies that if there is indeed an upward trend in vocabulary usage, it is too slight or inconsistent to be confirmed rigorously.

Middle ⟶ Complex

Moving from Middle to Complex, TTR changes remain very slight, with p = .7980, which is also far from the typical cut-off for significance. Thus, the lexical variety does not appear to shift in a reliable manner when complexity steps up from moderate to high for Group 3. One potential explanation is that participants might have found a lexical repertoire that sufficed at the Middle level and simply continued using it under more complex demands, rather than broadening or restricting it substantially.

Simple ⟶ Complex

Directly comparing the Simple and Complex tasks, we see p = .3748, again above .05, signifying no statistically significant difference. Even if the graph might indicate a gentle incremental gain in TTR, the data as a whole is too variable or the effect too small to pass standard thresholds of significance.

In summary, though Group 3's line graph points to a mildly increasing TTR with rising task complexity, none of the stepwise comparisons meet significance criteria. We cannot confidently conclude that more demanding tasks in Group 3 lead to higher lexical variety; at best, we can say there is a possible positive trend that fails to reach conventional significance. Reasons might include heterogeneous strategies among participants, partial offset by attentional limits, or simply that the tasks' complexity did not strongly affect their vocabulary usage patterns.

4.1.2 Syntactic Complexity-Mean length of AS-unit

The Mean Length of AS-unit (MLAS) is a linguistic metric commonly employed to evaluate the syntactic complexity of spoken or written language. An AS-unit (Analysis of Speech unit) comprises a main clause and any subordinate clauses or phrases linked to it. MLAS is calculated by determining the average number of words or syllables within each AS-unit in a given speech or text sample. This metric is frequently used in language proficiency assessments to gauge the complexity and elaboration of a speaker's or writer's language production. Higher MLAS values typically indicate more intricate and sophisticated language use, reflecting greater syntactic complexity.

The following table presents the MLAS results for all three groups across the three levels of Cognitive Task Complexity (CTC), offering insights into the relationship between task complexity and syntactic elaboration.

MEAN LENGTH OF AS-UNIT FOR THREE GROUPSS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	6.12	6.87	9.12	P1	6.6	9.75	7.8	P1	6.3	5.8	5.73
P2	5.14	5.91	9.44	P2	9.33	6.41	9.64	P2	8.08	12.9	7.69
P3	4.75	7.5	6.55	P3	7.67	6.38	6.73	P3	4.36	7	6.88
P4	6.45	6.1	5.46	P4	4.89	3.91	3.8	P4	5.26	7.44	7.1
P5	3.54	4.69	6.44	P5	5.56	5.36	5.75	P5	7	9.39	7.55
P6	5.13	10.5	5.79	P6	8.15	7.2	6	P6	7.17	6.15	5.46
P7	7.11	6.88	6.1	P7	7.87	5.06	5.94	P7	6.14	5.96	7.88
P8	4.35	8	6.58	P8	6.39	5.15	5.7	P8	6.25	9.63	6.67
P9	6.5	4.71	9.73	P9	4.36	4.92	4.67	P9	3.54	5.22	3.67
P10	4.55	6.25	8.47	P10	6.36	7.13	9.14	P10	8.48	7.19	7.5
P11	6.15	8.3	6.6	P11	7.46	5.75	4.71	P11	9.73	9.38	8.48
P12	6.73	7.33	4.61	P12	5.97	5.48	5.5	P12	7.75	9	8.06
P13	4.28	6.33	5.29	P13	8.38	5.29	6.92	P13	8.42	6.4	8.81
P14	5.25	7.29	5.29	P14	5	4.35	9.1	P14	4.71	8.29	7.22
P15	5.383	4.307	7.209	P15	5.41	6.06	8.55	P15	5.00	9.09	7.48
P16	5.15	6.913	5.285	P16	6.24	3.38	7.61	P16	4.56	5.42	4.97
P17	4.80	4.96	6.92	P17	8.09	5.62	4.07	P17	4.53	10.28	5.98
P18	5.62	8.76	7.11	P18	5.42	7.16	5.24	P18	6.59	7.27	9.37
P19	5.87	8.63	7.27	P19	6.04	5.59	6.43	P19	9.70	5.32	7.45
P20	5.76	7.85	7.12	P20	9.09	7.40	7.28	P20	9.55	9.66	7.04
AVE	5.432	6.904	6.819	AVE	6.714	5.867	6.529	AVE	6.656	7.839	7.049

Figure 5-Mean Length of AS-unit for All three groups

The following line graphs visually depict the trends in syntactic complexity, as measured by the Mean Length of AS-unit (MLAS), for the three groups across the simple, middle, and complex tasks.

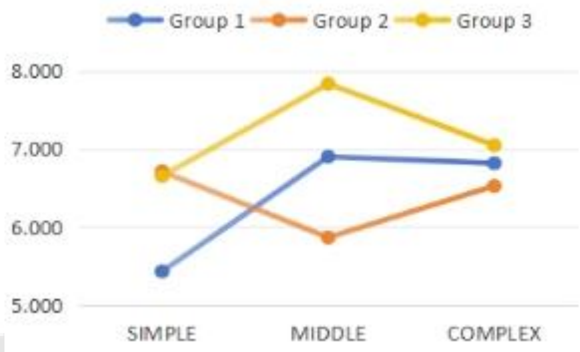

Figure 6-Mean Length of AS-unit Trend for All Three Groups

4.1.2.1 MLAS for Group 1_SMC

Examining Group 1_SMC, the line graph shows distinct changes in Mean Length of AS-unit (MLAS) with rising cognitive task complexity. Three key pairwise comparisons help clarify how syntactic complexity evolves in this group:

Simple ⟶ Middle

Moving from Simple to Middle tasks, the data reveal a substantial increase in MLAS. Statistical testing yields p = .0033, well below the .05 threshold, indicating a significant difference. Thus, Group 1 participants appear to construct more elaborate AS-units when complexity escalates from a relatively low to a moderate level. One plausible reason is that moderate complexity challenges them to expand clauses—perhaps by adding subordinate phrases—without overwhelming their attentional capacity.

Middle ⟶ Complex

By contrast, from Middle to Complex tasks, the MLAS slightly decreases. The p-value of .6966 is not statistically significant, implying no reliable difference. This suggests that after an initial jump in syntactic elaboration at moderate complexity, participants do not continue extending their AS-units further when faced with even higher cognitive demands. They may have reached a plateau in syntactic resource allocation, or else they shift attention toward maintaining accuracy or lexical access rather than further increasing length.

Simple ⟶ Complex

When comparing Simple vs. Complex tasks directly, the data again demonstrate a clear upward shift in MLAS. With $p = .0152 < .05$, the jump is deemed significant, reinforcing that overall, participants produce longer AS-units when tasks move from the simplest to the highest complexity. This result dovetails with the strong effect found in the Simple–Middle comparison, implying a “moderate jump” that persists into the final complexity level, even if the step from Middle to Complex is not evident on its own.

Overall, Group 1 exhibits a significant boost in syntactic elaboration from Simple to Middle tasks, and a similarly significant difference between Simple and Complex tasks—though no additional gain from Middle to Complex. In practical terms, moderate complexity seems especially conducive to more elaborated syntactic production, while moving to the highest level does not yield a further improvement. This pattern can be interpreted as partial support for the Cognition Hypothesis, insofar as moderate complexity pushes participants to produce more intricate AS-units, yet also hints at an upper bound where the gains in syntactic complexity no longer materialize under heavier cognitive load.

4.1.2.2 MLAS for Group 2_CMS

Turning to Group 2_CMS, we see a different pattern in MLAS changes as task complexity shifts from Complex ⟶ Middle ⟶ Simple (based on the group’s CTC ordering). The line graph captures three main comparisons:

Simple ⟶ Middle

The data show a noticeable dip in MLAS from Simple-level tasks to Middle-level tasks, with a p-value of .1404. Since .1404 exceeds the .05 threshold, we label this difference not statistically significant, though the mild downward trend might suggest participants briefly reduce their AS-unit length under Middle tasks. Possibly, they reorganize resources or shift attention to other aspects (like accuracy) when stepping down in complexity from a prior challenge.

Middle ⟶ Complex

From Middle to Complex, the line graph suggests a minor recovery in MLAS, but the p-value of .2962 is similarly not significant. In practical terms, there is no strong evidence that participants expand (or reduce) their AS-units going from a moderate to a high level of complexity. One might speculate that, having begun with a more difficult task or toggling between complexities, they develop a stable syntactic approach that does not vary much with final complexity changes.

Simple ⟶ Complex

Directly comparing Simple and Complex tasks, the line graph remains relatively steady, with p = .7687 also failing to reach significance. Hence, Group 2's MLAS does not differ reliably from the simplest to the most complex tasks in their sequence arrangement. This general stability could be interpreted in two ways:

participants maintain a consistent approach to syntax, unaffected by further changes in complexity.

There is considerable inter-participant variability overshadowing any small differences.

All p-values lie above .05, meaning none of the pairwise complexity comparisons yield significant MLAS changes. The minor fluctuations in the line graph do not translate into robust evidence that Group 2's syntax is strongly modulated by complexity. This contrasts with Group 1's clear gains, suggesting Group 2 might either distribute their cognitive resources differently (e.g., focusing on lexical choices or fluency) or lack the impetus to adjust syntax substantially for more demanding tasks.

4.1.2.3 MLAS for Group 3_MCS

Lastly, Group 3_MCS provides another perspective on how MLAS responds to rising cognitive demands. The line graph indicates a possible "peak" around the middle complexity, followed by a slight dip at the complex level, though p-values clarify the significance:

Simple ⟶ Middle

The data show an increase in MLAS from Simple to Middle tasks, with a reported p-value of .0682. While this does not meet the .05 significance criterion, it hovers in what some might call a “marginal” or “trend” region. We might interpret it as suggestive that participants elaborate syntax somewhat more at moderate complexity, but the effect is not strong enough to claim definitive significance.

Middle ⟶ Complex

Subsequently, from Middle to Complex, MLAS appears to decline. The corresponding p-value of .1449 is also above .05, so again not statistically significant. This indicates that any downward trend is similarly not robust enough to be declared a clear difference. The pattern could reflect participants reaching a threshold at the Middle level, and then pulling back or plateauing once faced with even higher demands.

Simple ⟶ Complex

A direct comparison between Simple and Complex tasks yields p = .5147, well above .05, indicating no reliable difference in average MLAS. Thus, although the line graph suggests a potential “rise then fall”, the statistics do not confirm a strong shift in AS-unit length from the simplest to the most demanding tasks. The group’s variance or varying participant strategies might overshadow any consistent pattern.

Overall, Group 3’s MLAS data hint at a mild uptick in syntactic elaboration when moving from Simple to Middle tasks but does not reach significance. The subsequent drop at Complex likewise fails to produce a statistically clear result. Essentially, Group 3 does not show conclusive changes in AS-unit length across complexity levels, pointing to an ambiguous or relatively weak link between higher cognitive load and syntactic expansion or contraction. One explanation could be that some participants adapt better, while others are more overwhelmed, yielding a net effect that cancels out at the group level.

4.1.3 Syntactic Complexity-Mean length of Clause

The Mean Length of Clause (MLC) is a widely used metric for assessing the syntactic complexity of language, applicable to both spoken and written forms. It is frequently utilized in linguistic research to measure the elaboration and sophistication of language, with longer clauses typically indicating higher levels of syntactic complexity and more advanced language use.

The following table presents the MLC results for all three groups across the three levels of Cognitive Task Complexity (CTC), providing insights into how clause length varies with increasing task complexity.

MEAN LENGTH OF CLAUSE FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	4.36	5.8	7.31	P1	5.14	4.72	3.9	P1	4.85	4.77	5.25
P2	5.04	4.51	4.23	P2	6.31	5.9	4.82	P2	4.41	6.89	6.41
P3	4.55	6.15	4.08	P3	6.08	6.00	4.61	P3	2.29	5.00	4.62
P4	5.25	5.08	4.86	P4	3.98	3.67	2.71	P4	4	4.62	4.81
P5	3.32	4.29	4.25	P5	5	4.69	2.76	P5	3.26	4.42	5.19
P6	5.07	7.85	5	P6	6.47	5.79	3.38	P6	3.07	5.88	5.07
P7	3.69	5.55	4.85	P7	4.91	4.06	3.95	P7	3.21	4.68	4.47
P8	3.71	6.33	5.85	P8	5.13	4.53	3.19	P8	2.88	6.21	5
P9	4.75	4.65	7.2	P9	3.06	4.57	2.55	P9	1.92	4.33	2.88
P10	2.36	5.56	4.08	P10	3.07	5.53	3.88	P10	3.95	5.34	5.92
P11	4.62	6.72	5.67	P11	5.72	4.19	3.48	P11	3.4	6.88	7.39
P12	5.03	6.21	4.24	P12	2.87	4.57	2.75	P12	3.44	5.14	5.96
P13	5.98	4.57	5.29	P13	3.89	4.59	2.77	P13	3.21	4	6.74
P14	2.45	5.18	4.7	P14	2.83	3.29	4.79	P14	2.79	4	4.81
P15	5.14	4.28	6.10	P15	3.75	5.38	2.92	P15	3.17	4.69	7.12
P16	3.92	6.21	4.37	P16	4.69	2.79	4.76	P16	3.39	4.25	4.56
P17	4.80	4.21	3.96	P17	4.92	4.39	4.07	P17	2.83	7.09	4.14
P18	3.72	5.86	5.58	P18	4.83	5.99	2.56	P18	2.85	6.82	4.95
P19	4.58	6.29	4.13	P19	4.28	4.80	3.78	P19	2.66	3.61	4.49
P20	3.63	6.77	6.55	P20	5.17	5.00	3.16	P20	5.10	4.46	6.68
AVE	4.299	5.604	5.115	AVE	4.604	4.721	3.539	AVE	3.334	5.154	5.323

Figure 7 Mean Length of Clause for All Three Groups

The following line graphs visually illustrate the trends in syntactic complexity, as measured by the Mean Length of Clause (MLC), for the three groups across the simple, middle, and complex tasks. These graphs provide a clear depiction of how clause length changes with varying levels of cognitive task complexity.

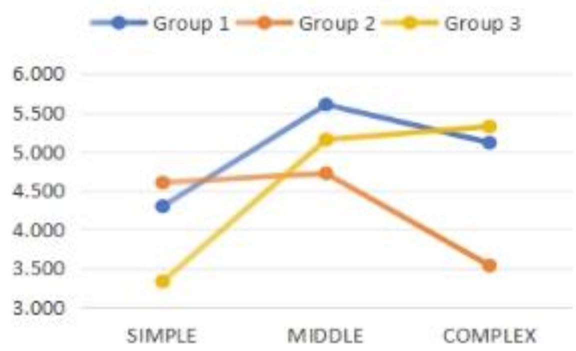

Figure 8 Mean Length of Clause Trend for All Three Groups

4.1.3.1 MLC for Group 1_SMC

Examining Group 1_SMC, the line graph reveals clear changes in Mean Length of Clause (MLC) as cognitive task complexity (CTC) escalates from Simple to Middle to Complex tasks. We can parse these shifts through three pairwise comparisons:

Simple ⟶ Middle

The data indicate a substantial increase in MLC when moving from Simple to Middle tasks. A p-value of .0024 is well below the .05 threshold, denoting a statistically significant difference. This suggests that Group 1 participants respond to moderate increases in task complexity by producing longer clauses—likely because they incorporate more subordinate structures or additional modifiers once the tasks demand slightly more elaborate speech.

One plausible explanation is that moderate complexity, compared to a very simple baseline, prompts learners to invest in more complex clause construction without overly straining their attention.

Middle ⟶ Complex

From Middle to Complex tasks, the MLC shows a slight decline in the line graph. Statistically, p = .2204 > .05, indicating no significant difference. Despite a visible dip, the evidence does not confirm that participants reliably shorten their clauses at higher complexity.

A potential interpretation is that, after an initial jump in complexity (from Simple to Middle), participants might stabilize their approach to clause building. They do not necessarily produce even longer sentences for the Complex task; they may redistribute resources to other language dimensions such as accuracy or fluency, resulting in no consistent further extension of clause length.

Simple ⟶ Complex

Directly comparing Simple with Complex, the MLC again displays a noticeable rise overall. The p-value of .0527 is near the .05 boundary; some might classify this as a borderline significant or marginal outcome. While not as decisively significant as the Simple–Middle contrast, it suggests a likely trend of increased clause length at the highest complexity level relative to the simplest.

Although the numerical difference does not meet a strict threshold, the pattern supports the notion that more demanding tasks can promote more complex clause structures—at least when measured from the simplest baseline.

Overall, Group 1 demonstrates a significant jump in clause length from Simple to Middle tasks, and likely some increase (albeit borderline) from Simple to Complex. However, the Middle–Complex comparison is not statistically significant, indicating no consistent rise once tasks surpass a moderate complexity. In practical terms, moderate difficulty fosters noticeably richer clauses; pushing complexity further may not necessarily yield longer clauses, possibly due to attentional trade-offs or participants reaching a plateau in syntactic elaboration.

4.1.3.2 MLC for Group 2_CMS

Shifting focus to Group 2_CMS, which follows a distinct complexity order—Complex, then Middle, then Simple—our analysis must interpret MLC with attention to how these tasks are sequenced.

Simple ⟶ Middle

The line graph indicates a modest increase from the Simple-level tasks to the Middle-level tasks, but the p-value of .7797 is far above .05, signifying no statistical significance. Despite the graph showing a modest upward shift, we cannot conclude there is a reliable difference in clause length.

One explanation might be that participants found both tasks sufficiently manageable, so their average clause length remains generally comparable, overshadowing any subtle changes.

Middle ⟶ Complex

In the group's actual "middle ⟶ complex" step, the line graph suggests a sharp drop in MLC. The p-value is reported as .0007, which is highly significant. This indicates that participants produce notably shorter clauses under the Complex task than they do at the Middle level. Possibly, as complexity climbs further, Group 2 participants allocate resources to controlling accuracy or lexical items, thus curtailing the structural elaboration of clauses.

This strong effect stands out and contrasts with some other sub-dimensions, implying that Group 2 is indeed sensitive to higher complexity in terms of syntactic structuring.

Simple ⟶ Complex

Comparing Simple directly with Complex, the line graph similarly reflects a downward shift in MLC, and p = .0156 is below .05, confirming a significant difference. This complements the Middle–Complex result: from either vantage (Simple or Middle), participants produce shorter clauses at Complex tasks. Overall, the data reveal that more advanced or demanding tasks push Group 2 to reduce clause length, possibly due to increased cognitive strain that limits syntactic expansion.

Group 2 shows no difference between Simple and Middle tasks but exhibits significant drops in MLC going from Middle to Complex or from Simple to Complex. This suggests that the highest complexity tasks hamper participants' ability to maintain or expand their clause structures, reinforcing the idea that Group 2 is particularly impacted by high levels of cognitive load, leading to more succinct syntax under heavier demands.

4.1.3.3 MLC for Group 3_MCS

Finally, Group 3_MCS provides yet another perspective on how Mean Length of Clause (MLC) responds to rising cognitive task complexity (CTC). The group tackles tasks in the order of Middle ⟶ Complex ⟶ Simple, so the comparisons among the three levels follow a unique pattern. The line graph suggests a generally upward trajectory in clause length from Simple to Complex, though step-by-step contrasts clarify where changes do or do not reach statistical significance:

Simple ⟶ Middle

As depicted in the line graph, shows a consistent upward trajectory in syntactic complexity as task complexity rises. From Simple to Middle tasks, the line graph demonstrates a pronounced increase in MLC, reflecting enhanced syntactic elaboration. This trend is strongly supported by the statistical findings, with $p=.00001$ indicating a very high degree of confidence and a large positive trend underscoring the significant jump in clause length.

Middle ⟶ Complex

Moving from Middle to Complex tasks, the line graph shows relatively stable MLC, indicating minimal variation. The reported p-value is .6729, well above the .05 threshold, thus not statistically significant. In plain terms, Group 3's participants do not reliably shift their clause length when moving from a moderate to a high level of complexity. They may already be near some plateau in syntactic structuring, or else distribute attention to other linguistic dimensions such as lexis or accuracy.

Simple ⟶ Complex

By contrast, directly comparing the simplest and the most complex tasks indicates a clearly visible upward jump in the line graph. The reported p-value is .0067, which is well below .05, thus showing a significant difference. Consequently, Group 3 participants do produce longer clauses when tasks shift from very basic demands to the highest complexity. One interpretation is that only a large gap in task difficulty prompts a tangible syntactic adjustment in clause length, whereas a smaller increment (Middle ⟶ Complex) does not consistently spur further expansion.

Overall, Group 3's MLC results paint a somewhat mixed picture. The difference between Middle and Complex remains non-significant, suggesting no reliable adjustment in clause length at the moderate-to-high step of complexity. Yet when comparing Simple to Complex, we see a statistically significant leap, indicating that a sufficiently large jump in cognitive load can indeed motivate participants to craft longer clauses. This partial alignment with the Cognition Hypothesis (i.e., more complexity fosters more elaborate syntax) emerges only if the difference in difficulty is pronounced, pointing to a threshold effect in Group 3's response to rising complexity.

4.1.4 Accuracy_Error-Free Clause Ratio (EFCR)

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.

The following table presents the EFCR results for all three groups across the three levels of Cognitive Task Complexity (CTC), offering insights into how task complexity influences accuracy.

ERROR-FREE CLAUSE RATIO FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	0.923	0.867	0.823	P1	0.785	0.846	0.929	P1	0.818	0.7	0.865
P2	0.769	0.909	0.923	P2	0.842	0.824	0.812	P2	0.889	0.947	0.765
P3	0.8	0.9	0.916	P3	0.712	0.769	0.9	P3	0.627	0.737	0.832
P4	0.727	0.8	0.889	P4	0.75	0.818	0.9	P4	0.889	0.7	0.769
P5	0.846	0.75	0.8	P5	0.75	0.722	0.733	P5	0.695	0.731	0.9
P6	0.923	0.5	0.778	P6	0.818	0.917	0.857	P6	0.867	0.725	0.8
P7	0.818	0.845	0.9	P7	0.867	0.727	0.833	P7	0.652	0.782	0.833
P8	0.692	0.909	0.769	P8	0.913	0.9	0.75	P8	0.756	0.923	0.929
P9	0.538	0.786	0.824	P9	0.913	0.857	0.846	P9	0.9	0.727	0.778
P10	0.846	0.833	0.8	P10	0.909	0.923	0.813	P10	0.8	0.969	0.919
P11	0.75	0.909	0.925	P11	0.889	0.842	0.95	P11	0.725	0.913	0.965
P12	0.6	0.75	0.85	P12	0.778	0.786	0.857	P12	0.757	0.882	0.915
P13	0.75	0.833	0.857	P13	0.75	0.833	0.857	P13	0.725	0.963	0.906
P14	0.789	0.833	0.8	P14	0.944	0.894	0.933	P14	0.944	0.825	0.904
P15	0.829	0.892	0.92	P15	0.775	0.8	0.79	P15	0.88	0.675	0.941
P16	0.734	0.695	0.65	P16	0.865	0.645	0.83	P16	0.9	0.88	0.81
P17	0.63	0.65	0.815	P17	0.78	0.675	0.64	P17	0.87	0.69	0.74
P18	0.933	0.75	0.85	P18	0.724	0.62	0.75	P18	0.86	0.69	0.925
P19	0.751	0.82	0.875	P19	0.85	0.665	0.65	P19	0.645	0.925	0.895
P20	0.701	0.845	0.790	P20	0.94	0.903	0.893	P20	0.6	0.865	0.725
AVE	0.767	0.804	0.838	AVE	0.828	0.798	0.826	AVE	0.790	0.812	0.856

Figure 9 Error-Free Clause Ratio for All Three Groups

The following line graphs visually depict the trends in accuracy, as measured by the Error-Free Clause Ratio, for the three groups across the simple, middle, and complex tasks. The graph provide a clear representation of how phonetic performance varies with increasing levels of cognitive task complexity.

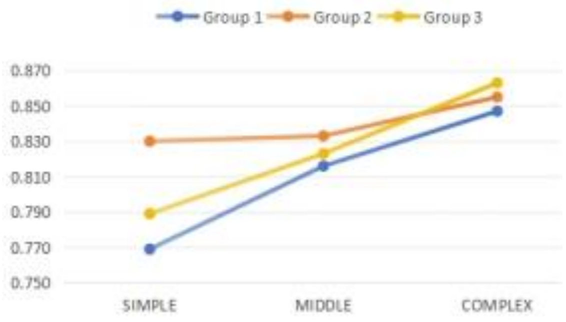

Figure 10- Error-Free Clause Ratio Trends for Three Groups

4.1.4.1 EFCR for Group 1_SMC

Turning first to Group 1_SMC, we examine how the Error-Free Clause Ratio (EFCR) varies with increases in cognitive task complexity (CTC). The line graph illustrates a generally upward pattern in EFCR from Simple to Middle to Complex, implying participants in Group 1 become more accurate at higher task demands—though we must confirm whether these changes reach statistical significance.

Simple ⟶ Middle

The data show a rise in EFCR from Simple to Middle tasks, with a p-value of .2614—above the conventional .05 threshold. Consequently, this difference is not statistically significant. Visually, the line graph indicates a moderate upward shift, but the variability among participants or the size of the effect do not suffice to confirm a reliable change in clause-level accuracy at this step.

Middle ⟶ Complex

From Middle to Complex tasks, EFCR appears to further increase, with the test yielding p = .3467, again above .05, hence also not significant. Although the line graph suggests a continued upward trend, we lack evidence to state definitively that Group 1's accuracy is substantially higher at Complex than at Middle complexity. It is possible that individual differences or other performance factors overshadow any small gains.

Simple ⟶ Complex

In contrast to those two adjacent comparisons, the direct comparison between Simple and Complex tasks yields p = .0253, which is below .05, indicating a significant difference. This result means that, despite neither partial step (Simple⟶Middle or Middle⟶Complex) being conclusive on its own, participants demonstrate a notably higher EFCR at the highest complexity relative to the simplest. One interpretation is that Group 1 invests resources across the entire progression—though the incremental improvements from Simple to Middle and Middle to Complex individually might be too small or variable to register as significant, the cumulative jump is large enough to be detected when comparing the extremes.

Overall, while the line graph shows a steady upward slope in EFCR, only the Simple–Complex contrast meets the significance threshold. Thus, we can assert that Group 1's accuracy, measured via EFCR, improves meaningfully by the time they reach the most complex tasks compared to the simplest tasks, even if each intermediate step is not independently reliable. This pattern is consistent with the possibility that participants gradually adapt and allocate attentional resources to reduce clause-level errors over the entire task progression, culminating in a measurable gain at the highest complexity.

4.1.4.2 EFCR for Group 2_CMS

Next, we assess Group 2_CMS, whose complexity ordering runs Complex ⟶ Middle ⟶ Simple, to see if EFCR changes across tasks with different demands.

Simple ⟶ Middle

The data indicate minimal change, yielding p = .9200, far above .05. Hence, there is no evidence of a significant difference in EFCR between these two complexity levels. Any small variations in accuracy are likely overshadowed by participant-to-participant variability, or simply do not occur consistently.

Middle ⟶ Complex

Moving from Middle to Complex tasks, the line graph suggests a minor increase in EFCR, but p = .3710 also exceeds .05, meaning not significant. Thus, we find no reliable improvement or decline in accuracy as participants move to the highest complexity, at least not when referencing the immediate prior level.

Simple ⟶ Complex

Finally, directly comparing Simple and Complex tasks yields p = .3595, again well above .05, signifying no significance. Similarly, the comparison between Simple and Complex tasks shows a moderate increase in EFCR, yet the p-value (p=.3595) again implies limited confidence in this difference.

All pairwise p-values are above .05, indicating that none of the complexity steps produce a clear shift in EFCR. The line graph might show small fluctuations in clause-level accuracy, but these fail to meet standard significance criteria. We can conclude that Group 2's accuracy, as measured by EFCR, does not reliably change when tasks become more (or less) cognitively demanding. One plausible explanation is that Group 2 participants maintain a certain error rate across tasks, focusing on other aspects—like lexical choices or syntax expansions—and not devoting additional attention to error reduction.

4.1.4.3 EFCR for Group 3_MCS

Lastly, Group 3_MCS addresses tasks in the order Middle ⟶ Complex ⟶ Simple. We analyze whether EFCR exhibits any discernible patterns under these shifting complexities.

Simple ⟶ Middle

Between Simple and Middle tasks, the line graph shows a slight increase in EFCR, suggesting a potential improvement in linguistic accuracy. However, the statistical findings indicate a low degree of confidence in detecting this difference (p=.3833), implying that the observed increase does not meet the conventional threshold for significance.

Middle ⟶ Complex

From Middle to Complex tasks, the line graph points to a further upward shift in EFCR, yet the data yield p = .2434, again exceeding .05. Therefore, we cannot claim a significant improvement or drop in accuracy when complexity steps up from moderate to high. This might mean some participants do improve while others falter, averaging out to no consistent pattern at the group level.

Simple ⟶ Complex

Directly comparing Simple vs. Complex tasks, the p = .0282, which is below the .05 boundary, implying a significant difference. That suggests the overall jump in complexity from the simplest to the most demanding tasks is enough to produce a visible improvement in EFCR. This lines up with the notion that participants can reorganize resources over a large gap in difficulty, potentially focusing more on ensuring accurate clauses.

Similar to Group 1, Group 3 does not see a significant shift at each intermediate step (Simple⟷ Middle or Middle⟷ Complex), but the Simple–Complex comparison is reported as significant, indicating an overall improvement in EFCR when tasks shift from very easy to very demanding. The group, therefore, displays a pattern where accuracy changes become evident only when contrasting the lowest and highest complexities, reinforcing the idea that smaller incremental shifts may be too subtle or inconsistent to detect, but a larger jump in task difficulty can yield a noticeable impact on error-free production.

4.1.5 Total Number of Words_TNW

The Total Number of Words (TNW) is a fundamental metric used to quantify the length of a spoken or written text. This measure counts every individual word within a given sample, offering a straightforward means to assess the overall volume of language produced. TNW is widely employed in linguistic analysis to evaluate text length, measure verbosity, and compare the extent of expression across different samples or experimental conditions. While a basic measure, TNW is essential for understanding language production and serves as a preparatory step for further analyses, including fluency evaluations.

The following table presents the TNW results for all three groups across the three levels of Cognitive Task Complexity (CTC), providing insights into how task complexity influences the overall language performance.

TOTAL NUMBER OF WORDS FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	82	112	114	P1	75	83	78	P1	92	99	139
P2	76	72	96	P2	125	123	106	P2	64	122	153
P3	67	74	64	P3	64	85	111	P3	81	69	127
P4	71	65	105	P4	48	45	38	P4	84	130	144
P5	48	62	50	P5	51	58	69	P5	78	75	73
P6	85	70	114	P6	110	115	108	P6	113	138	115
P7	71	66	60	P7	106	88	95	P7	127	83	106
P8	56	83	81	P8	120	64	131	P8	88	91	110
P9	93	83	114	P9	62	67	84	P9	96	135	94
P10	57	104	114	P10	90	109	128	P10	120	95	81
P11	75	83	69	P11	100	122	80	P11	99	115	118
P12	104	122	107	P12	67	65	66	P12	123	103	101
P13	38	33	40	P13	108	90	83	P13	106	111	123
P14	113	122	111	P14	73	86	91	P14	71	119	148
P15	57	100	97	P15	78	116	97	P15	67	80	131
P16	88	73	116	P16	68	108	112	P16	103	88	98
P17	84	70	96	P17	78	64	73	P17	116	72	89
P18	101	61	53	P18	138	80	86	P18	74	107	86
P19	53	113	58	P19	81	74	78	P19	61	143	77
P20	60	76	111	P20	80	79	105	P20	109	127	136
AVE	73.95	82.20	88.50	AVE	86.10	86.05	90.95	AVE	93.60	105.10	112.45

Figure 11 Total Number of Words for All Three Groups

The following line graphs visually illustrate the trends in productivity, as measured by the Total Number of Words (TNW), for the three groups across the simple, middle, and complex tasks. These graphs provide a clear depiction of how language performance varies with increasing levels of cognitive task complexity.

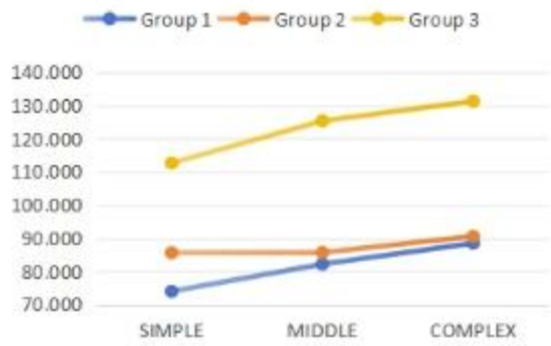

Figure 12 Total Number of Words Trend for All Three Groups

4.1.5.1 TNW for Group 1_SMC

Focusing on Group 1_SMC, we examine how the Total Number of Words (TNW) produced by participants shifts with increases in cognitive task complexity (CTC) from Simple to Middle to Complex tasks. The line graph suggests a modest upward slope, but the statistical tests clarify whether these changes meet significance:

Simple ⟶ Middle

Moving from Simple to Middle tasks, TNW shows a visible increase, yet the p-value of .2833 exceeds .05, indicating no statistically significant difference. Although the graph depicts a mild upward trend, it appears the variation across individuals is too large or the effect size too small to conclude a reliable jump in word count at this moderate rise in complexity.

Middle ⟶ Complex

From Middle to Complex, TNW remains relatively stable or continues with only a slight increase, accompanied by p = .6357—again above .05, hence not significant. In practical terms, Group 1 participants do not consistently expand their overall word count as tasks become more demanding beyond the moderate level. Some may speak more words due to increased challenge, while others might shorten responses, leaving no net effect.

Simple ⟶ Complex

A direct comparison between Simple and Complex tasks yields p = .1233, which also does not fall below .05, so it is not significant despite showing a somewhat larger numerical gap. This implies that, although the line graph hints at an upward trend from simplest to most demanding, the data do not confirm a strong or consistent enough effect to meet typical significance criteria.

Overall, while the line graph depicts a modest rise in TNW across ascending complexity levels, none of the stepwise comparisons proves to be statistically significant. This suggests that participants in Group 1, despite an apparent inclination to produce more words as tasks become more challenging, do not show a consistent pattern that can be deemed reliable at standard alpha levels. Their productivity may be influenced by a host of factors—e.g., how quickly they adapt or how they balance accuracy, complexity, and fluency—leading to a mild but inconclusive increase in total production.

4.1.5.2 TNW for Group 2_CMS

For Group 2_CMS, which encounters tasks in the sequence Complex ⟶ Middle ⟶ Simple, we evaluate whether and how TNW fluctuates under varying cognitive loads.

Simple ⟶ Middle

The data show almost no change in TNW, with p = .9353 well above .05, hence clearly not significant. The line graph similarly indicates near-flat levels of total word output, suggesting participants produce about the same volume of speech at these two levels.

Middle ⟶ Complex

Shifting from Middle to Complex tasks, the line graph portrays a slight increase in TNW, but p = .6098 also falls above .05, meaning no significant difference. This result indicates that, for Group 2, raising complexity beyond Middle does not systematically induce them to speak more words. They might approach tasks with a steady baseline of word output, focusing their cognitive capacity on other aspects (e.g., accuracy or syntactic complexity).

Simple ⟶ Complex

Comparing the extremes—Simple vs. Complex—p = .5624, again not significant. Despite the tasks being quite different in difficulty, participants do not reliably generate more or fewer words overall. One possibility is that having started at a high-complexity task early in the session, Group 2 quickly stabilizes in their approach and remains unaffected by additional shifts in complexity.

The results show no significant changes in TNW among any of the complexity pairings. The line graph basically reveals small or negligible differences, aligning with p-values all well above .05. Thus, Group 2's total word output remains broadly unchanged when tasks move from simpler to more complex or vice versa, hinting that other performance dimensions (e.g., fluency rate or lexical variety) might be more sensitive for this group than raw word count.

4.1.5.3 TNW for Group 3_MCS

Lastly, Group 3_MCS addresses tasks in Middle ⟶ Complex ⟶ Simple order. The question is whether total word production shifts notably with ascending or descending complexity levels.

Simple ⟶ Middle

From Simple to Middle tasks, the graph shows a minimal upward trend in TNW, aligning with statistical findings. A p-value of .5192, above the conventional 0.05 threshold, indicates a low degree of confidence in detecting a difference, and the small positive trend underscores limited variation in word count.

Middle ⟶ Complex

Between Middle and Complex tasks, the line graph suggests stability in TNW, with only a slight increase observed. This corresponds to the statistical analysis, where p=.7965 reflects low confidence in detecting differences, and the small effect size further confirms minimal variation in productivity at these levels.

Simple ⟶ Complex

For the comparison between Simple and Complex tasks, the line graph shows a modest upward trend, consistent with the statistical findings. The p-value of .4345 again suggests low confidence in detecting a difference, and the small positive trend indicates only a slight increase in TNW.

As with Group 2, none of the pairwise complexity contrasts produce a statistically significant shift in word count for Group 3. Despite the line graph's mild movements up or down, the p-values all exceed .05, indicating no robust evidence that complexity strongly influences the total number of words. In effect, Group 3 participants may maintain relatively stable verbosity regardless of whether tasks are moderately or highly demanding, or whether they end on a simpler note.

4.1.6 Total Number of Syllables_TNS

The Total Number of Syllables is a linguistic metric that quantifies the phonetic components of spoken or written language by counting the total syllables in a given sample. This measure provides valuable insights into the rhythm, complexity, and fluency of language production. It is widely utilized in linguistic analysis, language learning, and speech evaluation to assess the structure and delivery of language. Additionally, the total number of syllables serves as a foundational metric for calculating other key measures, such as speech rate and articulation rate, which are critical indicators of fluency and language proficiency. By analyzing the total syllables, researchers can better understand the relation between language complexity and delivery in various contexts.

TOTAL NUMBER OF SYLLABLES FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	135	157	193	P1	107	109	108	P1	114	120	150
P2	109	127	142	P2	154	144	126	P2	136	123	154
P3	85	127	104	P3	109	121	135	P3	100	130	128
P4	87	89	160	P4	67	54	44	P4	112	131	145
P5	55	82	66	P5	70	72	78	P5	100	93	140
P6	91	97	155	P6	148	140	131	P6	125	138	116
P7	85	82	72	P7	138	98	108	P7	128	94	143
P8	63	124	111	P8	186	72	163	P8	110	134	111
P9	107	121	157	P9	71	85	100	P9	113	136	120
P10	63	145	162	P10	128	142	169	P10	121	124	141
P11	92	116	96	P11	142	138	88	P11	101	130	119
P12	124	138	137	P12	71	72	86	P12	124	104	137
P13	45	42	49	P13	150	106	99	P13	110	115	150
P14	129	131	147	P14	82	109	110	P14	80	125	149
P15	86	120	129	P15	100	129	101	P15	118	119	132
P16	103	106	140	P16	122	114	120	P16	122	100	130
P17	94	116	138	P17	127	106	101	P17	117	111	150
P18	127	85	98	P18	148	98	102	P18	90	112	120
P19	80	131	112	P19	93	92	114	P19	99	144	112
P20	85	119	133	P20	107	97	113	P20	120	128	137
AVE	90.71	112.71	125.07	AVE	115.93	104.36	110.36	AVE	112.00	120.55	134.20

Figure 13 Total Number of Syllables for All Three Groups

The following line graphs visually depict the trends in productivity, as measured by the Total Number of Syllables, for the three groups across the simple, middle, and complex tasks. These graphs provide a clear representation of how phonetic output varies with increasing levels of cognitive task complexity.

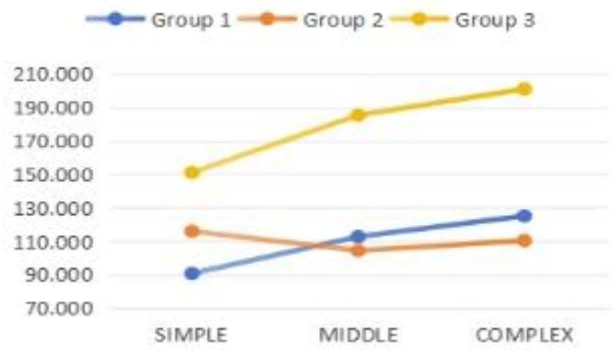

Figure 14 Total Number of Syllables Trend for All Three Groups

4.1.6.1 TNS for Group 1_SMC

For Group 1_SMC, we examine whether the Total Number of Syllables (TNS) produced by participants varies significantly as they move from Simple to Middle to Complex tasks. The line graph visually suggests a gradual increase in TNS across these complexity levels, but the pairwise comparisons clarify where changes are (or are not) statistically significant.

Simple ⟶ Middle

From Simple to Middle tasks, TNS shows an upward shift, with a p-value of .0580. This number is slightly above .05, often considered marginal or approaching significance, though not strictly below the conventional threshold. The pattern may indicate a possible trend toward producing more syllables when complexity rises from a low to a moderate level, yet the data stop short of confirming it as definitively reliable.

Middle ⟶ Complex

Moving from Middle to Complex tasks, TNS continues to increase, albeit more modestly, with p = .3870—well above .05, hence not statistically significant. While the line graph still points to a small positive slope, the variability across participants likely obscures a consistent effect. In other words, some participants may speak more syllables at the highest complexity, while others do not.

Simple ⟶ Complex

Directly comparing the simplest and the most demanding tasks yields p = .0180, which is below .05 and thus significant. This result implies that, considered over the entire gap in complexity, participants in Group 1 produce significantly more syllables at the complex level than at the simplest level. Put differently, even though each incremental step (Simple⟶Middle or Middle⟶Complex) did not individually pass the threshold, the cumulative difference from Simple to Complex is large enough to be detected statistically.

Overall, Group 1 exhibits a significant jump in total syllables when comparing Simple vs. Complex tasks, but the changes from one step to the next (Simple⟶Middle or Middle⟶Complex) are not definitively significant—though the Simple⟶Middle difference hovers near marginal. This suggests that an overall increase in TNS emerges once participants reach the highest complexity, consistent with the idea that Group 1's speaking output expands meaningfully over the full complexity range.

4.1.6.2 TNS for Group 2_CMS

Turning to Group 2_CMS, the complexity ordering is effectively reversed (Complex ⟶ Middle ⟶ Simple), so we look at how TNS evolves under these changing demands.

Simple ⟶ Middle

The line graph shows minimal differences, and the p-value of .4858 is well above .05, so there is no statistical significance. Participants produce roughly the same total syllables at these two levels, suggesting that moderate changes in task difficulty do not reliably alter Group 2's total speech performance.

Middle ⟶ Complex

Looking at the Middle-to-Complex contrast, the line graph suggests a slight increase, though p = .7538 also far exceeds .05, confirming no significance. Hence, we find no strong evidence that raising complexity beyond a moderate level leads to a systematic upward in TNS. It appears that for many participants, the total syllable count remains relatively stable or unpredictable across these complexities.

Simple ⟶ Complex

Lastly, comparing the extremes in Group 2 yields p = .6867, also not significant. Despite the large nominal difference in complexity from Simple to Complex, participants do not show a consistent shift in TNS. Possibly, their baseline approach to speaking remains the same or is overshadowed by other performance factors such as controlling for accuracy, or dealing with syntactic complexity.

In sum, no pairwise contrast among Simple, Middle, and Complex tasks yields a significant TNS difference for Group 2. The line graph might indicate small ups and downs, but none approach the standard significance threshold, suggesting that total syllables in Group 2's speech remains effectively unaffected by the tasks' complexity changes. This group thus seems to maintain a rather stable phonetic output volume, regardless of whether tasks are simpler or more demanding.

4.1.6.3 TNS for Group 3_MCS

Finally, Group 3_MCS faces tasks in the sequence Middle ⟶ Complex ⟶ Simple. We ask whether TNS shifts in a meaningful way with the group's varied complexity levels.

Simple ⟶ Middle

From Simple to Middle tasks, the line graph illustrates a noticeable upward trend in TNS, reflecting enhanced productivity. This aligns with the statistical findings, where p=.2315—above the conventional 0.05 threshold—suggests a no significance. Although not statistically significant, the moderate positive trend supports a steady improvement.

Middle ⟶ Complex

Between Middle and Complex tasks, the line graph shows relative stability in TNS, with modest increase observed. This observation is consistent with the statistical analysis, where p=.7133 reflects a no significance, and the small positive trend indicates limited variation in syllable production.

Simple ⟶ Complex

Examining the extremes yields p = .1953, also greater than .05, so again not significant. Despite the line graph's suggestion of a possible difference, it is not large or consistent enough to pass the threshold. Participants may or may not speak more syllables in the Complex tasks relative to the Simple ones, but the net effect is inconclusive.

Across all comparisons, none produce a significant shift in TNS for Group 3. The line graph typically indicates small or moderate fluctuations, yet the statistics do not confirm them as reliable changes. Overall, Group 3's total syllable count appears stable under both moderate and high complexities, and finishing on a simpler task does not result in any notable difference in overall phonetic output. Hence, the TNS metric does not show clear sensitivity to complexity variation in this group.

4.1.7 Fluency_Number of Words Per Minute (WPM)

The Number of Words Per Minute (WPM) is a key metric for assessing speech rate, calculated by counting the number of words produced within a one-minute interval. This measure serves as a valuable indicator of spoken fluency, with higher WPM suggesting faster and more fluent speech, and lower WPM reflecting a slower, more measured delivery. WPM provides a straightforward yet effective gauge of production speed and fluency in language performance evaluations.

The following table presents the WPM results for all three groups across the three levels of Cognitive Task Complexity (CTC), offering insights into how task complexity influences speech rate and fluency.

NUMBER OF WORDS PER MIN FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	91.1	90.8	106.7	P1	153.85	134.6	90	P1	113.15	116.13	130
P2	101.6	116.7	132.0	P2	118.32	110.87	75.71	P2	111.9	108.33	113.12
P3	125.5	105.7	132.0	P3	90.9	79.37	53.02	P3	130.9	106.83	110.91
P4	101.2	118.2	122.0	P4	107.78	62.62	36.31	P4	105.3	63.72	67.94
P5	82.1	103.3	90.0	P5	104	99.48	94.09	P5	91.5	103.8	89.5
P6	91.0	95.0	121.0	P6	91.47	88.45	63.53	P6	72.88	60.87	67.06
P7	115.1	118.3	102.0	P7	110.53	117.33	95	P7	44	56	54.55
P8	51.3	74.1	73.0	P8	71.05	81.74	77.06	P8	125	81.82	101.83
P9	96.5	90.2	115.0	P9	92	62.81	67.2	P9	109.52	89.09	86.14
P10	86.2	81.1	61.0	P10	87.5	79.73	64	P10	103.15	106.86	111.55
P11	125.5	93.5	99.0	P11	132.05	104.54	109.09	P11	112.31	104.07	108.21
P12	93.0	85.1	78.0	P12	104.06	122	120	P12	113.41	84.85	76.06
P13	91.5	133.3	162.0	P13	127.21	101.92	103.75	P13	105.6	106.93	103.3
P14	114.2	72.6	88.0	P14	113.33	105.26	86.67	P14	57.95	73.75	76.67
P15	100.1	97.8	89.1	P15	92.50	66.21	101.20	P15	102.98	117.12	117.99
P16	114.8	73.9	112.0	P16	111.30	111.40	70.99	P16	66.93	91.04	90.99
P17	110.0	88.1	119.7	P17	98.14	83.80	66.50	P17	86.08	64.89	85.14
P18	87.6	107.9	123.1	P18	117.65	76.30	69.06	P18	111.14	77.09	90.10
P19	64.6	96.3	78.2	P19	101.30	112.09	89.33	P19	105.88	118.04	80.09
P20	109.8	126.7	114.5	P20	123.73	128.70	93.87	P20	125.05	72.94	91.08
AVE	97.63	98.43	105.91	AVE	107.434	96.461	81.319	AVE	99.732	90.209	92.612

Figure 15 Words Per Minute for All Three Groups

The following line graphs visually depict the trends in fluency, as measured by the Words per Minute, for the three groups across the simple, middle, and complex tasks. The graph provide a clear representation of how phonetic output varies with increasing levels of cognitive task complexity.

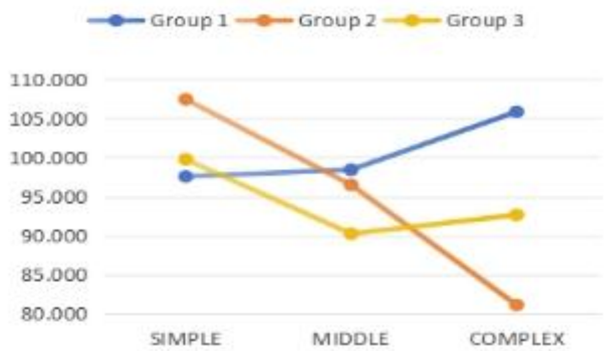

Figure 16 Words Per Minute Trend for All Three Groups

4.1.7.1 WPM for Group 1_SMC

Analyzing Group 1_SMC, we look at how Words Per Minute (WPM) changes when task complexity (CTC) moves from Simple to Middle to Complex. The line graph suggests a modest upward trend, but we must check if any pairwise differences reach statistical significance:

Simple ⟶ Middle

The shift from Simple to Middle tasks shows only a minimal rise in WPM. A p-value of .9035 lies well above the .05 threshold, meaning there is no statistically significant difference here. While the line graph might depict a small positive slope, the data do not indicate a reliable jump in speaking rate from a lower to a moderate level of complexity.

Middle ⟶ Complex

From Middle to Complex tasks, we see a slight continued uptick in WPM, but p = .4042 also fails to meet significance. Thus, even though some participants may speed up their speech slightly under higher demands, the group overall does not exhibit a consistent, measureable gain in WPM that clears the .05 cutoff.

Simple ⟶ Complex

Comparing the extremes—Simple vs. Complex—yields p = .3607, again above .05, so no significant difference is detected. Despite the line graph's indication of a modest increase across the full range, the data do not show it to be robust or uniform enough to conclude significance.

Overall, none of the complexity-level comparisons produce a significant effect on WPM for Group 1. Though the line graph might show a gentle upward slope, all p-values are above .05. We can infer that Group 1 participants do not systematically accelerate their speaking speed in response to rising complexity—at least not in a way that yields a statistically reliable difference.

4.1.7.2 WPM for Group 2_CMS

Shifting focus to Group 2_CMS, we assess whether participants' speech rate, measured by WPM, varies across tasks arranged in Complex ⟶ Middle ⟶ Simple order.

Simple ⟶ Middle

The data reveal a p-value of .1870, which is above .05 but under .2, sometimes viewed as "showing a possible but not significant trend". Statistically, we interpret it as not significant. The line graph might point to a moderate downward shift, but the data do not firmly confirm a reliable difference in WPM between these two complexities.

Middle ⟶ Complex

From Middle to Complex tasks, the line graph indicates a relatively sharper drop in WPM, with p = .0801. This is still above .05, albeit not by a large margin. Some researchers might term this marginal or approaching significance, but within standard criteria, we say it is not conclusively significant. That said, it suggests that participants might speak more slowly at the highest complexity, although not with enough consistency for a strong claim.

Simple ⟶ Complex

A direct comparison of Simple vs. Complex yields p = .0041, which is below .05, clearly significant. This means that, overall, participants exhibit a notable decrease in WPM when tasks jump from the simplest to the most demanding. The line graph strongly corroborates this gap, implying that at the extremes, Group 2 slows down speech rate under high task complexity, likely due to increased cognitive load or attentional shifts (e.g., focusing on accuracy or lexical choices).

For Group 2, the only significant effect on WPM emerges when comparing the simplest and the most complex tasks. Intermediate steps (Simple⟶Middle or Middle⟶Complex) individually fail to clear .05, but the overall S–C difference is robust. This pattern suggests participants might maintain or slightly adjust their speech rate at intermediate complexities but substantially slow down at the highest level, producing a significant net difference between the easiest and hardest tasks.

4.1.7.3 WPM for Group 3_MCS

Lastly, Group 3_MCS tackles tasks in the order Middle ⟶ Complex ⟶ Simple, and the question is whether WPM changes significantly across these varied complexities.

Simple ⟶ Middle

Although the line graph might show a drop, the relevant p-value is .2767, which is above .05, indicating no statistical significance. Consequently, we cannot assert that participants speak notably slower at Middle-level tasks compared to the simplest tasks, at least not in a consistent, group-wide manner.

Middle ⟶ Complex

Moving from Middle to Complex tasks, the line graph suggests a mild upward shift, but p = .7644 also well exceeds the .05 cutoff, confirming no significant difference. The group, on average, does not reliably alter speech rate in response to the higher complexity. Some individuals may attempt to speak faster or slower, but the effect does not manifest strongly at the group level.

Simple ⟶ Complex

Directly comparing the simplest and most complex tasks yields p = .4295, again not meeting standard significance. Although the line graph might show a modest overall difference, the data are too variable or the effect too small to claim a clear shift in WPM. Group 3 thus does not exhibit any robust evidence of speed changes from the easiest to the hardest tasks.

For Group 3, none of the pairwise comparisons show a significant difference in WPM, paralleling the scenario seen with some other performance metrics in this group. While the line graph indicates minor fluctuations between Middle, Complex, and Simple tasks, the statistical results confirm these variations do not exceed chance. Hence, Group 3 participants' speech rate remains relatively unaffected by changes in cognitive task complexity.

4.1.8 Fluency_Number of Syllables Per Second (SPS)

The number of syllables per second is a measure used to assess the rate of speech by calculating how many syllables are spoken in one second. This metric is commonly used in linguistic analysis and speech studies to evaluate fluency, articulation speed, and overall speech tempo. A higher number of syllables per second typically indicates faster speech, which can be associated with greater fluency or more rapid articulation. Conversely, a lower number of syllables per second may suggest slower, more deliberate speech.

The following table presents the SPS results for all three groups across the three levels of Cognitive Task Complexity (CTC), offering insights into how task complexity influences speech rate and fluency.

NUMBER OF SYLLABLES PER SEC FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX		SIMPLE	MIDDLE	COMPLEX
P1	2.5	2.12	3.06	P1	2.74	2.89	1.85	P1	2.66	2.32	2.8
P2	2.6	3.43	3.46	P2	2.03	1.66	1.59	P2	3	3.71	3.07
P3	2.65	3.02	3.47	P3	2.95	2.63	1.61	P3	2.77	2.13	2.55
P4	2.17	2.7	2.86	P4	2.16	1.59	1.02	P4	2.6	1.6	1.55
P5	1.66	2.28	1.94	P5	2.33	2.06	1.77	P5	1.97	2.45	2.12
P6	1.62	2.2	2.63	P6	1.74	1.47	1.28	P6	1.99	1.43	1.59
P7	2.3	2.49	2.32	P7	2.42	2.18	1.8	P7	0.92	1.28	1.21
P8	1.02	1.85	1.79	P8	1.89	1.53	1.6	P8	2.64	2.03	2.32
P9	1.81	2.29	2.49	P9	1.58	1.33	1.23	P9	2.16	1.94	2.06
P10	1.62	1.88	1.34	P10	1.91	1.73	1.41	P10	2.59	2.4	3.07
P11	2.56	1.94	2.59	P11	2.94	1.97	2	P11	2.85	2.23	2.84
P12	1.08	2.04	1.65	P12	2.09	2.25	2.61	P12	2.33	1.89	1.91
P13	1.81	2.8	2.88	P13	2.59	1.98	2.06	P13	2.29	2.74	3.01
P14	1.79	1.44	1.93	P14	2.54	2.22	1.75	P14	1.75	1.37	2.11
P15	1.73	2.58	1.68	P15	2	2.49	1.66	P15	3.25	2.02	2.98
P16	3	1.73	3.21	P16	2.87	2.67	2.07	P16	2.39	3.46	2.13
P17	3.02	1.54	2.18	P17	2.84	2.26	3.09	P17	2.03	2.45	1.62
P18	3.45	1.93	1.53	P18	2.72	3.03	3.21	P18	3.49	1.51	2.68
P19	1.52	2.87	3.48	P19	3.26	3.25	1.64	P19	2.37	3.48	2.11
P20	2.64	1.86	2.37	P20	2.37	1.57	2.59	P20	1.57	2.88	2.55
AVE	2.128	2.250	2.443	AVE	2.399	2.138	1.892	AVE	2.381	2.266	2.314

Figure 17 Syllables Per Second for All Three Groups

The following line graphs visually depict the trends in fluency, as measured by the Syllables per Second, for the three groups across the simple, middle, and complex tasks. The graph provide a clear representation of how phonetic output varies with increasing levels of cognitive task complexity.

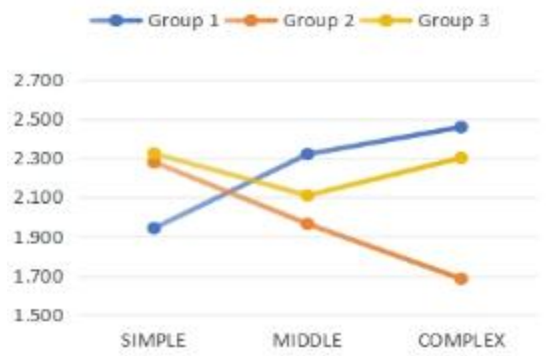

Figure 18 Syllables per Second Trends for Three Groups

4.1.8.1 SPS for Group 1_SMC

For Group 1_SMC, we evaluate whether Syllables Per Second (SPS) changes as tasks become more cognitively demanding, moving from Simple to Middle to Complex. The line graph indicates a moderate upward slope, suggesting participants might speak faster under higher complexity, but we must see if pairwise comparisons are statistically significant:

Simple ⟶ Middle

The jump from Simple to Middle tasks corresponds to $p = .0711$, which is slightly above .05 and thus not conventionally significant. However, it lies in a range some researchers call “borderline” or “marginal”, hinting at a possible upward trend that does not quite clear standard criteria. One interpretation is that participants may have begun to speed up speech somewhat at moderate complexity, though not consistently enough to confirm.

Middle ⟶ Complex

When shifting from Middle to Complex, the line graph shows very little upward in SPS, and $p = .5470$ is well above .05, confirming no significant difference. This result suggests that once tasks pass a certain moderate threshold, participants in Group 1 do not further accelerate speech rate at the highest complexity; they might channel attentional resources to other linguistic dimensions.

Simple ⟶ Complex

Comparing the simplest to the most demanding tasks, $p = .0321$ is below .05, indicating a significant difference. Overall, participants produce more syllables per second under Complex tasks than Simple ones, presumably reflecting that some measure of "rising challenge" fosters a net boost in speech rate across the broad gap. The stepwise contrasts alone were not both significant, but the entire range from S to C is large enough to detect.

Although the Simple–Middle increment does not strictly meet significance ($p \sim .07$) and Middle–Complex fails to show significance, the Simple–Complex comparison is significant. This suggests participants speed up over the course of the entire complexity range, but the effect is only statistically evident when comparing the two extremes. Possibly, moderate tasks elicit partial acceleration, while going beyond moderate complexity does not yield further incremental gains.

4.1.8.2 SPS for Group 2_CMS

Group 2_CMS encounters tasks in the reverse order—Complex ⟶ Middle ⟶ Simple—so the question is whether participants modulate their speech rate differently under changing complexity. The line graph shows a declining slope overall, implying a slowdown at higher complexity, but we check each comparison:

Simple ⟶ Middle

The p-value is .0681, somewhat above .05, meaning it does not reach conventional significance. One could label it a borderline or marginal effect. While the line graph might illustrate a downward shift (or slight difference), the result does not confirm it as a robust distinction in speech rate.

Middle ⟶ Complex

From Middle to Complex, the data yield $p = .0954$, again above .05, so not significant. It suggests a continued drop in SPS for some participants but not uniform enough to surpass the threshold. In short, Group 2 does not show a reliably consistent slowdown from moderate to complex tasks, even though the line graph hints at it.

Simple ⟶ Complex

The most striking difference arises when directly comparing the simplest to the most demanding tasks, with p = .0008 clearly well below .01. This indicates a significant drop in syllables per second. In other words, Group 2 participants speak significantly more slowly under high-complexity tasks compared to the simplest tasks, which aligns with the notion that they face heavier cognitive burdens and thus reduce speech rate to cope.

Similar to some earlier patterns, Group 2's stepwise changes do not individually meet .05 significance from Simple⟶Middle or Middle⟶Complex, though each is near .07–.09. However, the overall gap from S to C is large enough that the difference in SPS is significant. We interpret that as participants drastically slowing their speech at the highest complexity, producing a robust difference between the simplest and the hardest conditions, even if partial increments are not each validated statistically.

4.1.8.3 SPS for Group 3_MCS

Finally, Group 3_MCS addresses tasks in Middle ⟶ Complex ⟶ Simple. The line graph depicts slight dips or rises in SPS, but the pairwise p-values clarify significance:

Simple ⟶ Middle

We can compare these two complexities. The test yields p = .3162, above .05, meaning no significant difference. Visually, the line may show a moderate shift, but it remains insufficient to conclude a stable effect on speech rate for Group 3 between these levels.

Middle ⟶ Complex

Transitioning from moderate to high complexity yields p = .3883, also not significant. This result indicates Group 3 does not reliably speed up or slow down in response to the higher demands at Complex tasks. Some participants may show small changes, but the overall pattern is not robust.

Simple ⟶ Complex

Directly comparing the simplest and the hardest tasks yields p = .6200, again well above .05, indicating no significance. Group 3's line graph display a near-flat or slightly changing slope, but the data do not confirm any real difference in SPS between the two extremes of complexity.

In sum, none of the three comparisons (Simple⟶Middle, Middle⟶Complex, or Simple⟶Complex) achieves significance. This points to a generally stable or inconsistent approach to speech rate for Group 3; raising or lowering complexity does not elicit a large enough effect on syllables per second to surpass chance variation. Hence, Group 3's fluency in terms of SPS remains effectively the same across all levels of task difficulty.

4.1.9 Summary and Implications for *Task-Level Effect*

Based on the data results presented above across these sub-dimensions of oral performance, we can now delve deeper to gain a clearer understanding of the underlying patterns and insights revealed by the data, and how the results match with the predictions of the Cognition Hypothesis and Limited Attentional Capacity models. The following table presents the fluctuation for all language performance dimensions across three groups (Task-Level Effect), (Simple-Middle, Middle-Complex, Simple-Complex).

	TTR	MLAS	MLC	TNW	TNS	WPM	SPS	EFCR
Group 1	(-,-,-)	(+,-,+)	(+,-,+)	(+,+,+)	(+,+,+)	(+,+,+)	(+,+,+)	(+,+,+)
Group 2	(+,+,+)	(-,+,-)	(+,-,-)	(=,+,+)	(-,+,-)	(-,-,-)	(-,-,-)	(-,+,=)
Group 3	(+,+,+)	(+,-,-)	(+,+,+)	(+,+,+)	(+,+,+)	(-,+,-)	(-,+,-)	(+,+,+)

Cognitive Task Complexity & L2 Performance (Task-Level Effect)

Group 1: This pattern strongly supports the Cognition Hypothesis while partially reflecting attention limits at the highest task level. The Cognition Hypothesis, which posits that increased task complexity leads to greater syntactic complexity and lexical variety, aligns with Group 1's steady improvement in fluency, accuracy, and syntactic complexity across tasks. This group shows significant growth in MLAS, MLC, and EFCR, particularly between simpler and more complex tasks, indicating that participants were able to enhance their syntactic and lexical performance as task complexity increased. This trend suggests that Group 1 participants may have been able to allocate attentional resources effectively even as cognitive demands grew, supporting the Cognition Hypothesis. However, the observed decrease in TTR and a slight decline in complexity metrics at the highest task level hints at the limitations imposed by attentional capacity, where lexical variety and other dimensions may experience some constraints due to cognitive load.

Group 2: Group 2's performance validates the Limited Attentional Capacity theory, showing that once complexity increases, the group prioritizes maintaining basic performance over syntactic or lexical enhancements. Group 2's sharp declines in WPM, SPS, and significant challenges in maintaining TTR and MLC at the complex level support the Limited Attentional Capacity theory. As task complexity rose, this group struggled to sustain fluency and syntactic complexity, suggesting that their attentional resources were increasingly taxed. The stability observed in TNW but the decline in TNS indicates that while participants may have maintained overall productivity, the complexity of their oral production suffered, aligning with the theory that limited cognitive resources restrict the ability to enhance complexity and fluency simultaneously. This trend suggests that as cognitive demands increase, Group 2 may prioritize maintaining a basic oral production level rather than achieving higher syntactic complexity or fluency.

Group 3: Overall, the group's balanced adaptation suggests a hybrid outcome, partially consistent with Cognition Hypothesis but tempered by practical attention redistribution. Group 3 demonstrates a unique pattern of gradual improvement in productivity and accuracy but fluctuates in fluency, particularly with a dip in the middle complexity level and recovery at the complex level. The increase in TTR and MLC, though not strongly significant, suggests an attempt to improve lexical and syntactic variety as task complexity rose, partially aligning with the Cognition Hypothesis. However, the low confidence in fluency metrics, along with Group 3's moderate success in complexity, may imply a strategic redistribution of attentional resources. Group 3's gradual improvements suggest a balance between cognitive demands and attentional resources, adapting with minor fluctuations but without substantial decreases in performance.

In summary, Group 1 largely aligns with the Cognition Hypothesis, Group 2 highlights the constraints of limited attentional capacity under higher cognitive loads, and Group 3 demonstrates a balanced, adaptive approach. These results underscore the nuanced relationship between cognitive demands and attentional resources, with each group responding in distinct ways.

Building on these insights into complexity-driven patterns, the next section shifts attention to the role of task sequence—examining how different orders of task presentation may likewise influence L2 performance dynamics.

4.2 Dynamic Changes in Speech Data-Along Task Sequence

This section presents the results of the CALF dimensions for this experiment, alongside the corresponding p-values, emphasizing the dynamic changes observed in response to varying task sequences. These findings provide a comprehensive understanding of the impact of task order on L2 performance. The calculated p-values and visualized charts will be explained concisely, highlighting their statistical significance and revealing potential relationships between task sequence and language performance, while laying the groundwork for further detailed analysis.

Note: In this section,

In Group 1, Task 1 = Simple, Task 2 = Middle, Task 3 = Complex;

In Group 2, Task 1 = Complex, Task 2 = Middle, Task 3 = Simple;

In Group 3, Task 1 = Middle, Task 2 = Complex, Task 3 = Simple.

4.2.1 Lexical Complexity_TTR

The following table presents the TTR results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences lexical variety.

TTR FOR THREE GROUPS_Sequence											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	Task 1	Task 2	Task 3		Task1	Task 2	Task 3		Task 1	Task 2	Task 3
P1	0.768	0.670	0.509	P1	0.667	0.597	0.5	P1	0.679	0.708	0.619
P2	0.684	0.750	0.544	P2	0.638	0.455	0.625	P2	0.662	0.513	0.546
P3	0.697	0.700	0.712	P3	0.623	0.624	0.616	P3	0.677	0.607	0.667
P4	0.676	0.815	0.518	P4	0.842	0.819	0.81	P4	0.635	0.714	0.59
P5	0.771	0.6	0.627	P5	0.681	0.64	0.635	P5	0.53	0.584	0.514
P6	0.588	0.63	0.550	P6	0.741	0.55	0.5	P6	0.619	0.645	0.558
P7	0.465	0.71	0.603	P7	0.695	0.76	0.64	P7	0.567	0.587	0.511
P8	0.875	0.48	0.627	P8	0.580	0.77	0.55	P8	0.511	0.6	0.613
P9	0.613	0.578	0.465	P9	0.524	0.57	0.565	P9	0.694	0.826	0.652
P10	0.804	0.635	0.488	P10	0.516	0.63	0.612	P10	0.438	0.591	0.45
P11	0.840	0.76	0.59	P11	0.638	0.57	0.563	P11	0.447	0.459	0.493
P12	0.692	0.57	0.463	P12	0.698	0.67	0.661	P12	0.589	0.556	0.505
P13	0.895	0.66	0.565	P13	0.687	0.6	0.496	P13	0.565	0.478	0.381
P14	0.743	0.58	0.613	P14	0.615	0.69	0.632	P14	0.559	0.435	0.726
P15	0.652	0.871	0.488	P15	0.602	0.532	0.479	P15	0.701	0.578	0.518
P16	0.759	0.887	0.623	P16	0.827	0.518	0.781	P16	0.744	0.574	0.703
P17	0.795	0.512	0.589	P17	0.480	0.903	0.516	P17	0.727	0.621	0.536
P18	0.699	0.577	0.570	P18	0.805	0.607	0.572	P18	0.437	0.643	0.668
P19	0.608	0.530	0.577	P19	0.442	0.742	0.699	P19	0.479	0.576	0.563
P20	0.821	0.539	0.527	P20	0.763	0.532	0.556	P20	0.419	0.565	0.366
AVE	0.722	0.653	0.562	AVE	0.653	0.639	0.600	AVE	0.584	0.593	0.559

Figure 19 TTR for All Three Groups (along task sequence)

The following line graphs illustrate the trends in lexical complexity, measured by the TTR, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic output fluctuates with task processing.

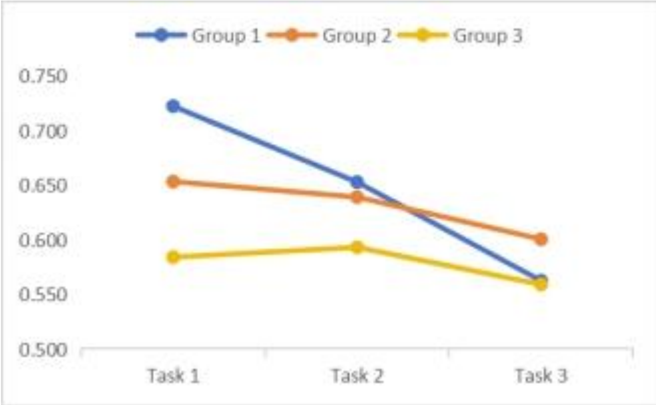

Figure 20 TTR Trends for All Three Groups (along task sequence)

Group 1 (Task Sequence: S ⟶ M ⟶ C)

From the table, p = .0013 indicates a significant effect of task sequence on TTR across the three tasks in Group 1. Specifically, TTR values decline from Task 1 (Simple) to Task 2 (Middle) to Task 3 (Complex), suggesting that as participants progress through this ascending complexity order, their lexical variety gradually diminishes. The line graph confirms a downward trajectory, implying participants might devote more attention to structural or accuracy-related aspects in later, harder tasks, thereby reducing the range of vocabulary they employ.

Overall, the sequence from simpler to more complex tasks correlates with a steady reduction in TTR, which aligns with the idea that cumulative cognitive demands can gradually constrain lexical variety.

Group 2 (Task Sequence: C ⟶ M ⟶ S)

For Group 2, p = .0043 also lies below .05, indicating a significant effect of task sequence on TTR. However, the direction differs from Group 1. The average TTR values reveal a slight downward trend from Task 1 (Complex) through Task 2 (Middle) to Task 3 (Simple), though each step's decline may not be large. One interpretation is that participants initially face high complexity and thus might start with somewhat narrower lexical usage; as tasks become easier (Middle, then Simple), they could remain cautious or fatigued, not necessarily expanding their vocabulary.

This pattern could be partially attributed to the group's adaptation or possible fatigue. Even though tasks lighten in complexity, participants do not show an increase in lexical diversity. The overall result is statistically significant across the three tasks, suggesting that the unique sequence might lead to a net decline in TTR.

Group 3 (Task Sequence: M ⟶ C ⟶ S)

In Group 3, the overall effect is p = .1423, which does not meet the .05 threshold, thus not significant. The line graph shows TTR values that fluctuate slightly from Task 1 (Middle) to Task 2 (Complex) to Task 3 (Simple), but these changes are insufficient to be declared reliable. Although participants might display small increases or decreases in lexical variety, the variability across individuals prevents a clear directional conclusion.

Hence, for Group 3, any observed changes in TTR along the sequence appear more random or modest, failing to yield a statistically significant overall effect.

Across all three groups, TTR shows a downward trend as tasks progress, indicating that task sequence noticeably affects lexical performance. Group 1 shows a significant TTR decrease across tasks, consistent with progressive lexical constraints from S⟶M⟶C. Group 2 also has a significant effect, featuring a mild downward slope from C⟶M⟶S. Group 3 does not show a significant overall shift, implying stable or inconsistent TTR across M⟶C⟶S.

4.2.2 Syntactic Complexity_Mean Length of AS-unit

The following table presents the MLAS results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences syntactic variety.

MEAN LENGTH OF AS-UNIT FOR THREE GROUPS_sequence											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	Task 1	Task 2	Task 3		Task 1	Task 2	Task 3		Task 1	Task 2	Task 3
P1	6.12	6.87	9.12	P1	7.8	9.75	6.6	P1	5.8	5.73	6.3
P2	5.14	5.91	9.44	P2	9.64	6.41	9.33	P2	12.9	7.69	8.08
P3	4.75	7.5	6.55	P3	6.73	6.38	7.67	P3	7	6.88	4.36
P4	6.45	6.1	5.46	P4	3.8	3.91	4.89	P4	7.44	7.1	5.26
P5	3.54	4.69	6.44	P5	5.75	5.36	5.56	P5	9.39	7.55	7
P6	5.13	10.5	5.79	P6	6	7.2	8.15	P6	6.15	5.46	7.17
P7	7.11	6.88	6.1	P7	5.94	5.06	7.87	P7	5.96	7.88	6.14
P8	4.35	8	6.58	P8	5.7	5.15	6.39	P8	9.63	6.67	6.25
P9	6.5	4.71	9.73	P9	4.67	4.92	4.36	P9	5.22	3.67	3.54
P10	4.55	6.25	8.47	P10	9.14	7.13	6.36	P10	7.19	7.5	8.48
P11	6.15	8.3	6.6	P11	4.71	5.75	7.46	P11	9.38	8.48	9.73
P12	6.73	7.33	4.61	P12	5.5	5.48	5.97	P12	9	8.05	7.75
P13	4.28	6.33	5.29	P13	6.92	5.29	8.38	P13	6.4	8.81	8.42
P14	5.25	7.29	5.29	P14	9.1	4.35	5	P14	8.29	7.22	4.71
P15	5.383	4.307	7.209	P15	8.55	6.06	5.41	P15	9.09	7.48	5.00
P16	5.15	6.913	5.285	P16	7.61	3.38	6.24	P16	5.42	4.97	4.56
P17	4.80	4.96	6.92	P17	4.07	5.62	8.09	P17	10.28	5.98	4.53
P18	5.62	8.76	7.11	P18	5.24	7.16	5.42	P18	7.27	9.37	6.59
P19	5.87	8.63	7.27	P19	6.43	5.59	6.04	P19	5.32	7.45	9.70
P20	5.76	7.85	7.12	P20	7.28	7.40	9.09	P20	9.66	7.04	9.55
AVE	5.432	6.904	6.819	AVE	6.529	5.867	6.714	AVE	7.839	7.049	6.656

Figure 21 MLAS for All Three Groups (along task sequence)

The following line graphs illustrate the trends in syntactic complexity, measured by the MLAS, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic performance fluctuates with task processing.

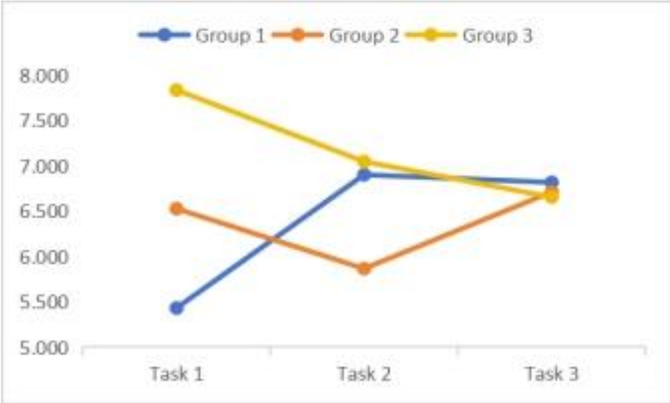

Figure 22 MLAS Trends for All Three Groups (along task sequence)

Group 1 (Task Sequence: S ⟶ M ⟶ C)

According to the statistical results, p = .0018 is below .05, implying a significant overall effect of task sequence on MLAS. The line graph indicates a substantial jump from Task 1 (S) to Task 2 (M), followed by a slight decrease in Task 3 (C), but still well above Task 1's baseline. Thus, the ascending sequence fosters a notable peak in syntactic complexity at the middle task, though the final (most complex) stage sees a small fallback. The net effect across the three tasks is significant, demonstrating how the interplay of growing familiarity and rising demands shapes MLAS.

Group 2 (Task Sequence: C ⟶ M ⟶ S)

For Group 2, p = .1165 is above .05, so the overall effect of task sequence on MLAS is not significant. Despite the line graph possibly showing a dip from Task 1 (Complex) to Task 2 (Middle) followed by a partial recovery at Task 3 (Simple), the variations do not meet significance criteria. Therefore, we conclude that there is no statistically consistent pattern in how Group 2's syntactic complexity changes across tasks ordered C⟶M⟶S.

Group 3 (Task Sequence: M ⟶ C ⟶ S)

In Group 3, p = .0351 is below .05, indicating a significant influence of task sequence on MLAS. The average values suggest a relatively high starting point at Task 1 (Middle), followed by a dip at Task 2 (Complex), and then a further decline in

Task 3 (Simple). Hence, Group 3 sees MLAS gradually decreasing through the sequence, with the differences collectively reaching significance over the three tasks.

Overall, the line graph illustrates distinct patterns for each group: Group 1: A significant effect, with MLAS peaking around Task 2 but still higher than baseline by Task 3. Group 2: Shows no significant variation in MLAS across the C⟶M⟶S order. Group 3: Demonstrates a significant downward pattern from M to C to S.

4.2.3 Syntactic Complexity_Mean Length of Clause

The following table presents the MLC results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences syntactic variety.

MEAN LENGTH OF CLAUSE FOR THREE GROUPS_sequence											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	Task 1	Task 2	Task 3		Task 1	Task 2	Task 3		Task 1	Task 2	Task 3
P1	4.36	5.8	7.31	P1	3.9	4.72	5.14	P1	4.77	5.25	4.85
P2	5.04	4.51	4.23	P2	4.82	5.9	6.31	P2	6.89	6.41	4.41
P3	4.55	6.15	4.08	P3	4.61	6.00	6.08	P3	5.00	4.62	2.29
P4	5.25	5.08	4.86	P4	2.71	3.67	3.98	P4	4.62	4.81	4
P5	3.32	4.29	4.25	P5	2.76	4.69	5	P5	4.42	5.19	3.26
P6	5.07	7.85	5	P6	3.38	5.79	6.47	P6	5.88	5.07	3.07
P7	3.69	5.55	4.85	P7	3.95	4.05	4.91	P7	4.68	4.47	3.21
P8	3.71	6.33	5.85	P8	3.19	4.53	5.13	P8	6.21	5	2.88
P9	4.75	4.65	7.2	P9	2.55	4.57	3.05	P9	4.33	2.88	1.92
P10	2.36	5.56	4.08	P10	3.88	5.53	3.07	P10	5.34	5.92	3.95
P11	4.62	6.72	5.67	P11	3.48	4.19	5.72	P11	6.88	7.39	3.4
P12	5.03	6.21	4.24	P12	2.75	4.57	2.87	P12	5.14	5.96	3.44
P13	5.98	4.57	5.29	P13	2.77	4.59	3.89	P13	4	6.74	3.21
P14	2.45	5.18	4.7	P14	4.79	3.29	2.83	P14	4	4.81	2.79
P15	5.14	4.28	6.10	P15	2.92	5.38	3.75	P15	4.69	7.12	3.17
P16	3.92	6.21	4.37	P16	4.76	2.79	4.69	P16	4.25	4.56	3.39
P17	4.80	4.21	3.96	P17	4.07	4.39	4.92	P17	7.09	4.14	2.83
P18	3.72	5.86	5.58	P18	2.56	5.99	4.83	P18	6.82	4.95	2.85
P19	4.58	6.29	4.13	P19	3.78	4.80	4.28	P19	3.61	4.49	2.66
P20	3.63	6.77	6.55	P20	3.16	5.00	5.17	P20	4.46	6.68	5.10
AVE	4.299	5.604	5.115	AVE	3.539	4.721	4.604	AVE	5.154	5.323	3.334

Figure 23 MLC for All Three Groups (along task sequence)

The following line graphs illustrate the trends in syntactic complexity, measured by the MLC, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic performance fluctuates with task processing.

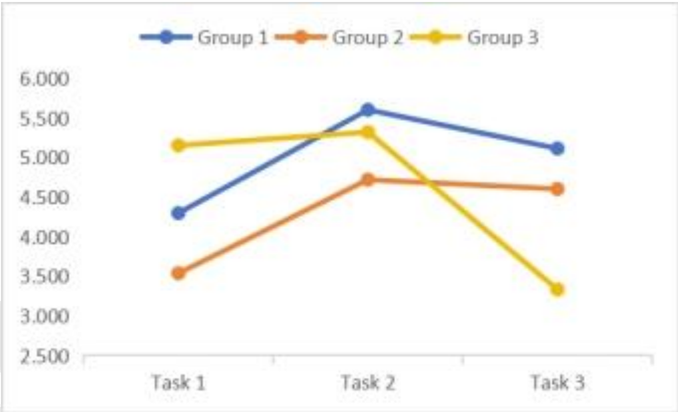

Figure 24 MLC Trends for All Three Groups (along task sequence)

Group 1 (Task Sequence: S ⟶ M ⟶ C)

In Group 1, p = .0224 lies below .05, suggesting a significant overall impact of task sequence on MLC. The line graph typically shows a strong rise from Task 1 (S) to Task 2 (M), followed by a slight decline by Task 3 (C). Hence, the overall three-task pattern is robust enough to be deemed significant, indicating participants do achieve longer clauses in the second stage, though the final complex task sees a mild fallback.

Group 2 (Task Sequence: C ⟶ M ⟶ S)

For Group 2, p = .0002 is well below .01, affirming a highly significant effect of sequence on MLC. The line graph highlights a sharp surge in MLC from Task 1 (Complex) to Task 2 (Middle), then a small dip at Task 3 (Simple), but still above the initial complex-level measure. Overall, the sequence from complex to simpler tasks ironically leads to a net increase in MLC across tasks, showing that stepping down in difficulty might free participants to produce more elaborate clauses, overshadowing any potential fatigue factor.

Group 3 (Task Sequence: M ⟶ C ⟶ S)

In Group 3, $p < .00001$ denotes an extremely significant effect of sequence on MLC. The line graph shows stability or a slight rise between Task 1 (M) and Task 2 (C), then a steep drop by Task 3 (S). As a result, comparing all three tasks reveals a strong cumulative difference. Consequently, the overall shape is a mild plateau or small gain followed by a drastic decline, which across the three tasks is highly significant. This suggests that while moderate or complex tasks encourage more elaborate clauses, shifting to the final simple task leads to notably shorter clauses, possibly indicating less motivation or a simpler approach as the sequence concludes.

Overall, all three groups exhibit a similar trend: clause length generally peaks during Task 2 before declining in Task 3, reflecting a consistent pattern shaped by task sequence. Group 1 shows a sharp increase then a slight decrease, Group 2 a clear rise then a modest dip, and Group 3 relative stability before a steep drop. While each group's magnitude of change differs, the collective results imply that task sequence plays a pivotal role in clause construction—likely due to initial improvement or adaptation in the middle task, followed by potential fatigue or constraints in the final stage.

4.2.4 Accuracy_Error-Free Clause Ratio

The following table presents the EFCR results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences accuracy.

ERROR-FREE CLAUSE RATIO FOR THREE GROUPS											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	Task 1	Task 2	Task 3		Task 1	Task 2	Task 3		Task 1	Task 2	Task 3
P1	0.923	0.867	0.823	P1	0.929	0.846	0.785	P1	0.7	0.865	0.818
P2	0.769	0.909	0.923	P2	0.812	0.824	0.842	P2	0.947	0.765	0.889
P3	0.8	0.9	0.916	P3	0.9	0.769	0.712	P3	0.737	0.832	0.627
P4	0.727	0.8	0.889	P4	0.9	0.818	0.75	P4	0.7	0.769	0.889
P5	0.846	0.75	0.8	P5	0.733	0.722	0.75	P5	0.731	0.9	0.695
P6	0.923	0.5	0.778	P6	0.857	0.917	0.818	P6	0.725	0.8	0.867
P7	0.818	0.845	0.9	P7	0.833	0.727	0.867	P7	0.782	0.833	0.652
P8	0.692	0.909	0.769	P8	0.75	0.9	0.913	P8	0.923	0.929	0.756
P9	0.538	0.786	0.824	P9	0.846	0.857	0.913	P9	0.727	0.778	0.9
P10	0.846	0.833	0.8	P10	0.813	0.923	0.909	P10	0.969	0.919	0.8
P11	0.75	0.909	0.925	P11	0.95	0.842	0.889	P11	0.913	0.965	0.725
P12	0.6	0.75	0.85	P12	0.857	0.786	0.778	P12	0.882	0.915	0.757
P13	0.75	0.833	0.857	P13	0.857	0.833	0.75	P13	0.963	0.906	0.725
P14	0.789	0.833	0.8	P14	0.933	0.894	0.944	P14	0.825	0.904	0.944
P15	0.829	0.892	0.92	P15	0.79	0.8	0.775	P15	0.675	0.941	0.88
P16	0.734	0.695	0.65	P16	0.83	0.645	0.865	P16	0.88	0.81	0.9
P17	0.63	0.65	0.815	P17	0.64	0.675	0.78	P17	0.69	0.74	0.87
P18	0.933	0.75	0.85	P18	0.75	0.62	0.724	P18	0.69	0.925	0.86
P19	0.751	0.82	0.875	P19	0.65	0.665	0.85	P19	0.925	0.895	0.645
P20	0.701	0.845	0.790	P20	0.893	0.903	0.94	P20	0.865	0.725	0.6
AVE	0.767	0.804	0.838	AVE	0.826	0.798	0.828	AVE	0.812	0.856	0.790

Figure 25 EFCR for All Three Groups (along task sequence)

The following line graphs illustrate the trends in accuracy, measured by the EFCR, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic performance fluctuates with task processing.

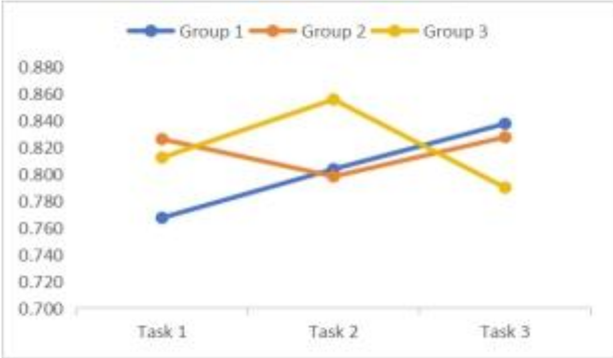

Figure 26 EFCR Trends for All Three Groups (along task sequence)

Group 1 (Sequence: S ⟶ M ⟶ C)

Based on the analysis, p = .041 lies below .05, indicating a significant impact of task sequence on EFCR for Group 1. The line graph shows a gradual rise in EFCR from Task 1 (Simple) to Task 2 (Middle) to Task 3 (Complex). Specifically: Task 1 yields the lowest EFCR, consistent with participants' initial adaptation phase, Task 2 sees a moderate climb, suggesting some improvement in accuracy as participants become more attuned to the task demands, Task 3 records the highest EFCR, implying a culmination of practice and increasing comfort even though the tasks are the most complex at this point.

Overall, this ascending order (S⟶M⟶C) seems to foster incremental improvements in clause-level accuracy, potentially due to combined familiarity and the impetus provided by rising cognitive demands.

Group 2 (Sequence: C ⟶ M ⟶ S)

For Group 2, p = .638 is well above .05, hence not significant. The average EFCR values show only minor fluctuations from Task 1 (Complex) to Task 2 (Middle) to Task 3 (Simple). Although the line graph may display small ups and downs, these differences fail to reach any conventional threshold. We can conclude that, for Group 2, the order of tasks—starting with a complex one and ending with a simple one—does not reliably affect EFCR. Participants' accuracy level remains relatively stable across the sequence.

Group 3 (Sequence: M ⟶ C ⟶ S)

In Group 3, p = .247 also exceeds .05, indicating no statistically significant effect of task sequence on EFCR. The line graph suggests a slight fluctuation, perhaps peaking around Task 2 (Complex) before dipping again at Task 3 (Simple), but the changes do not yield a firm conclusion. Thus, Group 3's error-free clause ratio remains effectively the same on average, regardless of whether tasks progress from moderate to complex and finally to simple.

Overall, these EFCR values reveal distinct patterns across the three groups. Group 1: Shows a significant increase in EFCR across S⟶M⟶C, implying participants steadily improve accuracy over the task series. Group 2: No significant shift; EFCR stays roughly constant despite going from hardest to easiest tasks. Group 3: Also no significant difference, pointing to stable clause-level accuracy across M⟶C⟶S ordering.

4.2.5 Productivity_Total Number of Words

The following table presents the TNW results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences productivity.

TOTAL NUMBER OF WORDS FOR THREE GROUPS											
	Group1_SMC				Group 2_CMS				Group 3_MCS		
	Task1	Task 2	Task 3		Task 1	Task 2	Task 3		Task 1	Task 2	Task 3
P1	82	112	114	P1	78	83	75	P1	99	139	92
P2	76	72	96	P2	106	123	125	P2	122	153	64
P3	67	74	64	P3	111	85	64	P3	69	127	81
P4	71	65	105	P4	38	45	48	P4	130	144	84
P5	48	62	50	P5	69	58	51	P5	75	73	78
P6	85	70	114	P6	108	115	110	P6	138	115	113
P7	71	66	60	P7	95	88	106	P7	83	106	127
P8	56	83	81	P8	131	64	120	P8	91	110	88
P9	93	83	114	P9	84	67	62	P9	135	94	96
P10	57	104	114	P10	128	109	90	P10	95	81	120
P11	75	83	69	P11	80	122	100	P11	115	118	99
P12	104	122	107	P12	66	65	67	P12	103	101	123
P13	38	33	40	P13	83	90	108	P13	111	123	106
P14	113	122	111	P14	91	86	73	P14	119	148	71
P15	57	100	97	P15	97	116	78	P15	80	131	67
P16	88	73	116	P16	112	108	68	P16	88	98	103
P17	84	70	96	P17	73	64	78	P17	72	89	116
P18	101	61	53	P18	86	80	138	P18	107	86	74
P19	53	113	58	P19	78	74	81	P19	143	77	61
P20	60	76	111	P20	105	79	80	P20	127	136	109
AVE	73.95	82.20	88.50	AVE	90.95	86.05	86.10	AVE	105.10	112.45	93.60

Figure 27 TNW for All Three Groups (along task sequence)

The following line graphs illustrate the trends in productivity, measured by the TNW, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic performance fluctuates with task processing.

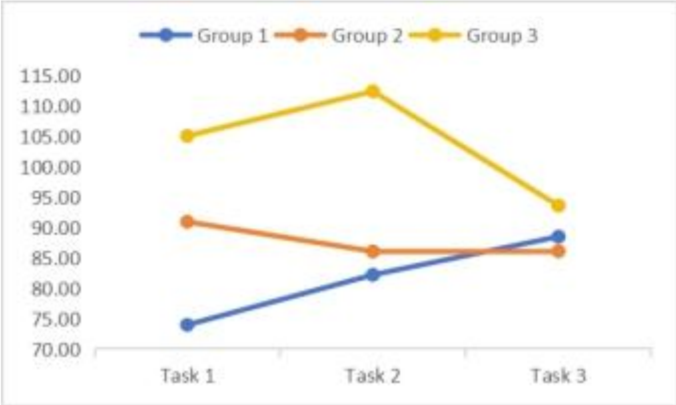

Figure 28 TNW Trends for All Three Groups (along task sequence)

Group 1 (Sequence: S ⟶ M ⟶ C)

The test for Group 1 yields p = .2122, above .05, hence not significant. Although the line graph might show an upward inclination—meaning participants seem to speak more words by Task 3—the difference across tasks does not reach standard significance. Participants might slightly increase their total words as they become more familiar, but the effect is not consistent or large enough to surpass chance variation.

Group 2 (Sequence: C ⟶ M ⟶ S)

In Group 2, p = .6376 is also well above .05, signifying no statistically meaningful difference in total word output from Task 1 to Task 3. Even though the tasks lighten from Complex to Simple, participants do not reliably change their overall verbosity; the line graph indicates only minor fluctuations that remain below significance thresholds. This suggests that Group 2's word count is not strongly tied to the ordering or perceived difficulty of tasks.

Group 3 (Sequence: M ⟶ C ⟶ S)

For Group 3, the p-value of .0578 falls slightly above .05, sometimes regarded as a borderline or marginal scenario, but officially not significant under typical alpha conventions. The line graph reveals a sharper increase from Task 1 to Task 2, followed by a drop at Task 3, indicating possible adaptation or a short-lived boost in word count under the Complex task. However, the data do not confirm it as a definitive effect. The final p-value suggests we cannot conclude a reliable overall sequence impact on Group 3's TNW.

Across all groups, the average number of words exhibits varying patterns, implying that the impact of task sequence on productivity differs notably. Group 3 shows the most pronounced fluctuations, whereas Groups 1 and 2 demonstrate more moderate or minimal changes.

4.2.6 Productivity_Total Number of Syllables

The following table presents the TNS results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences productivity.

TOTAL NUMBER OF SYLLABLES FOR THREE GROUPS_sequence											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	Task 1	Task 2	Task 3		Task 1	Task 2	Task 3		Task 1	Task 2	Task 3
P1	135	157	193	P1	108	109	107	P1	120	150	114
P2	109	127	142	P2	126	144	154	P2	123	154	136
P3	85	127	104	P3	135	121	109	P3	130	128	100
P4	87	89	160	P4	44	54	67	P4	131	145	112
P5	55	82	66	P5	78	72	70	P5	93	140	100
P6	91	97	155	P6	131	140	148	P6	138	116	125
P7	85	82	72	P7	108	98	138	P7	94	143	128
P8	63	124	111	P8	163	72	186	P8	134	111	110
P9	107	121	157	P9	100	85	71	P9	136	120	113
P10	63	145	162	P10	169	142	128	P10	124	141	121
P11	92	116	96	P11	88	138	142	P11	130	119	101
P12	124	138	137	P12	86	72	71	P12	104	137	124
P13	45	42	49	P13	99	105	150	P13	115	150	110
P14	129	131	147	P14	110	109	82	P14	125	149	80
P15	86	120	129	P15	101	129	100	P15	119	132	118
P16	103	105	140	P16	120	114	122	P16	100	130	122
P17	94	116	138	P17	101	105	127	P17	111	150	117
P18	127	85	98	P18	102	98	148	P18	112	120	90
P19	80	131	112	P19	114	92	93	P19	144	112	99
P20	85	119	133	P20	113	97	107	P20	128	137	120
AVE	92.25	112.70	125.05	AVE	109.80	104.80	116.00	AVE	120.55	134.20	112.00

Figure 29 TNS for All Three Groups (along task sequence)

The following line graphs illustrate the trends in productivity, measured by the TNS, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic production fluctuates with task processing.

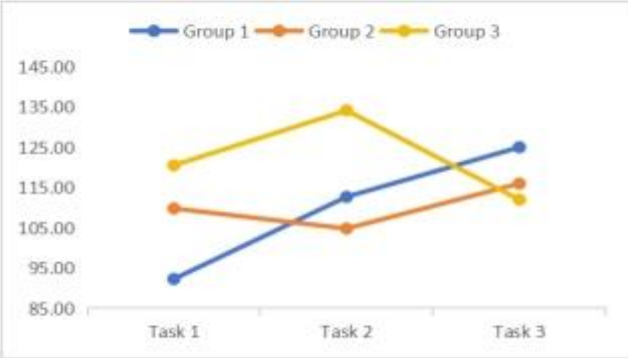

Figure 30 TNS Trends for All Three Groups (along task sequence)

Group 1 (Sequence: S ⟶ M ⟶ C)

For Group 1, $p < .0001$ indicates a very strong and significant effect of the task sequence on TNS. The line graph shows a consistent increase from Task 1 (S) to Task 2 (M) to Task 3 (C), implying participants progressively produce more syllables across the sequence. This might reflect both increased comfort/familiarity and a drive to articulate more content as tasks become more complex. Task 1: Lowest TNS, possibly due to initial caution or simpler demands, Task 2: Moderately higher TNS, Task 3: Highest TNS, suggesting participants achieve their largest phonetic output by the final, most complex stage.

Group 2 (Sequence: C ⟶ M ⟶ S)

In Group 2, $p = .5488$ is above .05, therefore not significant. Although the line graph might reveal some fluctuations—perhaps a small dip followed by a rise—the data do not confirm any reliable pattern in TNS across tasks. Participants may or may not produce fewer/more syllables under simpler tasks, but the net effect is indistinguishable from chance variations.

Group 3 (Sequence: M ⟶ C ⟶ S)

For Group 3, p < .0001 also denotes a highly significant influence of task sequence on TNS. The line graph typically shows a notable jump from Task 1 (M) to Task 2 (C), then a drop at Task 3 (S). Overall, the differences among the three tasks are large enough to exceed random variability. Task 1 ⟶ Task 2: Sharp rise in TNS, suggesting that once they move to a more complex task, participants expand their phonetic production, Task 2 ⟶ Task 3: A pronounced decline, pointing to reduced syllable production in the final, simpler task—perhaps due to less perceived need for elaboration or a sense of fatigue.

Overall, these results show that Group 1: Exhibits a significant overall increase in TNS from Task 1 to Task 3, culminating in the highest oral production at the last stage. Group 2: No significant change across tasks, maintaining a stable or fluctuating TNS that does not meet the .05 threshold. Group 3: Shows a significant effect, with TNS peaking in the middle (Complex) task and falling sharply by the final (Simple) stage.

4.2.7 Fluency_Number of Words Per Minute

The following table presents the NWM results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences fluency.

NUMBER OF WORDS PER MIN FOR THREE GROUPS_sequence											
	Group 1_SMC				Group 2_CMS				Group 3_MCS		
	Task1	Task2	Task3		Task1	Task 2	Task3		Task1	Task2	Task3
P1	91.1	90.8	106.7	P1	90	134.6	153.85	P1	116.13	130	113.15
P2	101.6	116.7	132.0	P2	75.71	110.87	118.32	P2	108.33	113.12	111.9
P3	125.5	105.7	132.0	P3	53.02	79.37	90.9	P3	106.83	110.91	130.9
P4	101.2	118.2	122.0	P4	36.31	62.62	107.78	P4	63.72	67.94	105.3
P5	82.1	103.3	90.0	P5	94.09	99.48	104	P5	103.8	89.5	91.5
P6	91.0	95.0	121.0	P6	63.53	88.45	91.47	P6	60.87	67.06	72.88
P7	115.1	118.3	102.0	P7	95	117.33	110.53	P7	56	54.55	44
P8	51.3	74.1	73.0	P8	77.06	81.74	71.05	P8	81.82	101.83	125
P9	96.5	90.2	115.0	P9	67.2	62.81	92	P9	89.09	86.14	109.52
P10	86.2	81.1	61.0	P10	64	79.73	87.5	P10	106.86	111.55	103.15
P11	125.5	93.5	99.0	P11	109.09	104.54	132.05	P11	104.07	108.21	112.31
P12	93.0	85.1	78.0	P12	120	122	104.06	P12	84.85	76.06	113.41
P13	91.5	133.3	162.0	P13	103.75	101.92	127.21	P13	106.93	103.3	105.6
P14	114.2	72.6	88.0	P14	86.67	105.26	113.33	P14	73.75	76.67	57.95
P15	100.1	97.8	89.1	P15	101.20	66.21	92.50	P15	117.12	117.99	102.98
P16	114.8	73.9	112.0	P16	70.99	111.40	111.30	P16	91.04	90.99	66.93
P17	110.0	88.1	119.7	P17	66.50	83.80	98.14	P17	64.89	85.14	86.08
P18	87.6	107.9	123.1	P18	69.06	76.30	117.65	P18	77.09	90.10	111.14
P19	64.6	96.3	78.2	P19	89.33	112.09	101.30	P19	118.04	80.09	105.88
P20	109.8	126.7	114.5	P20	93.87	128.70	123.73	P20	72.94	91.08	125.05
AVE	97.63	98.43	105.91	AVE	81.319	96.461	107.434	AVE	90.209	92.612	99.732

Figure 31 NWM for All Three Groups (along task sequence)

The following line graphs illustrate the trends in fluency, measured by the NWM, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic output fluctuates with task processing.

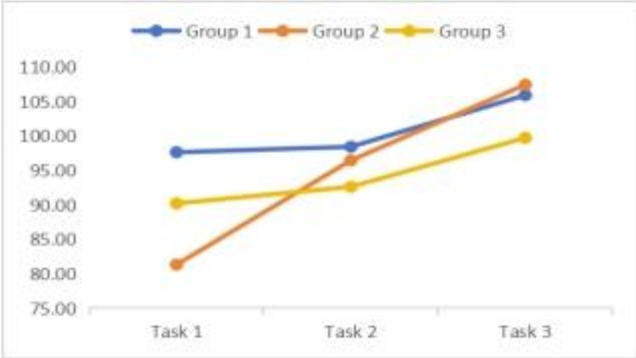

Figure 32 NWM Trends for All Three Groups (along task sequence)

Group 1 (Sequence: S ⟶ M ⟶ C)

For Group 1, the statistical test yields p = .3867, which is well above the .05 significance threshold. While the line graph may show a gradual rise in WPM from Task 1 (Simple) to Task 3 (Complex), these differences do not reach statistical significance. One interpretation is that although participants might speak somewhat faster by the final (most difficult) task, the variation or effect size is too modest to surpass chance-level fluctuations.

Group 2 (Sequence: C ⟶ M ⟶ S)

Turning to Group 2, p = .0006 lies far below .05, indicating a significant impact of task sequence on WPM. Specifically, the line graph suggests a consistent upward trend: participants start with relatively low WPM in Task 1 (Complex), then speak faster in Task 2 (Middle), and reach their highest WPM by Task 3 (Simple). This outcome implies that as tasks become simpler in sequence, Group 2 experiences a clear net acceleration in speech rate.

Group 3 (Sequence: M ⟶ C ⟶ S)

For Group 3, p = .4493 again exceeds .05, thus not significant. The line graph might show a modest upward slope in WPM over the three tasks, but the data do not confirm it as statistically reliable. Hence, Group 3 participants do not exhibit a consistent pattern of speech rate changes aligned to the order in which tasks appear, possibly reflecting heterogeneous individual strategies or minimal sensitivity to sequence in terms of WPM.

Overall, all three groups exhibit a similar upward trend in WPM across tasks, Group 1: No significant shift in WPM across S⟶M⟶C. Group 2: A significant rise in WPM from C⟶M⟶S, producing the clearest evidence that simpler tasks at the end can boost speaking speed. Group 3: Fails to exhibit a significant effect, with only minor or inconsistent changes.

4.2.8 Fluency_Number of Syllables Per Second

The following table presents the NSS results for all three groups across the three stages of the task sequence, offering insights into how task sequence influences fluency.

NUMBER OF SYLLABLES PER SEC FOR THREE GROUPS_sequence

	Group1_SMC				Group 2_CMS				Group 3_MCS		
	Task 1	Task 2	Task 3		Task 1	Task 2	Task 3		Task 1	Task 2	Task 3
P1	2.5	2.12	3.06	P1	1.85	2.89	2.74	P1	2.32	2.8	2.66
P2	2.6	3.43	3.46	P2	1.59	1.66	2.03	P2	3.71	3.07	3
P3	2.65	3.02	3.47	P3	1.61	2.63	2.95	P3	2.13	2.55	2.77
P4	2.17	2.7	2.86	P4	1.02	1.59	2.16	P4	1.6	1.55	2.6
P5	1.66	2.28	1.94	P5	1.77	2.06	2.33	P5	2.45	2.12	1.97
P6	1.62	2.2	2.63	P6	1.28	1.47	1.74	P6	1.43	1.59	1.99
P7	2.3	2.49	2.32	P7	1.8	2.18	2.42	P7	1.28	1.21	0.92
P8	1.02	1.85	1.79	P8	1.6	1.53	1.89	P8	2.03	2.32	2.64
P9	1.81	2.29	2.49	P9	1.23	1.33	1.58	P9	1.94	2.06	2.16
P10	1.62	1.88	1.34	P10	1.41	1.73	1.91	P10	2.4	3.07	2.59
P11	2.56	1.94	2.59	P11	2	1.97	2.94	P11	2.23	2.84	2.85
P12	1.08	2.04	1.65	P12	2.61	2.25	2.09	P12	1.89	1.91	2.33
P13	1.81	2.8	2.88	P13	2.06	1.98	2.59	P13	2.74	3.01	2.29
P14	1.79	1.44	1.93	P14	1.75	2.22	2.54	P14	1.37	2.11	1.75
P15	1.73	2.58	1.68	P15	1.66	2.49	2	P15	2.02	2.98	3.25
P16	3	1.73	3.21	P16	2.07	2.67	2.87	P16	3.46	2.13	2.39
P17	3.02	1.54	2.18	P17	3.09	2.26	2.84	P17	2.45	1.62	2.03
P18	3.45	1.93	1.53	P18	3.21	3.03	2.72	P18	1.51	2.68	3.49
P19	1.52	2.87	3.48	P19	1.64	3.25	3.26	P19	3.48	2.11	2.37
P20	2.64	1.86	2.37	P20	2.59	1.57	2.37	P20	2.88	2.55	1.57
AVE	2.128	2.250	2.443	AVE	1.892	2.138	2.399	AVE	2.266	2.314	2.381

Figure 33 NSS for All Three Groups (along task sequence)

The following line graphs illustrate the trends in fluency, measured by the NSS, for the three groups across Task 1, Task 2, and Task 3. These graphs clearly demonstrate how phonetic production fluctuates with task processing.

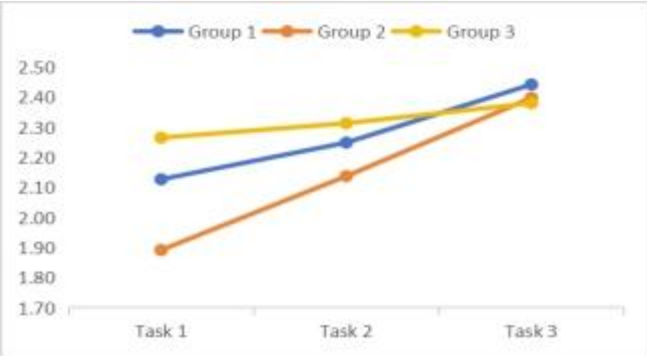

Figure 34 NSS Trends for All Three Groups (along task sequence)

Group 1 (Sequence: S ⟶ M ⟶ C)

Group 1's p = .041 is below .05, indicating a significant effect of task sequence on SPS. The line graph shows a moderate climb in SPS from Task 1 (Simple) through Task 2 (Middle) to Task 3 (Complex). On average, participants appear to speak more syllables per second as they progress through tasks in ascending complexity, possibly reflecting growing familiarity and an increased drive to maintain pace under heavier cognitive loads.

Group 2 (Sequence: C ⟶ M ⟶ S)

In Group 2, p = .003 also lies below .05, signifying a significant difference across tasks. The line graph suggests a consistent increase in SPS from Task 1 (Complex) to Task 3 (Simple). Much like WPM, participants gradually speed up their syllable production once tasks shift from more difficult to easier. By the final stage, they reach the highest SPS, underscoring that Group 2 responds positively—i.e., speaks more rapidly—when they finish with simpler demands.

Group 3 (Sequence: M ⟶ C ⟶ S)

For Group 3, p = .449 is above .05, hence not significant. The line graph might show mild upward in SPS, but nothing robust enough to be deemed reliable. Thus, the ordering M⟶C⟶S does not yield a consistent step-by-step change in how quickly participants articulate syllables, echoing previous findings on WPM and other measures for this group.

Overall, all three groups exhibit a similar upward trend in SPS across tasks, Group 1: Shows a significant ascending SPS trend across S⟶M⟶C. Group 2: Also demonstrates a significant upward shift from C⟶M⟶S, culminating in the highest SPS at the simplest final task. Group 3: No significant effect, indicating stable or inconsistent SPS patterns along M⟶C⟶S.

4.2.9 Summary and Implications for *Task-order Effect*

The following table presents the fluctuation for all language performance dimensions across three groups (Task-Order Effect), (Simple-Middle, Middle-Complex, Simple-Complex).

	TTR	MLAS	MLC	TNW	TNS	WPM	SPS	EFCR
Group 1	(-,-,-)	(+,-,+)	(+,-,+)	(+,+,+)	(+,+,+)	(+,+,+)	(+,+,+)	(+,+,+)
Group 2	(-,-,-)	(-,+,+)	(+,-,+)	(-,=,-)	(-,+,+)	(+,+,+)	(+,+,+)	(-,+,+)
Group 3	(-,-,-)	(-,-,-)	(+,-,-)	(+,-,-)	(+,-,-)	(+,+,+)	(+,+,+)	(+,-,-)

Task Sequence & L2 Performance (Task-Order Effect)

For Group 1, the results indicate that task sequence has a notable influence on various dimensions of language performance. Lexical complexity shows a clear downward trend, suggesting reduced diversity in later tasks, likely due to cognitive fatigue. Syntactic complexity exhibits significant changes, with increased AS-unit and clause lengths observed in the middle tasks before a slight decline in the final task. Productivity steadily rises across tasks, with participants producing more words and syllables, while fluency, reflected in both words per minute and syllables per second, improves gradually. Accuracy also demonstrates a consistent upward trend, indicating enhanced precision in clause production as tasks progress.

In Group 2, task sequence similarly affects performance, though the patterns are less pronounced. Lexical complexity shows a subtle downward trend, and syntactic complexity reveals moderate fluctuations, with some improvements in later tasks. Productivity and fluency increase steadily, particularly in fluency measures where participants show a strong positive response to task sequencing. However, accuracy remains relatively stable, with only minor changes across tasks, suggesting that task sequence has a limited impact on this dimension.

For Group 3, the influence of task sequence is more variable. While lexical complexity fluctuates with no clear trend, syntactic complexity demonstrates a gradual decline, particularly in the later stages, reflecting potential cognitive strain. Productivity shows noticeable fluctuations, with increases in earlier tasks followed by declines in the final task. Fluency improves slightly across tasks, although the statistical confidence is weaker compared to the other groups. Accuracy remains largely unchanged, indicating minimal sensitivity to task order.

In conclusion, the results across all three groups highlight the nuanced impact of task sequence on language performance. While Groups 1 and 2 display relatively consistent trends of improvement in fluency and productivity, Group 3 exhibits greater variability. Accuracy appears to benefit most in Group 1, with more limited effects in the other groups. These findings underscore the complex relationship between task sequence and linguistic output, providing valuable insights into how task design can shape language performance.

Task-level Effect vs. Task-order Effect

Building on the findings from both task-level (complexity) and task-order (sequence) analyses, we can see how each dimension (CALF) is shaped by *what* learners do (the difficulty of the tasks) as well as *when* they do it (the order in which those tasks are presented). Below is a concise synthesis showing how these two factors converge and sometimes interact across the three groups.

1. Group 1

Task-Level Effect (Complexity):

This group broadly supports the Cognition Hypothesis—as tasks move from simpler to more complex levels, participants display notable gains in syntactic complexity (MLAS, MLC), accuracy (EFCR), and even fluency (WPM, SPS). However, lexical complexity (TTR) eventually dips at the highest complexity, suggesting that although they can manage higher demands in structure and accuracy, their attentional resources may not stretch far enough to sustain full lexical diversity.

Task-Order Effect (Sequence):

Across tasks, Group 1's performance in productivity (TNW, TNS) and fluency shows a steady rise, while accuracy steadily improves. This pattern hints that encountering tasks in a certain sequence also helps them build momentum—yet the eventual drop in TTR aligns with the idea of cognitive fatigue late in the sequence.

Possible Correlation:

As complexity increases, Group 1 copes well overall (Cognition Hypothesis), but the final phase suggests both attentional limits (due to higher complexity) and fatigue (due to task order). For example, a demanding final task may compound the strain on lexical variety.

2. Group 2

Task-Level Effect (Complexity):

Group 2's results validate Limited Attentional Capacity (LAC): once task complexity rises, fluency (WPM, SPS) and syntactic complexity (MLC) suffer, and TTR proves difficult to maintain. They prioritize basic performance rather than enhancing complexity or variety.

Task-Order Effect (Sequence):

While fluency and productivity do show improvements over successive tasks, accuracy remains fairly stable, indicating that changing the order alone does not strongly lift accuracy for this group. Their slight gains in later tasks suggest a moderate benefit from sequencing, but not enough to override the heavier toll imposed by increased complexity.

Possible Correlation:

Even if Group 2 adapts somewhat over time (task order), high-complexity tasks deplete attentional resources the most. Thus, task sequence can offer incremental improvements in fluency or oral performance, but complexity dominates their overall limits—especially for higher-level syntactic or lexical features.

3. Group 3

Task-Level Effect (Complexity):

This group shows a hybrid pattern. They attempt to align with the Cognition Hypothesis—gradually improving accuracy and occasionally boosting lexical or syntactic metrics—but only to a moderate degree. Fluency is inconsistent, reflecting partial success in reallocating attentional resources under higher cognitive demands.

Task-Order Effect (Sequence):

The sequence data reveal variability: productivity increases early on, then declines; fluency improves slightly but with weaker statistical confidence; accuracy remains largely unaffected. These ups and downs imply they can adapt from task to task, but not in a strictly linear fashion.

Possible Correlation:

Because Group 3 is somewhat successful under higher complexity, task order can help them manage cognitive load—but it does not guarantee consistent gains. They show glimpses of higher lexical/syntactic performance yet remain sensitive to fluctuations. A tough final task or an overly challenging middle task can disrupt their otherwise steady adaptation.

Key Takeaways on Correlation

1) Lexical Complexity (TTR)

Often more vulnerable to task-level increases (especially at the highest complexity), as seen in Group 1 and Group 2's declines.

Can also be affected by sequence—for instance, Group 1 experiences fatigue in later tasks, reducing TTR even if complexity was manageable earlier.

2) Syntactic Complexity (MLAS, MLC)

Strongly tied to cognitive load (Task-Level). Groups 1 and 3 try to increase syntactic complexity under higher complexity tasks (Cognition Hypothesis), whereas Group 2 shows clear drop-offs (LAC).

Order can modulate whether learners get “warmed up” (leading to better complexity mid-sequence) or become fatigued.

3) Fluency & Productivity (WPM, SPS, TNW, TNS)

More immediately responsive to task sequence: repeated practice or familiarity can boost oral output and speed from one task to the next (Groups 1 and 2).

However, at higher complexity, limited attentional capacity might still degrade fluency if demands become overwhelming (Group 2 and partially Group 3).

4) Accuracy (EFCR)

Group 1 sees clear improvements with both higher complexity (they can allocate resources effectively) and the sequence factor (gradual mastery over tasks).

In Group 2 and Group 3, accuracy changes are minimal, implying either complexity or order alone does not significantly push them to refine clause-level precision.

Comparative Trends Across CAF Dimensions Under Different Task Conditions

Influence of Cognitive Task Complexity

The empirical findings show that varying cognitive task complexity produces distinct, and sometimes opposite, responses across the CAF dimensions. In general, moderate task complexity tends to elicit improvements in both syntactic complexity and fluency, whereas extremely high complexity can hinder certain aspects of performance. For instance, when tasks were moderately challenging, participants often produced longer and more complex sentences (higher syntactic complexity) while maintaining or even slightly increasing their speech rate. This alignment suggests that a moderate cognitive load can boost structural elaboration and spoken fluency in tandem, as learners are stimulated to express more ideas without being overwhelmed. Lexical variety, however, showed an opposite pattern under these same conditions. As task demands grew, measures like Type-Token Ratio (TTR), an index of vocabulary diversity, which tended to decline, especially from moderate to highly complex tasks. In other words, faced with greater cognitive load, participants resorted to a more limited range of words, indicating that cognitive strain constrained lexical diversity even when fluency and syntax were still advancing. This divergence highlights a trade-off: increasing complexity spurred faster and structurally richer speech, but at the cost of lexical richness, likely because speakers focused their mental resources on formulating complex sentences and keeping pace, leaving fewer resources for varied word choice.

Accuracy vs. Complexity Dynamics

Interestingly, the relationship between accuracy and syntactic complexity did not reflect a strict trade-off in the study's results. On the contrary, accuracy (measured by error-free clauses) sometimes improved alongside syntactic complexity as tasks became more challenging. For two of the three participant groups, error-free clause ratios rose modestly from simple to complex tasks, suggesting that these learners managed to uphold or even enhance grammatical accuracy while also producing more complex syntax. This indicates that under moderate task complexity, accuracy and complexity can align, with learners paying attention to form even as they expand structure: a finding consistent with the absence of an accuracy & complexity conflict in those groups. However, at the highest level of task complexity, a subtle divergence emerged: syntactic complexity gains plateaued or reversed (sentences became slightly simpler or shorter under extreme load), even though accuracy did not significantly drop. This pattern implies that when cognitive demands peak, learners prioritize maintaining accuracy over pushing greater complexity, resulting in stable accuracy but no further syntactic elaboration. Notably, none of the groups showed a sharp accuracy decline as complexity increased: a testament that high task complexity did not universally trigger an accuracy breakdown. Instead, participants adjusted other dimensions (like reducing lexical variety or simplifying syntax) to cope with difficulty while keeping their speech relatively error-free. These trends underscore that CAF dimensions link under cognitive load: a challenging task might simultaneously strengthen some areas (accuracy, syntax) while weakening others (lexical diversity), depending on how learners allocate their limited attentional resources.

Effects of Task Sequence on Performance

The study also reveals clear patterns when comparing performance across the sequence of three tasks (regardless of complexity level). Some CAF dimensions improved consistently over the sequence, hinting at practice effects, while others showed decline or non-linear trends due to fatigue or repetitive content. Fluency demonstrated a robust upward trend across successive tasks for all groups. Measured by words per minute (WPM) or syllables per second, fluency increased from the first to the final task, indicating that as participants became more familiar with the speaking task format and content, they spoke more rapidly and fluidly. This sequential improvement was most pronounced in one group (which showed a strong gain in WPM over the three tasks) and evident albeit to a lesser degree in the other groups. The universal direction of change suggests a practice or habituation effect: repeated task exposure allowed speakers to gain confidence and efficiency, thus boosting fluency independent of task complexity. In stark contrast, lexical variety consistently declined with each successive task. All three groups showed lower TTR in later tasks than in the first, meaning their speech became lexically simpler over time. This drop in lexical diversity can be attributed to cumulative fatigue or reduced novelty, as participants repeat task scenarios or topics, they might rely on familiar words and phrases, leading to less varied vocabulary in later performances. The opposing trajectories of fluency and lexical complexity over task repetition are striking: speech became faster but lexically less diverse as the sequence progressed. This suggests that while practice makes speaking more fluid, it may also encourage a kind of routinization where speakers settle into using a limited repertoire of words.

Syntactic Complexity and Accuracy over Task Sequence

Unlike the linear patterns observed for fluency and lexis, syntactic complexity followed a curvilinear trend across the task sequence. Across groups, participants often peaked in syntactic complexity in the second task before dropping off in the third. In other words, the middle task in the sequence elicited the most syntactically complex speech, after which participants produced slightly simpler sentences in the final task. This pattern likely reflects a short-term boost as learners warmed up and hit their stride by the second performance, followed by a mild decline in the last task due to fatigue or task repetition effects. By the third round, participants might have been mentally taxed or less inclined to elaborate syntax, leading to a simplification of utterances. Accuracy trends under task repetition were more subtle and varied by group. One group showed a steady improvement in accuracy (error-free clause ratio increased with each task, suggesting that continued practice helped them speak more correctly), whereas the other two groups' accuracy remained largely flat or oscillated slightly across the three tasks. For those latter groups, performing the task repeatedly did not substantially change how accurate their language was, perhaps because they were already operating near a personal stable level of accuracy or because any potential gains were offset by fatigue. The group that did improve in accuracy over time illustrates that task rehearsal can bolster accuracy, potentially as speakers become more comfortable and pay greater attention to form in later attempts.

Meanwhile, the absence of a strong sequential effect in the other groups indicates that accuracy is the most resistant dimension to change from mere repetition, especially if no explicit feedback is given between tasks.

Group Differences and connections

Across the three participant groups, there were both notable consistencies and divergences in these patterns. Some trends were robust across all groups, for example, the decline in lexical variety over task repetitions and the general fluency gains with practice were observed in each group, underscoring common effects of task sequence on these dimensions. This consistency suggests that certain performance aspects (like using diverse vocabulary) are universally susceptible to fatigue or resource depletion, while others (like fluency) universally benefit from practice and familiarity. On the other hand, responses to increasing task complexity varied by group, reflecting how different task sequences or individual proficiencies modulated the CAF interplay. One group (those who tackled tasks from simple to complex in order) appeared to adapt optimally to rising complexity: they managed to speak faster and more accurately with more complex tasks, and showed the largest syntactic gains at moderate complexity. Another group (who began with the hardest task first) showed the greatest difficulty under high complexity, evidenced by a pronounced drop in fluency at the most complex task and an overall leveling-off of complexity and accuracy measures: they maintained their accuracy but only by slowing down and simplifying their speech under pressure. The third group displayed mixed adaptations, with moderate improvements and smaller fluctuations, indicating a more stable but less pronounced response to both complexity and repetition. Despite these differences, the overarching finding is that no single CAF dimension dominates or uniformly directs performance; instead, the dimensions connect in dynamic ways. For example, all groups exhibited some form of trade-off between lexical variety and other dimensions: when participants pushed themselves to produce language more fluently or with more complex structures (whether due to a challenging task or simply by the second task in the sequence), they often did so by drawing on a narrower lexical repertoire. Conversely, gains in accuracy did not necessarily impede complexity or fluency in many cases, which is a positive sign that learners can improve form and flow together under supportive conditions (like moderate challenge or repeated practice).

Overall relationship of CAF Dimensions

These empirical patterns paint a cohesive picture of how CAF dimensions align and diverge in a task-based performance context. Fluency and syntactic complexity often improve in parallel under optimal conditions (moderate cognitive complexity or through practice), suggesting a synergy where being comfortable with a task enables speakers to talk both faster and in more complex sentences. In contrast, lexical diversity tends to move inversely relative to those gains : when tasks become demanding or as speakers grow accustomed to a task, they appear to sacrifice variety in word choice, likely as a strategy to reduce cognitive load. Accuracy, for its part, demonstrates a more independent trajectory, generally holding steady or gradually improving regardless of fluctuations in the other dimensions. The fact that accuracy did not dramatically deteriorate even when other aspects were pushed (and even improved with complexity for some groups) indicates that learners can maintain a baseline of correctness while adjusting other performance areas. In sum, the CAF dimensions interact in response to task conditions in a compensatory manner: certain dimensions can be enhanced together (e.g. fluency with syntactic or accuracy gains), whereas others may be constrained as those improvements take place (e.g. lexical variety). The presence of three distinct groups in the study further highlights that individual or contextual factors (such as task order or learner profiles) moderate these relationships. Across all groups, repeated task exposure tended to facilitate fluency (through practice) but constrain lexical novelty, whereas increasing cognitive task complexity up to a moderate level fostered more complex and accurate language without universally hurting fluency. Beyond that moderate point, however, performance trade-offs became more evident as participants coped with maximal task demands.

Overall, these findings provide a nuanced overview of how complexity, accuracy, and fluency dimensions converge or diverge under different task conditions, reinforcing the idea that task-based language performance is the result of a dynamic balancing act among multiple competing linguistic demands.

4.3 Questionnaire Analyze (Emotional Aspect)

This section presents a comprehensive analysis of the survey data derived from the Foreign Language Fear (FLF) and Foreign Language Enjoyment (FLE) questionnaires. Participants completed these questionnaires immediately after each task, enabling an evaluation of their emotional responses specific to the task just completed. The study involved 60 participants, each of whom completed three tasks, resulting in a total of 180 questionnaire responses.

The primary objective of this analysis is to investigate how factors such as cognitive task complexity and task sequencing influence participants' levels of Foreign Language Fear (FLF) and Foreign Language Enjoyment (FLE). By analyzing the data, we aim to uncover the relationships between these emotional responses and the perceived difficulty or order of tasks. Additionally, the survey responses provide a basis for evaluating fluctuations in participants' emotions throughout the experiment, offering valuable insights into the dynamic nature of their foreign language learning experiences.

This analysis not only illuminates the connection between task demands and emotional experiences but also underscores the potential impact of emotions such as fear and enjoyment on task performance. By examining the evolution of these emotions during the process, we can identify patterns of emotional adaptation or resilience, which may inform the development of more effective teaching strategies and task designs in foreign language learning contexts. The insights derived from the 180 completed questionnaires will contribute to a deeper understanding of the emotional dimensions of foreign language learning, particularly in relation to the challenges introduced by task complexity and sequencing.

4.3.1 Foreign Language Fear (FLF)

The FLF questionnaire in this study assesses the intensity and progression of FLF emotions experienced before, during, and after task performance. Drawing on the collected data, this section provides evidence supporting the presence of fear-related emotions as conceptualized within the FLF framework, distinguishing them from the broader construct of foreign language anxiety. Furthermore, the analysis investigates two key factors influencing FLF emotions: *the task-level effect*, which captures emotional fluctuations associated with cognitive task complexity, and *the task-order effect*, which examines changes in emotions resulting from the sequence of task performance. Both aspects are explored within the scope of FLF emotions to provide a nuanced understanding of their dynamics.

4.3.1.1 FLF for Three Groups

The table below presents the average scores for each question based on the responses of 20 participants in each group to the five-point Likert scale questions in the FLF questionnaire. Group 1 completed the tasks in the order of simple-middle-complex, Group 2 followed the order of complex-middle-simple, and Group 3 completed the tasks in the order of middle-complex-simple. The data highlights the variations in FLF emotion scores across the three groups, reflecting their emotional responses under different task sequences.

		Question 1	Question 2	Question 3	Question 4	Question 5	Average value
Group 1 NP=20	Simple Level	2.244	2.567	2.048	2.567	3.404	2.566
	Middle Level	2.739	2.913	2.683	3.043	3.739	3.023
	Complex Level	2.889	3.142	2.833	3.267	4.128	3.252
Group 2 NP=20	Simple Level	3.25	2.817	2.75	2.988	3.458	3.053
	Middle Level	3.111	2.875	2.708	3.042	3.375	3.022
	Complex Level	3.231	2.692	2.308	2.692	3.154	2.815
Group 3 NP=20	Simple Level	3.178	3.261	2.668	3.261	3.422	3.158
	Middle Level	2.333	3.100	2.426	2.684	3.421	2.793
	Complex Level	3.360	3.540	3.000	3.200	3.960	3.412

Figure 35 FLF Scores for All Three Groups

The following two line charts illustrate the changes in FLF emotions across the three groups. The first chart depicts the variation in FLF emotions with changes in cognitive complexity, while the second chart presents the fluctuations in FLF emotions based on task sequence. Both charts are plotted using the average scores from the questionnaire responses, providing a clear visual representation of the trends in FLF emotions under different conditions.

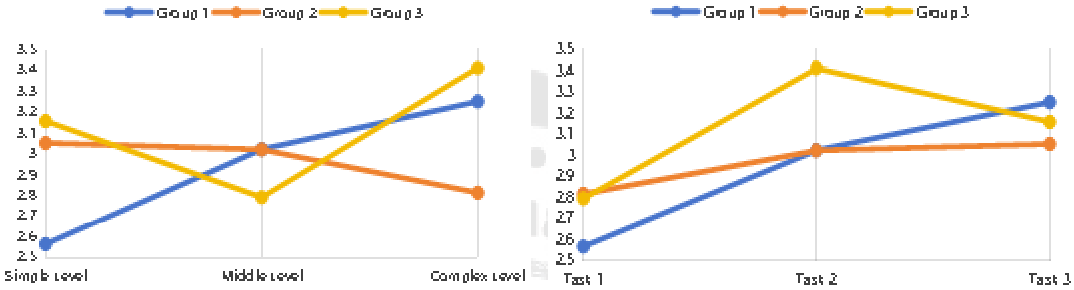

Figure 36 FLF Scores for All Three Groups (along CTC levels and sequence)

Group 1

Task-Level (CTC) Analysis

A statistical examination of cognitive task complexity (CTC) and FLF emotion ratings in Group 1 (Simple, Middle, Complex) revealed a progressive increase in FLF emotions as tasks became more demanding. Specifically, average ratings rose from 2.566 in Simple to 3.023 in Middle (p = 0.0011, significant), and then from 3.023 in Middle to 3.252 in Complex (p = 0.0064, also significant). The most notable overall shift appeared between Simple (2.566) and Complex (3.252) with p = 0.000043, underscoring a robust relationship between higher cognitive load and heightened fear-related responses. Thus, more demanding tasks significantly amplified FLF emotions in Group 1.

Task-Order (SMC) Analysis.

Turning to sequence effects, Group 1 tackled Task 1 (Simple), Task 2 (Middle), and Task 3 (Complex) in that order, yielding FLF scores of 2.566, 3.023, and 3.252, respectively. This forms a clear upward trend: FLF was lowest at the first task, rose by the second, and peaked by the third—matching the incremental jump in complexity. However, because task order (1⟶2⟶3) perfectly aligns with rising complexity (S⟶M⟶C), the steady growth in FLF may stem from both sequential buildup and increasing demands. Future comparisons with Groups 2 and 3—where sequences differ—may clarify whether complexity or sequence is the primary driver, or if both factors interact to elevate fear-related emotions.

Group 2

Task-Level (CTC) Analysis.

For Group 2, the relationship between CTC and FLF ratings followed a downward trajectory: FLF declined from 3.053 in Simple to 3.022 in Middle (p = 0.4738, not significant), then dropped further from 3.022 to 2.815 in Complex (p = 0.0853, also not significant though marginally close). Although these p-values do not meet conventional thresholds, they suggest a mild reduction in FLF at higher task complexities. Directly comparing Simple (3.053) and Complex (2.815) yields p = 0.0330, which does reach significance, indicating Group 2 shows lower FLF emotions at the most demanding level relative to the simplest. The largest shift emerges between those two extremes.

Task-Order (CMS) Analysis.

Regarding sequence, Group 2 approached tasks in the order Task 1 (Complex), Task 2 (Middle), Task 3 (Simple), producing FLF averages of 2.815, 3.022, and 3.053, respectively. This reveals a gradual increase over time: FLF started lowest at the Complex initial task, rose moderately by the Middle second, and climbed slightly further at the Simple final. This fluctuation may reflect factors such as mental fatigue or shifting emotional states across tasks, demonstrating that Group 2's fear built up during the sequence—even though isolated comparisons of complexity did not always intensify FLF. In other words, while cognitively complex tasks in isolation correlate with somewhat lower FLF, the actual order (C⟶M⟶S) fosters a mild upward drift in fear.

Group 3

Task-Level (CTC) Analysis.

Group 3 displayed a mixed pattern in terms of CTC. FLF ratings dropped from 3.158 in Simple to 2.793 in Middle ($p = 0.0747$, not meeting .05 but near a marginal zone), then rose substantially to 3.412 in Complex ($p = 0.0040$, significant). Comparing Simple (3.158) and Complex (3.412) yields $p = 0.0604$, which does not pass the .05 cutoff, though it suggests a possible upward shift. Overall, Group 3's most pronounced spike in FLF occurs between Middle and Complex tasks, with a smaller yet noteworthy change when contrasting Simple and Complex. The net effect is that complexity transitions do influence Group 3's fear, but not every step is individually significant.

Task-Order (MCS) Analysis.

For sequence, Group 3 proceeded as Task 1 (Middle), Task 2 (Complex), and Task 3 (Simple), yielding average FLF scores of 2.7928, 3.4120, and 3.1580, respectively. Thus, FLF is lowest at the start (Middle), peaks during the Complex second task, and then recedes slightly by the final (Simple). This "rise-then-dip" shape underscores fluctuating emotional responses influenced by an early jump to complex demands and a partial letdown or adaptation by the last stage. Task order, combined with the jump in complexity, appears to shape how fear escalates or abates across the session.

Summary

Taken together, these findings highlight distinct interactions between task complexity and task sequence in shaping FLF emotions across the three groups:

Group 1 (SMC) shows a straightforward upward trend in FLF aligned with both increasing complexity and sequential progression, making it challenging to disentangle whether the boost in fear is primarily due to rising difficulty or cumulative task order.

Group 2 (CMS) interestingly experiences less fear at higher complexity when viewed in isolation, but their FLF climbs gradually as they move from a tough initial task to a simpler final one—suggesting that sequence (starting at Complex) and potential fatigue or acclimatization shape their emotional arc.

Group 3 (MCS) exhibits a more fluctuating response, with a significant spike from Middle to Complex but a partial drop-off when ending on Simple, indicating a combined effect of sudden increased demands plus final relief or adaptation.

These patterns underscore the dual influence of what learners face (cognitive load) and when they face it (task order). In some cases, greater complexity directly heightens FLF; in others, an initially tough task may reduce fear if subsequent tasks are easier or if participants adapt over time. A broader inter-group comparison and further analysis of potential confounding factors (e.g., individual differences, motivation) can yield deeper insights into how best to balance task demands and sequencing to manage fear-related emotions in foreign language learning.

4.3.1.2 Emotional Dynamics of FLF Throughout the Task Process

This section addresses three key questions from the FLF questionnaire, focusing on the dynamics of FLF emotions throughout the entire process of completing the three tasks. The analysis encompasses multiple stages: before, during, and after each task (Task 1, Task 2, and Task 3). The table below presents the average scores for these dimensions of FLF emotions across all groups, based on responses from a total of 60 participants.

	Task 1 (Simple)			Task 2 (Middle)			Task 3 (Complex)		
Group 1	2.444	2.667	2.148	2.739	2.913	2.783	3.000	3.042	2.833
	Task 1 (Complex)			Task 2 (Middle)			Task 3 (Simple)		
Group 2	3.231	2.692	2.308	3.292	2.875	2.708	3.250	2.917	2.750
	Task 1 (Middle)			Task 2 (Complex)			Task 3 (Simple)		
Group 3	2.526	2.368	2.526	3.360	3.440	3.000	3.478	3.261	3.000

Figure 37 The Dynamic Changes for FLF Scores

To provide a clear view of the dynamic changes within each group, the line graph below visually represents the fluctuations in FLF emotions over time.

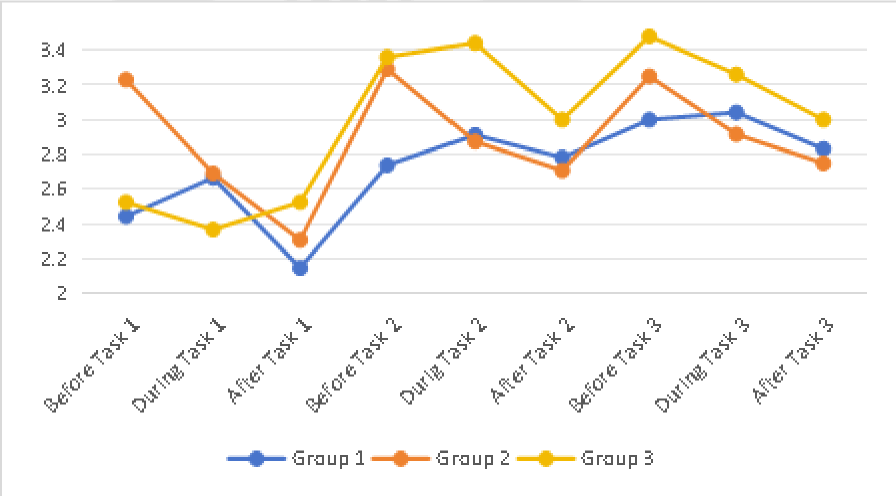

Figure 38 The Dynamic Changes for FLF Scores (line graph)

For this analysis, the term 'before the task' refers to the stage after participants finished watching the task's instructional videos and were preparing to answer the questions posed by the computer. 'During the task' refers to the stage when participants were actively speaking and engaging in conversation with the computer. 'After the task' denotes the period following the completion of all task-related questions, as participants prepared to proceed to the next task. According to the line graph, nine data points representing FLF emotion levels were recorded, revealing noticeable fluctuations both within individual groups and across the three groups. This pattern provides evidence of the presence of FLF emotions, indicating that FLA (Foreign

Language Anxiety) and FLF may co-occur throughout the experimental process. As discussed in the previous section, based on the APA Dictionary's definitions of anxiety and fear, FLA is expected to remain relatively stable over time. In contrast, the variability observed in FLF levels suggests a more dynamic and context-dependent emotional response.

The chart reveals a dynamic pattern of discomfort and emotional shifts across the different stages of task engagement, suggesting that participants' responses cannot be confined to a single emotional state. Rather, the data points to an relation between anxiety and fear. For example, the gradual changes in emotional responses before tasks might align with anxiety, reflecting participants' anticipation of the upcoming challenges. In contrast, more abrupt fluctuations during or immediately after tasks could signal the presence of fear, as participants confront specific, identifiable difficulties or immediate task demands.

This distinction reinforces the idea that anxiety and fear do not operate in isolation but are likely intertwined throughout the process. The emotional variability observed in the chart suggests that participants' discomfort arises from a combination of anticipatory anxiety and situational fear, each shaping their experience at different points in time. This perspective underscores the complexity of emotional experiences during cognitively demanding tasks, highlighting how the coexistence and interaction of anxiety and fear contribute to the participants' dynamic responses.

In the following section, the discussion of the line graph will focus on two key aspects:

The task-level effect, examining how FLF emotions vary with cognitive task complexity.

The task-order effect, exploring how the sequence of tasks influences fluctuations in FLF emotions.

‘

Group 1

Before Task 1: The average FLF emotion score starts at 2.444, indicating relatively low discomfort or apprehension before the first task.

During Task 1: As the task progresses, FLF emotions rise to 2.667, showing that participants experience a slight increase in discomfort while performing the task.

After Task 1: Following Task 1, FLF emotions drop to 2.148, suggesting that the completion of the task brings some relief or reduction in discomfort.

Before Task 2: The FLF emotion score rises again to 2.739 before the second task, indicating that participants feel a higher level of discomfort or apprehension compared to Task 1.

During Task 2: The FLF emotion continues to rise slightly during the task, reaching 2.913, implying that participants experience increased discomfort while engaged in the task.

After Task 2: The score stabilizes at 2.783 after the task, which is lower than the peak during Task 2, suggesting slight relief, but higher compared to the post-Task 1 period.

Before Task 3: The FLF emotion level peaks at 3.000 before Task 3, marking the highest pre-task level recorded, indicating that participants are feeling the most apprehensive at this point in the process.

During Task 3: FLF emotions remain high during the task, with the score rising to 3.042, the highest recorded level during the entire sequence of tasks, suggesting that discomfort is most pronounced while completing the final task.

After Task 3: Following the completion of Task 3, the FLF emotion level decreases to 2.833, which shows that some relief is felt, though the overall emotional burden remains elevated compared to the earlier stages of the process.

Key Observations for Group 1

Gradual Increase in FLF Emotion: The upward trend in FLF emotions throughout the task sequence is influenced by both the task order and the level of cognitive task complexity. As Task 1 is the simplest task, participants start with relatively low FLF emotions. However, as they move to Task 2 (middle complexity) and then to Task 3 (complex), both the increasing cognitive demands and the progression through the tasks contribute to the rising emotional discomfort. The combination of task complexity and the anticipation of more challenging tasks likely elevates FLF emotions.

Relief After Each Task: Although participants experience a temporary reduction in FLF emotions after completing each task, this relief is short-lived. The task order plays a significant role here, as the completion of an easier task like Task 1 offers more emotional relief than the later, more complex tasks. However, as participants approach Task 2 and Task 3, the growing cognitive complexity, along with the anticipation of more difficult tasks, leads to renewed FLF emotions. This suggests that the emotional relief might be overridden by the challenge posed by the next task.

Peak During Task 3: The highest FLF emotion levels are recorded before and during Task 3, which is the most complex task. The emotional strain is likely due to a combination of two factors: the cumulative effect of having completed the previous tasks, and the high cognitive demand of Task 3 itself. As the final task in the sequence, participants may feel a heightened sense of pressure or fear, knowing that the task is the most challenging, which contributes to the highest FLF levels observed.

Cumulative Build-up: The FLF emotion levels increase steadily across the tasks, indicating a cumulative emotional and cognitive load. Participants begin with Task 1 (simple), which acts as a relatively low-pressure entry point. However, as they progress through Task 2 and Task 3, the combined effects of cognitive fatigue from earlier tasks and the increasing complexity of the subsequent tasks lead to a build-up of emotional discomfort. This suggests that both the order in which tasks are completed and their growing complexity play critical roles in amplifying FLF emotions.

Summary for Group 1

Group 1's emotional experience is shaped by the interaction between task order and cognitive task complexity. As participants progress from Task 1 (simple) to Task 2 (middle complexity) and finally to Task 3 (complex), their FLF emotions rise due to both the increasing difficulty and the cumulative strain of successive tasks. Although there is some relief after each task, the anticipation of more challenging tasks in the sequence causes FLF emotions to rebound. The peak in FLF emotions during Task 3 underscores the combined impact of task complexity and accumulated emotional fatigue, highlighting the intertwined effects of task order and cognitive demands on participants' emotional states.

Group 2

Before Task 1: The average FLF emotion score starts at 3.231, reflecting a relatively high level of discomfort or apprehension before the first task, which is the most complex task in the sequence.

During Task 1: As participants engage with Task 1, the FLF emotion score decreases to 2.692, indicating that participants' discomfort reduces as they focus on the task.

After Task 1: The FLF emotion score drops to 2.308, marking the lowest point in the sequence. This significant reduction suggests a sense of relief and emotional release after successfully completing the most challenging task.

Before Task 2: The FLF emotion rises again to 3.292 before starting Task 2, indicating renewed apprehension or discomfort, slightly higher than the pre-task level for Task 1. This rise suggests that participants expect a new challenge, even though Task 2 is less complex.

During Task 2: During Task 2, the FLF emotion decreases to 2.875, implying that participants once again experience less discomfort once they engage in the task, similar to the pattern observed during Task 1.

After Task 2: The FLF emotion further decreases to 2.708 after Task 2, reflecting continued emotional relief after completing the task, though this relief is not as pronounced as the post-Task 1 period.

Before Task 3: The FLF emotion level remains relatively stable before Task 3 at 3.250, showing a moderate level of discomfort or apprehension as participants approach the final task, which is the simplest in this sequence.

During Task 3: During Task 3, the FLF emotion decreases slightly to 2.917, showing that participants feel less discomfort while performing the task.

After Task 3: After Task 3, the FLF emotion level drops to 2.750, reflecting that participants experience less discomfort or apprehension after completing the final task, though the relief is not as significant as that experienced after Task 1.

Key Observations for Group 2:

High Initial FLF Emotion: Group 2 starts with a high FLF emotion score of 3.231 before Task 1, which can be attributed to the cognitive challenge of starting with the most complex task. The anticipation of handling a difficult task right at the beginning likely heightens participants' discomfort.

Significant Emotional Relief After Task 1: After completing Task 1, the FLF emotion drops to its lowest point (2.308), indicating that participants experience a substantial emotional release after successfully finishing the most challenging task. This sharp reduction in FLF emotion suggests that once the cognitive load is lifted, participants feel relieved and less burdened by the task sequence.

Renewed Apprehension Before Task 2: FLF emotions rise again before Task 2 to 3.292, reflecting a moderate increase in discomfort as participants face the second task. Although Task 2 is less complex than Task 1, the renewed rise in FLF emotion could indicate apprehension about facing another task in the sequence.

Emotional Stabilization: As the task sequence progresses, FLF emotions gradually decrease during and after each task, but the decreases are not as significant as after Task 1. The final post-task score of 2.750 after Task 3 indicates that participants feel some relief after the task sequence is complete, but the emotional burden remains somewhat elevated compared to the immediate relief felt after Task 1.

Summary for Group 2

For Group 2, the emotional journey begins with high FLF emotions before starting Task 1 (complex), which is followed by a significant emotional relief after completing the most challenging task, as evidenced by the lowest FLF emotion score (2.308) after Task 1. However, participants experience renewed discomfort before Task 2 (middle complexity) and Task 3 (simple), though the overall FLF emotions gradually decrease as the tasks become less demanding. The task order plays a crucial role in shaping participants' emotional responses, with the initial high burden being alleviated early on by completing the hardest task, followed by more moderate relief after the simpler tasks.

Group 3

Before Task 1: The average FLF emotion score starts at 2.526, indicating moderate discomfort or apprehension before the first task, which is of middle complexity in this sequence.

During Task 1: As participants engage in Task 1, the FLF emotion decreases slightly to 2.368, suggesting that participants feel a bit less discomfort once they start working on the task.

After Task 1: After completing Task 1, the FLF emotion returns to 2.526, indicating that participants experience some stabilization in their emotional discomfort, but no significant relief is felt after this task.

Before Task 2: The FLF emotion rises sharply to 3.360 before starting Task 2, which is the most complex task in this sequence. This rise indicates heightened discomfort or apprehension as participants anticipate a more challenging task.

During Task 2: During Task 2, the FLF emotion continues to rise, reaching 3.440, marking the highest level of discomfort experienced during the entire sequence, indicating that participants struggle the most while performing the complex task.

After Task 2: After completing Task 2, the FLF emotion drops to 3.000, suggesting some relief, but the participants' discomfort remains elevated compared to the post-Task 1 period.

Before Task 3: The FLF emotion rises slightly to 3.478 before starting Task 3, indicating renewed apprehension, though the task is simpler compared to Task 2.

During Task 3: FLF emotions decrease to 3.261 during the task, showing a slight reduction in discomfort as participants engage in the final, simpler task.

After Task 3: After Task 3, the FLF emotion decreases further to 3.000, suggesting some emotional relief, but the overall level of discomfort remains higher than it was before the sequence began.

Key Observations for Group 3

Moderate Initial FLF Emotion: Group 3 begins with a moderate FLF emotion score of 2.526 before Task 1, which is of middle complexity. Participants likely feel moderate discomfort due to the cognitive challenge of starting with a moderately difficult task, but it is not overwhelming.

Emotional Stabilization After Task 1: After completing Task 1, the FLF emotion returns to the initial level (2.526), indicating that participants do not experience significant relief, but their discomfort remains stable. This suggests that finishing Task 1 does not drastically change their emotional state.

Sharp Increase in FLF Emotion Before and During Task 2: The FLF emotion rises sharply to 3.360 before Task 2 (complex task), and reaches its peak at 3.440 during the task, indicating that participants experience the highest levels of discomfort while engaging with the most challenging task in the sequence. The combination of cognitive demand and task complexity likely leads to heightened FLF emotions during this stage.

Partial Relief After Task 2: Although FLF emotions drop to 3.000 after completing Task 2, the decrease is not as significant as the initial emotional relief seen in other groups. This suggests that the emotional strain from Task 2 lingers, and participants do not fully recover after finishing the complex task.

Apprehension Before Task 3: Before starting Task 3 (simple task), the FLF emotion increases again to 3.478, indicating renewed apprehension, possibly due to accumulated emotional strain from the previous tasks, even though the upcoming task is simpler.

Gradual Relief During and After Task 3: During Task 3, FLF emotions decrease to 3.261, and drop further to 3.000 after the task is completed. This gradual decrease suggests that participants feel some emotional relief as the task becomes easier, though the relief is not as pronounced as the initial stages of the task sequence.

Summary for Group 3

For Group 3, the emotional journey begins with moderate FLF emotions before Task 1 (middle complexity), with participants experiencing slight stabilization in discomfort levels after completing the task. However, the emotional strain significantly increases before and during Task 2 (complex), where participants experience the highest levels of FLF emotions, reflecting the challenge posed by the most demanding task. Although there is some relief after Task 2, the discomfort remains relatively high, and participants face renewed apprehension before Task 3 (simple), likely due to accumulated emotional strain from earlier tasks. By the end of the sequence, FLF emotions decrease but remain elevated compared to the starting point, suggesting that the emotional burden persists despite the simplicity of the final task. The task order and increasing cognitive complexity significantly influenced Group 3's emotional dynamics. Starting with a moderately complex task provided a balanced introduction, followed by a noticeable increase in discomfort during the complex task. As the task sequence progressed, FLF emotions gradually decreased, suggesting a pattern of heightened fear during peak complexity and eventual emotional adjustment as the tasks concluded.

In summary, the analysis of FLF emotions across Groups 1, 2, and 3 reveals distinct trajectories shaped by task order and cognitive task complexity. Group 1, which completed tasks in the order of simple to complex, exhibited a steady increase in FLF emotions, peaking during Task 3. This pattern suggests that emotional discomfort heightened as task demands progressively increased. In contrast, Group 2, which started with the most complex task and ended with the simplest, showed a sharp decrease in FLF emotions after Task 1, with the lowest levels recorded following the initial task. This trend reflects a sense of emotional relief after overcoming the initial cognitive challenge, with discomfort continuing to diminish as tasks became less demanding. Group 3, beginning with a moderately complex task, progressing to the most complex, and ending with the simplest, experienced peak FLF emotions during Task 2, indicating that the greatest discomfort occurred in the middle of the sequence. While FLF emotions declined slightly by the end, they remained higher than at the start of the sequence.

Overall, the correlation between task order and cognitive complexity shapes the participants' emotional responses. Starting with more complex tasks tends to result in greater emotional relief as the sequence progresses, while starting with simpler tasks leads to a build-up of emotional discomfort as cognitive demands increase. These findings highlight the importance of considering both task difficulty and sequencing in managing emotional strain during task-based activities.

In the following section, the positive emotion dimension, specifically Foreign Language Enjoyment (FLE), will be analyzed in detail. This analysis will provide insights into how participants' enjoyment levels fluctuate throughout the task sequence and how factors such as task complexity and order contribute to their overall language learning experience.

4.3.2 Foreign Language Enjoyment (FLE)

This section discusses and analyzes data from the FLE questionnaire, which comprises five five-point Likert scale questions and one open-ended question, designed to investigate participants' enjoyment levels throughout the process of completing three tasks across the three groups. The analysis examines participants' levels of enjoyment, emphasizing the potential impacts of cognitive task complexity and task sequence on their experiences.

4.3.2.1 Dynamic Changes in FLE - Along CTC levels

The following section provides a detailed analysis of FLE feedback across three groups under different levels of cognitive complexity, based on questionnaire responses. This analysis includes raw data and visual representations, such as charts and graphs, to illustrate trends and patterns among the groups. Insights derived from the data are also explored, offering interpretations and explanations to reveal the influence of cognitive complexity on FLE. This comprehensive approach highlights the nuanced relationship between task complexity and language enjoyment.

The following charts display the Likert scale scores for five FLE questionnaire items, along with the final average scores, from a total of 60 participants across the three groups.

	Complexity Level	Question 1	Question 2	Question 3	Question 4	Question 5	Average value
	Simple Level	3.62	4.08	3.67	4	4	3.874
Group 1	Middle Level	3.79	3.96	3.88	3.92	4	3.91
	Complex Level	3.96	4.17	4.04	4.13	4.04	4.068
	Simple Level	3.6	4.08	3.84	3.92	3.84	3.856
Group 2	Middle Level	3.72	3.96	3.76	3.92	3.76	3.824
	Complex Level	3.37	3.92	3.5	3.71	3.71	3.642
	Simple Level	3.74	3.89	3.84	4	3.95	3.884
Group 3	Middle Level	3.33	3.81	3.48	3.71	3.71	3.608
	Complex Level	3.42	3.62	3.62	3.75	3.71	3.624

Figure 39 The FLE Scores for All Three Groups

The following line charts illustrate the trends in Foreign Language Enjoyment (FLE) across different levels of cognitive complexity.

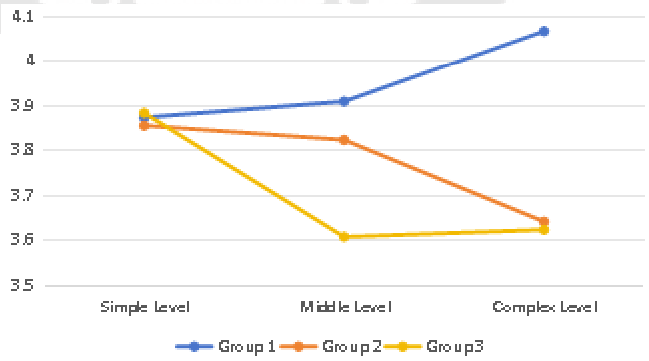

Figure 40 The FLE Scores Trends for All Three Groups

Group 1

Task Complexity (CTC) Analysis

Group 1 completed tasks in the simple-middle-complex order. Statistical tests examined how cognitive task complexity influenced the average Foreign Language Enjoyment (FLE) ratings across these three levels:

Simple ⟶ Middle

FLE scores showed a slight increase from 3.874 to 3.910, with p = 0.6148, which is above .05. This means the difference is not statistically significant. The small gap implies that transitioning from a simple to a moderately complex task did not substantially alter participants' enjoyment. Possibly, both tasks were deemed manageable enough that learners felt similarly comfortable, limiting any clear uptick in FLE.

Middle ⟶ Complex

A more substantial rise in FLE appears between Middle (3.910) and Complex (4.068), where p = 0.0072 is well below .05, indicating a significant increase. This result implies participants found the Complex tasks more engaging or stimulating, reflecting how well-calibrated difficulty can enhance involvement and satisfaction if it aligns with learners' capabilities.

Simple ⟶ Complex

Comparing the simplest (3.874) and most demanding tasks (4.068) yields p = 0.0451, which is just under .05, thus significant. Although the simpler task offered ease and familiarity, the additional challenge of the complex task evidently fostered greater accomplishment and enjoyment. This supports the notion that moderate-to-high difficulty—without being overwhelming—can be motivationally beneficial.

Overall, FLE remains roughly the same from Simple to Middle (no significance) but significantly increases going from Middle to Complex, and also when comparing Simple and Complex directly. These patterns underscore how a higher level of cognitive stimulation can heighten enjoyment once tasks push learners beyond a comfortable baseline yet remain achievable. Tasks that are too simple or exceedingly demanding may not elicit optimal engagement or satisfaction.

Group 2

Task Complexity (CTC) Analysis

Group 2 followed a complex-middle-simple ordering. The statistical tests here focused on how the Simple, Middle, and Complex levels of difficulty correlate with FLE:

Simple ⟶ Middle

Scores decreased slightly from 3.856 (Simple) to 3.824 (Middle), with $p = 0.4954$—well above .05, indicating no statistically significant change. Participants rated these two tasks similarly in terms of enjoyment, suggesting they perceived both levels as manageable and thus not drastically different in hedonic appeal.

Middle ⟶ Complex

A noticeable decline from 3.824 to 3.642 emerges, with $p = 0.0392$—below .05, thus significant. This suggests participants found the Complex tasks sufficiently demanding that it reduced their enjoyment. As difficulty intensified, satisfaction appeared to wane, hinting that tasks perceived as too challenging might limit engagement or comfort.

Simple ⟶ Complex

Comparing the extremes, from 3.856 (Simple) to 3.642 (Complex), yields $p = 0.0041$, also significant, underscoring a sharp decline in enjoyment between the easiest and hardest tasks. It highlights how tasks with heightened complexity can feel less accessible, dampening FLE.

Group 2 shows highest enjoyment at the simple level, with progressive drops as complexity rises. Notably, the biggest difference lies between Simple and Complex tasks, confirming participants prefer simpler, more manageable tasks, and experience decreased enjoyment as cognitive demands become pronounced.

Group 3

Task Complexity (CTC) Analysis

Group 3 tackled the tasks in a middle-complex-simple order. We examine FLE across the Simple, Middle, Complex tasks to see how difficulty affects enjoyment:

Simple ⟶ Middle

FLE decreased significantly from 3.884 to 3.608, where $p = 0.0084 < .05$. This sharp drop implies participants viewed the middle-level task as notably more demanding than the simplest one, resulting in lower enjoyment once they encountered moderate complexity.

Middle ⟶ Complex

Scores rose slightly from 3.608 to 3.624 ($p = 0.7915$), a non-significant difference. Participants likely perceived Middle and Complex tasks as comparably challenging, leading to no meaningful shift in enjoyment. They might have already been in a “more difficult” zone, so stepping further up in complexity did not further reduce their satisfaction—nor did it substantially increase it.

Simple ⟶ Complex

A direct comparison reveals a substantial decrease from 3.884 to 3.624, with $p = 0.0001 < .001$, indicating strong significance. The difference between simplest and most demanding tasks highlights how participants’ comfort and engagement drop considerably when faced with higher cognitive burdens. This underscores a preference for simpler tasks among Group 3, as enjoyment levels decline more dramatically at high complexity.

Group 3 rates the simplest task highest in enjoyment, with FLE sharply dropping as complexity increases, especially between the simplest and most demanding tasks. While the Middle⟶Complex step alone is not significant, the overall effect from Simple⟶Complex is quite pronounced. This aligns with the notion that participants in Group 3 prefer tasks that remain well within their comfort zone, and they register lower enjoyment once complexity surpasses that threshold.

In conclusion, Group 1: Shows minimal difference between Simple and Middle tasks but a clear significant increase in enjoyment at the Complex stage. This indicates a preference for moderate-to-challenging tasks that provide intellectual engagement without overwhelming learners. Group 2: Demonstrates significant declines in FLE as complexity rises from Simple to Complex, suggesting participants favor simpler demands for optimal enjoyment. Complex tasks, perceived as more strenuous, curtail satisfaction. Group 3: Likewise experiences the highest enjoyment at the simplest level, with a marked drop at higher complexity. The difference from Simple to Complex is strong, implying that heightened cognitive demands undermine enjoyment for most participants in this group.

These results highlight how varying degrees of complexity can either enhance or diminish Foreign Language Enjoyment, depending on whether the task demands align with participants' comfort and skill levels. While some learners (Group 1) thrive on more challenging tasks, others (Groups 2 and 3) find them less enjoyable. Educators may thus need to calibrate complexity carefully to maintain an optimal balance of engagement and enjoyment for different learner profiles.

4.3.2.2 Emotional Dynamics of FLE Throughout the Task Process

In this section, we analyze the Foreign Language Enjoyment (FLE) levels across the three tasks for each of the three groups, examining the dynamics of FLE throughout the entire task completion process. The table below presents the average FLE scores for each group, considering participants' cognitive levels and the specific task sequences they followed. This analysis is based on data collected from a total of 60 participants.

	Sequence	Average Value (FLE)
	Task 1 (Simple)	3.874
Group 1	Task 2 (Middle)	3.91
	Task 3 (Complex)	4.068
	Task 1 (Complex)	3.642
Group 2	Task 2 (Middle)	3.824
	Task 3 (Simple)	3.856
	Task 1 (Middle)	3.608
Group 3	Task 2 (Complex)	3.624
	Task 3 (Simple)	3.884

Figure 41 The FLE Scores for All Three Groups (along sequence)

To clearly observe the dynamic changes within each group, the line graph below offers a direct representation of the fluctuations across tasks.

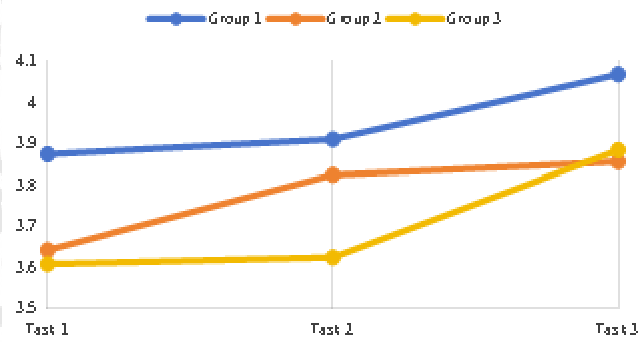

Figure 42 The FLE Scores Trends for All Three Groups (along sequence)

Group 1

When considering task sequence as a factor in the enjoyment levels of Group 1, the following trends can be observed based on the order of Task 1 (Simple), Task 2 (Middle), and Task 3 (Complex):

Task 1: The average FLE score of 3.874 reflects a moderate level of enjoyment at the beginning of the task sequence. While participants likely found the task manageable, their initial unfamiliarity with the process may have limited their engagement, resulting in moderate enjoyment.

Task 2: The average FLE score increases slightly to 3.910, indicating a small rise in enjoyment as participants progressed. This suggests that growing familiarity with the tasks and their format contributed to greater comfort and engagement, enhancing their enjoyment levels slightly.

Task 3: The average FLE score reaches its highest point at 4.068, indicating the peak of enjoyment by the final task. By this stage, participants had likely adapted to the task format, becoming more confident and comfortable. This familiarity and sense of accomplishment may have significantly boosted their enjoyment.

This analysis highlights the significant roles of cognitive task complexity and task sequence in shaping learners' Foreign Language Enjoyment (FLE). Contrary to the assumption that simpler tasks yield higher enjoyment, the findings reveal that participants' enjoyment tends to rise as tasks become more cognitively demanding, provided the difficulty remains within a manageable range. Challenging yet achievable tasks appear to foster greater engagement and a stronger sense of accomplishment, ultimately enhancing enjoyment.

Task sequence further influences emotional responses, as participants become more familiar with the task structure over time. This growing familiarity contributes to desensitization or adaptation, enabling learners to approach more complex tasks with increased confidence. Consequently, enjoyment often peaks during the final, most complex task, where the combination of heightened challenge and accumulated familiarity creates a more satisfying and engaging experience.

These findings suggest that effective task design should account for both the complexity of individual tasks and the sequence in which they are presented. A thoughtful arrangement of tasks can optimize learners' enjoyment and engagement, maximizing the potential benefits of task-based learning.

Group 2

When considering task sequence as a factor in the enjoyment levels of Group 2, the following observations can be made based on the order of Task 1 (Complex), Task 2 (Middle), and Task 3 (Simple):

Task 1: The average FLE score of 3.642 reflects a moderate level of enjoyment at the start of the sequence. Beginning with the most complex task likely posed high cognitive demands and unfamiliarity with the task structure, which may have dampened participants' initial enjoyment.

Task 2: The average FLE score rises to 3.824, indicating an increase in enjoyment as participants progressed. This improvement may be attributed to growing familiarity with the task structure, which likely reduced cognitive strain and enhanced engagement.

Task 3: The average FLE score peaks at 3.856, marking the highest level of enjoyment during the final task. By this stage, participants had likely adapted to the task format and gained confidence, allowing them to fully engage with and enjoy the simpler task.

The analysis of task sequence in Group 2 reveals that enjoyment levels steadily increased as participants progressed through the tasks. This upward trend in FLE scores suggests that growing familiarity with the task structure enabled participants to approach the tasks more confidently and comfortably. The highest enjoyment was recorded during the simplest task, indicating that the combination of task familiarity and reduced cognitive demands contributed to enhanced enjoyment. The gradual desensitization to the task structure and the ease of the final task likely played key roles in boosting participants' enjoyment throughout the sequence.

The findings from Group 2 underscore the importance of task simplicity in maximizing learner enjoyment. Contrary to the assumption that moderate challenges enhance engagement, the data clearly demonstrate that participants derived the most enjoyment from tasks with lower cognitive demands. This suggests that simpler tasks, particularly when placed at the end of a sequence, may create an optimal balance of confidence and satisfaction for learners.

In terms of task sequence, the results highlight the critical role of task familiarity in shaping learners' emotional responses. As participants progress through tasks, their growing familiarity with the format allows them to approach subsequent tasks with greater confidence and adaptability. This desensitization to the task structure, coupled with reduced cognitive strain during the final task, likely explains the peak in enjoyment observed in the last task.

These findings emphasize the importance of considering both cognitive demands and task sequence in task design. To optimize learners' enjoyment and engagement, it is crucial to offer a balance between challenging tasks and familiar structures. By carefully designing task sequences that build confidence and provide achievable challenges, educators can enhance learners' overall satisfaction and motivation in language learning activities.

Group 3

When examining task sequence as a factor in Group 3, based on the order of Task 1 (Middle), Task 2 (Complex), and Task 3 (Simple), the following observations can be made:

Task 1: The average FLE score of 3.608 reflects a moderate level of enjoyment at the start of the sequence. As the first task, participants may have been adjusting to the process, which could explain the moderate enjoyment level.

Task 2: The average FLE score remains relatively stable at 3.624, indicating little change in enjoyment as participants progressed to the second task. This stability may reflect a consistent level of engagement as participants became more accustomed to the task format.

Task 3: The average FLE score increases to 3.884, representing the highest enjoyment during the final task. By this point, participants likely felt more confident and familiar with the process, which contributed to their increased enjoyment.

The analysis of task sequence in Group 3 reveals that enjoyment levels were moderate at the beginning of the sequence and increased as participants progressed, peaking during the final task. This trend suggests that participants' comfort and confidence grew over time, leading to higher enjoyment by the end of the sequence. The data indicates that as participants became more familiar with the task structure and process, their enjoyment levels increased, with the final task being the most enjoyable. This underscores the role of task sequence in enhancing participants' overall experience as they advanced through the tasks.

The findings from Group 3 demonstrate that both cognitive task complexity and task sequence significantly influence learners' Foreign Language Enjoyment (FLE). Enjoyment was highest during the simplest task and lowest during the moderately complex tasks, indicating a preference for tasks with lower cognitive demands. Furthermore, the task sequence—beginning with more challenging tasks and concluding with simpler ones—played a pivotal role in boosting enjoyment toward the end of the sequence. These results emphasize the importance of carefully considering both task complexity and task order in learning design, as introducing simpler tasks later in the sequence may optimize learner engagement and satisfaction.

In general, based on the line chart, several key observations regarding the trend of Foreign Language Enjoyment (FLE) across the three groups can be made:

1) **All three groups show an upward trend in FLE as the tasks progress**, which may be a result of participants adapting to the task format and theme over time. As participants become more familiar with the nature of the tasks, their enjoyment increases, indicating a positive emotional adjustment as they proceed through the sequence.

2) **The magnitude of the increase in FLE differs significantly across the three groups.** This variation could be attributed to the different task orderings faced by each group, with different levels of cognitive complexity potentially impacting the rate at which participants adapt and increase their enjoyment. For example, Group 1

shows a steady and noticeable increase, while Group 3 demonstrates a more gradual rise, particularly between Tasks 1 and 2, followed by a sharp increase from Task 2 to Task 3.

3) **Group 1 consistently maintains the highest FLE scores across all tasks.** This could be explained by the task order sequence (Simple-Middle-Complex) aligning well with the participants' adaptability, allowing them to gradually build their confidence and comfort as they face increasingly complex tasks. The SMC sequence may provide an optimal balance of engagement and challenge, contributing to the highest levels of enjoyment compared to Groups 2 and 3.

4.4 Supplementary Data

In addition to the quantitative data from the recordings and questionnaires, qualitative data were also gathered. This includes two open-ended questions from the FLF and FLE questionnaires, respectively: "Please briefly tell us how you felt during the task" and "Please briefly tell us how you felt about the task". Additionally, video recordings from each experimental room were collected, allowing for the observation of participants' facial expressions and body language as they interacted with the computer.

4.4.1 Participants' Reflections on Open-ended Questions

For the two open-ended questions from the FLE and FLF questionnaires, we selected thematic form from the participants. These responses provide valuable insights into their emotional and cognitive experiences during the tasks. The following discussion will summarize and conclude these key points, focusing on participants' feelings of pressure, discomfort, engagement, enjoyment and adaptation throughout the process.

1) **Initial Pressure and Emotional Challenges:** Many participants reported feeling significant pressure at the beginning of the tasks, which affected their ability to clearly hear the computer's questions and led to a perceived decline in their language expression. For some, this discomfort was heightened by the content of the video, particularly the theme of theft, which made them uneasy. Several participants mentioned that this pressure peaked just before answering the computer's questions, as

the anticipation of responding heightened their anxiety. This initial tension was so intense for some that they felt the urge to complete the tasks quickly, with a few even experiencing physical symptoms such as trembling or agitation. Moreover, participants noted difficulties in recalling details from the videos or finding the right words to express their thoughts, which they attributed to the pressure and emotional strain of the situation.

2) **Enjoyment and Strategic Engagement:** Despite the initial pressure, many participants described the tasks as enjoyable and even exciting. The computer-based format provided a judgment-free environment, which participants appreciated as it allowed them to speak English without fear of criticism. Over time, participants adapted to the task structure, with some adopting strategic approaches such as planning their responses while watching the videos. These strategies reflected their growing comfort and engagement as the tasks progressed. However, some participants continued to feel frustrated or embarrassed by a perceived lack of vocabulary, which they believed hindered their ability to fully express themselves, limiting their enjoyment.

3) **Adaptation and Gradual Comfort:** As participants progressed through the tasks, many reported that their familiarity with the task format increased, leading to a gradual reduction in tension. This adaptation allowed them to better manage the challenges presented by the tasks. While the initial pressure made it difficult for some to focus or recall details, participants noted that as they became more comfortable, their ability to handle the demands of the tasks improved. This growing ease helped them navigate the tasks more effectively and contributed to a sense of progress.

4) **Sense of Accomplishment and Positive Emotional Responses:** Although the tasks initially caused significant pressure, many participants experienced moments of satisfaction when they successfully completed their responses. For example, one participant described the experience as “doing something good”, reflecting a positive emotional response tied to the task’s purpose. The context of assisting in a simulated police investigation added meaning to the tasks, creating a sense of purpose and contribution. This, in turn, helped to counterbalance the initial

discomfort and fostered a feeling of accomplishment, particularly during moments when participants felt they had successfully navigated the challenges.

Overall, the responses to the open-ended questions highlight the complex emotional and cognitive dynamics that participants experienced throughout the tasks. The initial pressure and discomfort they felt seem to have stemmed from both the task's unfamiliar format and the challenging nature of the crime-themed video content. However, as participants became more familiar with the tasks, many adapted and developed strategies to cope, leading to reduced discomfort and increased engagement. The sense of accomplishment reported by several participants, particularly those who felt they were contributing to a meaningful simulation, underscores the importance of task relevance and purpose in alleviating tension and fostering positive emotions. This mixture of pressure, engagement, and eventual enjoyment reflects the dynamic nature of task-based language learning, where challenges and discomfort can lead to personal satisfaction and growth. These insights suggest that providing purposeful, context-driven tasks can not only help learners manage their tension but also enhance their overall experience and sense of achievement.

4.4.2 Participants' Responses in Camera

The video recordings revealed additional insights into participants' behavior during the tasks, particularly in moments of heightened tension or word-finding difficulties. When participants struggled to express themselves, often due to pressure or not knowing the correct words, they would occasionally mix in their native language (Mandarin), using words such as "ranhou" (then), "jiushi" (exactly is), and "nage" (that). This behavior highlights how the urgency to communicate under pressure led to reliance on familiar linguistic tools.

Another common observation was the visible relief participants showed after answering a question or completing the entire task. Some could be seen exhaling deeply or audibly, signaling a reduction in stress. This physical response suggests that the pressure they experienced lessened once the task was completed.

Body language also provided useful insights. A few participants exhibited signs of discomfort during the task, such as scratching their heads, fidgeting, or appearing restless. Their posture tended to become more upright and focused while answering questions, with many leaning closer to the computer screen. Once the task was finished, participants often adopted more relaxed positions, leaning back in their chairs as they unwound from the task's demands. This shift in body language reflects the varying levels of pressure they experienced throughout the tasks.

The video recordings provide key insights into participants' non-verbal reactions under pressure. The use of native language fillers during moments of hesitation shows their struggle to maintain fluency when under stress. Visible signs of relief after answering questions, such as deep exhalations, further indicate the high pressure they experienced. Body language shifts, from tense and focused postures while responding to more relaxed positions after task completion, illustrate how their stress levels eased. These non-verbal cues, including language use and physical behavior, offer valuable understanding of how participants managed pressure and adapted during the tasks.

CHAPTER 5
CONCLUSION

This chapter synthesizes the key findings of the study, addressing the primary and secondary aims outlined at the beginning of the research. The empirical investigation focused on exploring the effects of cognitive task complexity on Foreign Language Fear (FLF), Foreign Language Enjoyment (FLE), and language performance. Meanwhile, the theoretical component aimed to establish distinctions between anxiety and fear, and further differentiate between fear-in and fear-of emotions within the context of foreign language learning. Grounded in frameworks such as Robinson's Cognitive Hypothesis and Skehan's Limited Attentional Capacity model, this study examined the nuanced link between cognitive demands and learners' emotional and linguistic responses.

The findings of this study provide insights into how cognitive complexity and task sequence influence learners' experiences and outcomes in technology-mediated, task-based language teaching (TMTBLT). Additionally, the study contributes to the theoretical discourse on the distinct roles of anxiety and fear in language learning, offering practical implications for designing effective learning environments that foster enjoyment and minimize fear-related barriers. The following sections summarize the major outcomes of the study, discuss their theoretical and pedagogical implications, and propose recommendations for future research.

5.1 The Cognitive Task Complexity vs. CALF Performance

The analysis of Type-Token Ratio (TTR) across cognitive task complexity reveals varying trends among the three groups. Group 1 shows a clear decline in lexical diversity as task complexity increases, highlighting the impact of heightened cognitive demands on reducing lexical variety. In contrast, Groups 2 and 3 demonstrate more stable or slightly increasing TTR values, suggesting that some participants were able to maintain or enhance their lexical diversity despite the increasing task complexity. Given the inconsistent trends observed across the three groups, it might be premature to conclude that cognitive task complexity could exert a uniform influence on lexical diversity. These findings suggest that individual differences in managing cognitive demands may play a significant role, leading to varied outcomes.

The analysis of syntactic complexity, measured through Mean Length of AS-unit (MLAS) and Mean Length of Clause (MLC), reveals consistent patterns across all groups, highlighting the significant influence of cognitive task complexity on syntactic elaboration. Notably, moderate task complexity emerges as a key factor, often eliciting the highest levels of syntactic complexity across measures. For Group 1, both MLAS and MLC increase significantly from simple to middle tasks before stabilizing or slightly declining at the complex level, indicating that syntactic performance peaks when tasks provide a moderate challenge. In Group 2, MLAS declines from simple to middle tasks but recovers slightly at the complex level, while MLC remains stable initially but drops sharply at the complex level, reflecting difficulties in sustaining syntactic elaboration under greater demands. Group 3 demonstrates distinct adaptability, with MLAS peaking at the middle level and showing only a modest decline under complex conditions, while MLC steadily increases across all levels of complexity. These findings underscore the unique role of moderate task complexity in fostering optimal syntactic complexity, while higher demands may lead to diminishing returns or varied performance depending on group-specific factors. Balancing cognitive challenge, therefore, appears essential for maximizing syntactic elaboration, with moderate complexity offering a particularly favorable condition.

The analysis of Total Number of Words (TNW) and Total Number of Syllables (TNS) reveals varied patterns of productivity across the three groups, reflecting differing responses to cognitive task complexity. Group 1 shows a steady increase in both TNW and TNS as task complexity rises, with notable changes in TNW between simple and middle tasks and in TNS between simple and complex tasks, suggesting that higher task complexity generally enhances productivity, albeit with moderate statistical confidence. Similarly, Group 3 demonstrates consistent upward trends in both measures, with the highest productivity observed at the complex level, indicating adaptability to increased cognitive demands despite limited statistical support. In contrast, Group 2 exhibits remarkable stability in TNW and minimal variation in TNS across all levels of task complexity. A slight increase in TNS from simple to middle tasks, followed by minor declines at the complex level, suggests that task complexity has little impact on productivity for this group. These findings collectively highlight that while higher task complexity tends to promote productivity in Groups 1 and 3, the stable oral output observed in Group 2 underscores the influence of group-specific dynamics and individual responses to cognitive demands.

The analysis of fluency, measured by Words Per Minute (WPM) and Syllables Per Second (SPS), reveals varied responses to task complexity across the three groups. Group 1 demonstrates a consistent upward trend in both measures as task complexity increases, with the most significant improvement observed between Simple and Complex tasks. While the SPS trend is supported by higher statistical confidence, the WPM trend shows limited statistical backing, suggesting that Group 1 adapts well to rising task demands and achieves greater fluency with increasing complexity. In contrast, Group 2 exhibits a steady decline in fluency across both measures as task complexity rises, culminating in the lowest levels at the Complex task. This downward trend is statistically significant for the Simple-to-Complex comparison, underscoring the challenges posed by increased task demands. Group 3 shows more fluctuating patterns, with declines at the Middle task followed by slight recoveries at the Complex level, though low statistical confidence suggests relative stability in fluency for this group. These findings collectively highlight that task complexity affects fluency differently across groups, with Group 1 showing the strongest adaptability, Group 2 the greatest difficulty, and Group 3 maintaining moderate stability.

The analysis of Error-Free Clause Ratio (EFCR) reveals distinct patterns across the three groups. Group 1 demonstrates consistent growth in EFCR as task complexity increases, with the most significant improvement observed between Simple and Complex tasks, supported by strong statistical confidence. However, comparisons between adjacent levels show low confidence, suggesting more gradual improvements.

Group 2 exhibits relatively stable EFCR across all complexity levels, with minimal variation and consistently low statistical confidence, indicating that linguistic accuracy remains largely unaffected by task complexity. In contrast, Group 3 shows steady improvements in EFCR as complexity rises, reaching the highest accuracy at the Complex level. While confidence is low for adjacent comparisons, the Simple-to-Complex trend is strongly supported statistically. These findings highlight that task complexity can enhance linguistic accuracy, particularly for Groups 1 and 3, but its impact varies depending on task sequence and group dynamics.

Having established how varying levels of cognitive task complexity can shape multiple dimensions of CALF performance, we now turn to the role of task sequence in influencing learners' linguistic output. While the previous section concentrated on complexity-driven changes, the following discussion explores how the order in which tasks are presented further modulates performance and interacts with the cognitive demands already identified.

5.2 The Task Sequence vs. CALF Performance

The analysis of Type-Token Ratio (TTR) across the task sequence reveals a consistent downward trend in lexical variety among all three groups, highlighting the influence of task order on lexical performance. Both Group 1 and Group 2 show a high degree of confidence in the effect of task sequence, with later tasks consistently demonstrating reduced lexical diversity, irrespective of complexity. Group 3, while displaying some fluctuations, also suggests a similar trend, albeit with moderate confidence in the results. These findings indicate that task sequence systematically constrains participants' ability to produce diverse vocabulary, likely due to accumulating cognitive demands or task-related fatigue. A possible explanation for this trend is that participants may initially attempt to use a wide range of vocabulary to describe scenes and answer questions. However, as the tasks progress, they may shift their focus toward ensuring accuracy and comprehensibility, opting for safer and more familiar words to maintain effective communication. Supporting this explanation, responses to the open-ended questionnaire revealed that many participants expressed dissatisfaction with their performance, citing limitations in their vocabulary as a significant factor that restricted their ability to express themselves fully.

The analysis of syntactic complexity, measured by the mean length of AS-units and clauses, reveals distinct yet interconnected patterns across the three groups, emphasizing the influence of task sequence on syntactic performance. For AS-units, Group 1 shows a sharp increase from Task 1 to Task 2, followed by a slight decline in Task 3, indicating significant variation in complexity as tasks progress. Group 2 exhibits a dip in Task 2 before recovering in Task 3, suggesting moderate variability with weaker evidence of task sequence effects. Group 3 demonstrates a gradual decline in AS-unit length across tasks, reflecting a steady reduction in syntactic elaboration. For clauses, all groups follow a consistent trend: clause length peaks during Task 2 before declining in Task 3. Group 1 and Group 2 show significant increases from Task 1 to Task 2, followed by modest decreases, while Group 3 maintains stability between Task 1 and Task 2 but experiences a steep drop in Task 3. These trends suggest that Task 2 often elicits the most syntactically complex responses, irrespective of cognitive complexity, likely due to improved fluency or adaptation, while the decline in Task 3 may result from cognitive fatigue or task-related constraints. Despite variations in magnitude, the overall pattern highlights a common influence of task sequence on syntactic complexity across groups.

The analysis of productivity, measured by the total number of words and syllables, reveals varied patterns across the three groups, reflecting differing responses to task sequence. For the total number of words, Group 1 shows an increasing trend across tasks, though with low confidence in the effect of task sequence. Group 2 exhibits a stable pattern with minimal fluctuations, providing little evidence of a significant task sequence impact. In contrast, Group 3 demonstrates a sharp increase from Task 1 to Task 2, followed by a notable decline in Task 3, indicating moderate confidence in the influence of task sequence on productivity. Regarding the total number of syllables, Groups 1 and 3 show strong evidence of task sequence effects, with both groups displaying a significant increase from Task 1 to Task 2, followed by a further increase in Group 1 and a decline in Group 3 during Task 3. For Group 2, the impact of task sequence on syllable production is less evident, with fluctuations lacking strong statistical support. These results highlight that task sequence significantly influences productivity for Groups 1 and 3, particularly in terms of syllable production, while Group 2 shows greater stability and less pronounced effects.

The analysis of fluency, measured by the number of words per minute (WPM) and syllables per second (SPS), reveals consistent upward trends across all three groups, highlighting the influence of task sequence on fluency. For WPM, all groups demonstrate an increase across tasks, though the magnitude and confidence levels vary. Group 2 exhibits the most pronounced trend with strong statistical confidence, indicating a significant impact of task sequence on verbal fluency. Groups 1 and 3 also show upward trends, but the evidence supporting the effect of task sequence is less robust. Similarly, for SPS, all groups display an increase across tasks, with Groups 1 and 2 showing stronger trends and higher confidence levels, suggesting that task sequence positively influences syllable production speed. Group 3 follows the same general trajectory but with lower statistical support. Despite differences in the magnitude and confidence of these effects, the shared pattern of increasing WPM and SPS across tasks underscores the positive role of task sequence in enhancing speech fluency over time, with Groups 1 and 2 showing the clearest improvements. Given the consistent upward trends across all three groups, it could be inferred that fluency might be influenced more by practice effects and participants' increasing familiarity with the task format over time, rather than by the complexity of the tasks themselves. This suggests that, compared to other aspects of language performance, fluency may benefit to a greater extent from repeated exposure to the task environment, allowing participants to gradually adapt and potentially improve their speech output with successive tasks.

The analysis of accuracy, measured by the Error-Free Clause Ratio (EFCR), reveals varied patterns across the three groups, highlighting differing influences of task sequence on accuracy. Group 1 exhibits a clear upward trend in EFCR across tasks, with a high degree of statistical confidence, suggesting that task sequence significantly enhances the accuracy of clause production. In contrast, Groups 2 and 3 display minor fluctuations in EFCR values, with low statistical confidence in the effect of task sequence, indicating limited or inconsistent impacts. These results suggest that while task sequence appears to positively influence accuracy in Group 1, the effects are less evident or negligible for Groups 2 and 3, reflecting variability in how different groups respond to task progression.

In summary, these findings collectively underscore the pivotal role of task sequence in shaping multiple dimensions of L2 speaking performance, albeit in distinct ways across groups. While lexical variety (TTR) consistently declined over successive tasks—likely reflecting cumulative cognitive load or fatigue—syntactic complexity often peaked at the second task before dropping, suggesting a short-lived boost in elaboration. Productivity results varied more by group, with some showing clear sequence-driven changes and others remaining stable. In contrast, fluency exhibited a generally upward trend across tasks, pointing to potential practice effects or increased familiarity with the task environment over time. Finally, accuracy benefited most strongly from sequence in Group 1, whereas Groups 2 and 3 showed minimal or inconsistent changes. Taken together, these patterns highlight how repeated task exposure can simultaneously facilitate certain facets of language use (e.g., fluency, accuracy for some groups) while constraining others (e.g., lexical variety), reinforcing the idea that cognitive adaptation, resource allocation, and task fatigue dynamically interact to shape L2 performance.

5.3 Cognitive Task Complexity vs. Emotion

The analysis of the relationship between cognitive task complexity (CTC) levels and Foreign Language Fear (FLF) emotions reveals distinct patterns across the three groups, highlighting varying responses to increasing task complexity. For Group 1, FLF ratings consistently rose as task complexity increased, with significant differences observed between Simple and Middle tasks and between Middle and Complex tasks. The most pronounced rise occurred between Simple and Complex tasks, demonstrating a strong relationship between higher task complexity and heightened fear-related emotions. In contrast, Group 2 exhibited a downward trend in FLF ratings as task complexity increased. While the changes between Simple and Middle tasks and Middle and Complex tasks were not statistically significant, the comparison between Simple and Complex tasks revealed a significant reduction in FLF emotions, suggesting a decrease in fear as tasks became more challenging. Group 3 displayed a unique pattern, with FLF ratings decreasing slightly from Simple to Middle tasks before rising significantly from Middle to Complex tasks. This increase was most notable between Middle and Complex tasks, indicating that more challenging tasks elicited stronger fear-related responses after an initial reduction. Overall, the results demonstrate that task complexity influences FLF emotions differently across groups, with Group 1 experiencing a steady increase, Group 2 showing a decline, and Group 3 exhibiting a more dynamic fluctuation influenced by the progression of complexity.

The analysis of Foreign Language Enjoyment (FLE) across the three groups reveals distinct patterns in response to task complexity, highlighting its significant role in shaping enjoyment levels. For Group 1, FLE remained relatively stable from Simple to Middle tasks, suggesting that both tasks were perceived as manageable and familiar. However, enjoyment increased noticeably from Middle to Complex tasks, indicating that the added challenge and cognitive stimulation of the Complex tasks enhanced participants' engagement and satisfaction. This pattern demonstrates how appropriately challenging tasks can foster a greater sense of accomplishment and enjoyment. In contrast, Group 2 exhibited a clear downward trend in FLE as task complexity increased. Enjoyment was highest during simpler tasks and declined progressively, with the sharpest decrease occurring between Simple and Complex tasks. This decline highlights the reduced comfort and engagement associated with more demanding tasks, suggesting a preference for simpler, more accessible activities. Group 3 showed a distinct pattern, with enjoyment decreasing sharply from Simple to Middle tasks, reflecting the increased demands of the middle-level task. While FLE remained relatively stable between Middle and Complex tasks, a significant decline was observed when comparing Simple and Complex tasks. This pattern also underscores a preference for simpler tasks, as heightened cognitive demands led to diminished satisfaction. Collectively, these findings demonstrate that task complexity plays a critical role in shaping FLE, with enjoyment peaking when tasks strike a balance between challenge and participants' abilities, while overly demanding tasks result in reduced satisfaction.

In summary, varying levels of cognitive task complexity clearly shape both fear and enjoyment in distinct ways across the three groups, suggesting that each learner cohort responds differently to increasing demands. Group 1 tends to exhibit heightened fear and boosted enjoyment under more challenging tasks, Group 2 generally shows reduced fear but diminished enjoyment as complexity rises, and Group 3 experiences more dynamic fluctuations in both emotional domains. These patterns underscore the nuanced connection between cognitive load and emotional engagement, where tasks that either exceed or align with participants' capacities can significantly modulate their fear and enjoyment. Building on these insights, the next section (5.4) shifts focus to task sequence, examining how the order of tasks further influences learners' emotional responses.

5.4 Task Sequence vs. Emotion

The analysis of FLF (Foreign Language Fear) emotions reveals distinct patterns shaped by task sequence across the three groups. In Group 1 (SMC), FLF emotions followed a steady upward trend, starting at the lowest levels in Task 1 (Simple), increasing in Task 2 (Middle), and peaking in Task 3 (Complex). This progression reflects a gradual intensification of FLF as tasks became more demanding. In Group 2 (CMS), FLF emotions also rose progressively, beginning at the lowest level in Task 1 (Complex), increasing in Task 2 (Middle), and reaching the highest point in Task 3 (Simple). This pattern suggests a cumulative build-up of FLF emotions, potentially influenced by mental fatigue or task-related strain. In Group 3 (MCS), FLF emotions increased from Task 1 (Middle) to Task 2 (Complex), peaking during the most challenging task, before showing a slight decline in Task 3 (Simple). This fluctuation highlights how task sequence and complexity interact, producing varied emotional responses. Overall, these findings emphasize the dynamic relationship between task sequence and FLF emotions, with each group exhibiting unique trends influenced by both task order and cognitive demands.

The analysis of task sequence across all three groups reveals its significant impact on Foreign Language Enjoyment (FLE), highlighting the connection between cognitive demands, familiarity, and task order. In Group 1, enjoyment steadily increased as participants progressed from simpler to more complex tasks, with the highest levels observed in the final task (Complex). This trend suggests that growing familiarity and confidence, coupled with an appropriately challenging task, fostered greater engagement and satisfaction. In Group 2, enjoyment followed a similar upward trajectory but was most pronounced during the final, simplest task, reflecting the positive influence of reduced cognitive demands and accumulated familiarity. Group 3 displayed a moderate level of enjoyment at the start, stability during the middle task, and a peak during the final, simplest task, further emphasizing the role of familiarity and the ease of the last task in enhancing satisfaction. These findings demonstrate that task sequence itself significantly shapes learners' emotional responses, as familiarity and adaptation over time enable participants to approach later tasks with greater ease and engagement. Thoughtful task sequencing, particularly placing simpler tasks toward the end, can optimize learners' enjoyment and create a more satisfying learning experience.

Given the similar overall structure of the three tasks, with differences only in specific events and response strategies, it would be expected that novelty decreases as the task sequence progresses. However, the results reveal an overall upward trend in Foreign Language Enjoyment (FLE) across the three groups. This suggests that participants' sense of enjoyment is primarily driven by their growing familiarity with the task format, a sense of control, and the accomplishment derived from handling tasks more effectively, rather than the novelty of the tasks themselves. As participants advanced through the sequence, increasing familiarity likely reduced cognitive load, enhanced their confidence, and strengthened their sense of mastery, ultimately contributing to greater enjoyment. This finding underscores the critical role of familiarity and task mastery in fostering positive emotional experiences in foreign language learning.

5.5 Linking Results with Prior Empirical Studies and Theoretical Frameworks

The present study's findings on CAF dimensions and emotional responses echo several previous task-based studies, while also diverging in key aspects—especially due to the introduction of Foreign Language Fear (FLF) as a distinct construct.

Language Performance (CAF):

Regarding linguistic output, the current study found that syntactic complexity increased under moderate task complexity, while lexical variety (TTR) tended to decline at higher cognitive load levels. These outcomes align with the Limited Attentional Capacity Model (LAC), which holds that learners must prioritize dimensions such as fluency or accuracy under cognitive pressure, often resulting in reduced lexical sophistication. (Xu et al., 2023) reported similar patterns: task complexity did not significantly enhance syntactic complexity, accuracy, or fluency, but did reduce lexical complexity and functional adequacy, thus supporting Skehan's LAC perspective. Similarly, (Donate, 2018) found that participants showed improved fluency and complexity under more difficult tasks but at the cost of accuracy, confirming the presence of performance trade-offs predicted by LAC.

In contrast, Cognition Hypothesis (CH) proponents argue that increased complexity should enhance syntactic complexity and accuracy, particularly when resource-directing features are manipulated. Robinson (2005) and (Kuiken & Vedder, 2008) found partial support for this in narrative tasks, especially in terms of grammatical complexity.

However, the current study's findings did not reveal a universal improvement in these dimensions across all groups, suggesting that learners' adaptation strategies and sequencing order played moderating roles. This nuanced result is consistent with Rahimi (2019) and Frear & Bitchener (2015), who also observed mixed effects across CAF dimensions in writing tasks with manipulated complexity.

Task Sequence:

The observed task-sequencing effects, particularly the facilitative role of progressing from simple to complex (as seen in Group 1), provide some support for Robinson's SSARC model. This model suggests that learners benefit when allowed to build fluency and accuracy on simpler tasks before restructuring and complexifying their oral production. However, consistent with findings from (Baralt et al., 2014) and (Malicka, 2020), the current study also indicates that task sequencing alone may not produce dramatic differences in overall performance unless combined with carefully scaffolded task design. Thus, while sequencing shapes learners' real-time processing and emotional engagement, the primary determinant of linguistic output remains the cognitive demand embedded in the task itself.

Emotion (FLE and FLF):

Emotionally, the study offers new insights by distinguishing Foreign Language Fear (FLF) from traditional Foreign Language Anxiety (FLA)—a point that has not been clearly articulated in earlier empirical studies. Most prior research, such as (Dewaele & Dewaele, 2017), measured enjoyment and anxiety using trait-oriented questionnaires, without isolating momentary fear during task performance. (Donate, 2018) acknowledged the dynamic nature of state anxiety but still did not separate fear as an independent affective state during language use.

The current study shows that FLF fluctuates in response to immediate task complexity, supporting the theoretical distinction that fear is an in-situ response to present threat, while anxiety is a broader anticipatory emotion. Learners in the present study reported rising fear during higher-complexity tasks, a pattern consistent with Donate's observation that fear levels increased with cognitive demands—but without a clear link to overall performance decline. The differentiation of FLF as a short-term, task-induced emotion enables a clearer interpretation of why learners may exhibit defensive behaviors in complex L2 situations—even when general FLA levels are low.

Foreign Language Enjoyment (FLE):

The enjoyment findings in this study reinforce (Dewaele & MacIntyre, 2014, 2016) view that FLE is shaped both by internal learner factors and task conditions. In the current study, moderate challenge enhanced FLE, especially when tasks were sequenced gradually. This mirrors earlier findings that learners tend to report more enjoyment when they feel sufficiently competent and are offered autonomy or stimulation in their task environment. However, when complex tasks appeared first (Group 2), enjoyment dropped sharply—a phenomenon not extensively explored in previous research. Thus, this study adds that enjoyment can be fragile when learners are not adequately scaffolded into difficulty.

In sum, the study's CAF results partially corroborate earlier work under both CH and LAC, showing clear trade-offs and occasional syntactic gains, but also highlighting the role of sequencing and individual differences. Emotionally, the contribution lies in the empirical validation of FLF as separate from FLA—a perspective supported theoretically but not previously operationalized in empirical TBLT studies. These distinctions broaden the scope for future research, suggesting that more refined emotional constructs can clarify how learners respond to task demands not only cognitively, but also affectively.

5.6 Pedagogical Implications

The findings of this study underscore that task design, particularly in terms of cognitive complexity and task sequence, has an appreciable impact on both language performance and learner emotions in TMTBLT contexts. First, the results highlight that moderate task complexity frequently yields optimal syntactic elaboration (e.g., Group 1's peak in MLC/MLAS at the middle level) and a sweet spot for enjoyment (FLE) without driving up fear (FLF). This indicates that, rather than assigning uniformly simple or excessively complex tasks, instructors could calibrate difficulty to fall within a zone that stretches learners' abilities while still feeling achievable. In practical terms, this means crafting tasks that demand increased cognitive engagement—yet remain transparently structured and offer sufficient scaffolding—so learners experience a sense of challenge and success instead of feeling overwhelmed.

Second, task sequence emerges as a powerful moderator of learner performance and emotions. Some groups (e.g., Group 1) benefit from ascending complexity, gradually building competence and confidence, thus reinforcing positive emotions and mitigating foreign language fear. However, other learners (e.g., Group 2) display stronger fluency gains or higher enjoyment when tasks start with more demanding activities and subsequently transition to simpler tasks, reflecting a "relief" or practice effect. Consequently, teachers should consider rotating complexity levels or arranging tasks such that learners spend time on moderately challenging tasks, gain mastery and comfort, and then either tackle an advanced form of the task or revisit simpler tasks for consolidation. By doing so, instructors can strategically mix ascending and descending complexities across a curriculum or unit, allowing students to cycle between pushing their upper limits and reinforcing foundational skills.

Lastly, the connection between negative and positive emotions—fear (FLF) and enjoyment (FLE)—provides insights into emotional support strategies. Complex tasks, if not aligned with learners' readiness, can heighten fear or reduce enjoyment. However, appropriately supported, well-sequenced tasks can transform this heightened demand into greater engagement and satisfaction. Teachers might therefore implement pre-task briefings, within-task scaffolding, or post-task reflection to buffer anxiety and enhance enjoyment. For instance, giving short, focused feedback or encouraging peer collaboration can help learners feel safer taking on more cognitively demanding tasks. Additionally, paying attention to learners who display signs of frustration or excessive fear can guide when to revert to simpler tasks, encourage more supportive feedback, or offer alternative approaches. Ultimately, thoughtful task design and sequencing—combined with targeted emotional support—can maximize both language outcomes and emotional well-being in TMTBLT settings.

5.7 Limitations and Future Directions

Despite the valuable insights gained regarding foreign language fear (FLF) and foreign language enjoyment (FLE), as well as the effects of cognitive task complexity and task sequencing on learners' performance, this study faces several limitations. First, the sample size and representativeness remain limited: the participants (N=60) were drawn from a single institution or region, which restricts the external validity of the findings and may not fully capture the diversity of learners in different educational contexts. Second, although the FLF and FLE questionnaires used in this study are practical and straightforward, they have not undergone extensive reliability and validity testing in larger or cross-cultural populations. Consequently, their internal consistency, factor structures, and cross-linguistic stability require further verification.

In addition, the study's task environment and setting focused primarily on computer-mediated spoken tasks, which may not fully replicate real-world classroom interactions or face-to-face communication. External factors such as technical glitches, ambient noise, or individual differences (e.g., fatigue, fluctuating attention) could also have influenced the emotional responses and performance data. Furthermore, although multiple task orders (SMC, CMS, MCS) were designed to distinguish between "complexity" and "sequence", it remains challenging to completely disentangle the interactive effects of these two variables. In some groups, the alignment of increasing complexity with sequential progression makes it difficult to ascertain whether heightened emotional responses stem from rising complexity or simple chronological order (or fatigue). Lastly, while this study introduced a theoretical distinction between "fear-in" and "fear-of," the empirical work focused on a broader assessment of FLF without operationalizing these two sources of fear separately, limiting deeper exploration of the specific origins and nuances of fear responses.

Future Directions

Building on these limitations, future research could enhance and expand the current findings in several ways. First, expanding the sample size and diversity—for instance, by recruiting learners of various academic backgrounds, language proficiencies, and cultural contexts—would increase external validity and illuminate potential group-level differences. Second, more rigorous development and validation of the FLF and FLE instruments are warranted. In particular, creating refined measures that capture distinctions such as “fear-in” versus “fear-of”, or more nuanced aspects of enjoyment, could clarify the conceptual framework and improve cross-cultural applicability.

Moreover, the inclusion of intervention studies (e.g., providing pre-task emotional support or post-task feedback) could shed light on how fear and enjoyment evolve over longer periods of language learning and how targeted strategies might mitigate negative emotions or bolster positive ones. It would also be valuable to test these emotional constructs in more varied contexts, such as realistic classroom settings, group discussions, role-plays, or face-to-face interactions, and to incorporate objective measures (e.g., facial expression analysis, physiological data) to complement self-report data and reduce potential response biases.

Finally, to disentangle the complex relation among task complexity, task sequencing, and multiple measurement time points (before/during/after tasks), future research could adopt more elaborate designs (e.g., incorporating additional complexity levels or randomized sequences) and advanced statistical modeling (e.g., multilevel modeling, structural equation modeling). By comprehensively addressing these aspects, scholars can develop a more in-depth understanding of how emotions shape and are shaped by task design, ultimately informing more effective foreign language pedagogy and emotional support strategies.

REFERENCES

Adams, R., & Nik Mohd Alwi, N. A. (2014). Chapter 3. Prior knowledge and second language task production in text chat. In *Technology-mediated TBLT*. John Benjamins.

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's Construct of Foreign Language Anxiety: The Case of Students of Japanese. *The Modern Language Journal*, *78*(2), 155-168.

Albert, A. (2011). Chapter 9. When individual differences come into play: The effect of learner creativity on simple and complex task performance. In R. Peter (Ed.), *Second Language Task Complexity* (pp. 239-266). John Benjamins Publishing Company.

Alenezi, S. M. (2020). Foreign language enjoyment and anxiety among the Northern Borders University EFL students: Links to gender and majors, *39*(185ج3), 1203-1233.

Almusharraf, A. M. (2021). Bridging the gap: Saudi Arabian faculty and learners' attitudes towards first language use in EFL classes. *Issues in Educational Research*, *31*(3), 679-698.

Alwi, N. A. N. M., Adams, R., & Newton, J. (2012). Writing to learn via text chat: Task implementation and focus on form. *Journal of second language writing*, *21*(1), 23-39.

Alwi, N. M., & Aloesnita, N. (2010). *Examining the language learning potential of a task-based approach to synchronous computer-meditated communication* Open Access Te Herenga Waka-Victoria University of Wellington].

Andon, N., & Eckerth, J. (2009). Chacun à son gout? Task-based L2 pedagogy from the teacher's point of view. *International journal of applied linguistics*, *19*(3), 286-310.

Aronson, E. (2002). Building empathy, compassion, and achievement in the jigsaw classroom. In *Improving academic achievement* (pp. 209-225). Elsevier.

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.

Baralt, M., Gilabert, R., & Robinson, P. (2014). Task complexity and task sequencing in traditional versus online language classes. *Task sequencing and instructed second language learning*, *59122*.

Baralt, M., & Gurzynski-Weiss, L. (2011). Comparing learners' state anxiety during task-based interaction in computer-mediated and face-to-face communication. *Language Teaching Research*, *15*(2), 201-229.

Barcomb, M., & Cardoso, W. (2020). Rock or Lock? Gamifying an online course management system for pronunciation instruction. *Calico Journal*, *37*(2), 127-147.

Bicen, H., & Kocakoyun, S. (2018). Perceptions of students for gamification approach: Kahoot as a case study. *International Journal of emerging technologies in learning*, *13*(2).

Blanchard, R. J., Kleinschmidt, C. F., Fukunaga-Stinson, C., & Blanchard, D. C. (1980). Defensive attack behavior in male and female rats. *Animal Learning & Behavior*, *8*, 177-183.

Bolles, R. C. (1970). Species-specific defense reactions and avoidance learning. *Psychological review*, *77*(1), 32.

Bolles, R. C., & Fanselow, M. S. (1980). A perceptual-defensive-recuperative model of fear and pain. *Behavioral and Brain Sciences*, *3*(2), 291-301.

Brindley, G. (1987). Factors affecting task difficulty. *Guidelines for the development of curriculum resources for the adult migrant education program/National Curriculum Resource Centre*.

Brown, G., Malmkjær, K., & Williams, J. (1996). *Performance and competence in second language acquisition*. Cambridge university press.

Brown, H. (2007). Douglas. Teaching by Principles: An Interactive Approach to Language Pedagogy. San Francisco State University. In: Pearson Education.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, *32*, 21.

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.

Calvo-Ferrer, J. R. (2017). Educational games as stand-alone learning tools and their motivational effect on L 2 vocabulary acquisition and perceived learning gains. *British Journal of Educational Technology*, *48*(2), 264-278.

Candlin, C. (1987). Towards task-based language learning. *Language learning tasks*, *7*, 5-22.

Carless, D., Shehadeh, A., & Coombe, C. (2012). TBLT in EFL settings. *Task-based language teaching in foreign language contexts*, 345-358.

Celce-Murcia, M. (2001). Teaching English as a second or foreign language. In: Heinle & Heinle.

Chang, S.-C. (2011). A contrastive study of grammar translation method and communicative approach in teaching English grammar. *English language teaching*, *4*(2), 13.

Chen, K. T.-C. (2021). The effects of technology-mediated TBLT on enhancing the speaking abilities of university students in a collaborative EFL learning environment. *Applied Linguistics Review*, *12*(2), 331-352.

Cheng, Y. s., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language learning*, *49*(3), 417-446.

Csikszentmihalyi, M., & Nakamura, J. (2011). Positive psychology: Where did it come from, where is it going?

Cutrone, P. (2009). Overcoming Japanese EFL learners' fear of speaking. *Language studies working papers*, *1*, 55-63.

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: a systematic review. *Computer Assisted Language Learning*, *34*(7), 934-957.

Dendrinos, B. (1992). *The EFL textbook and ideology*. Grivas.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments,

Dewaele, J.-M., & Alfawzan, M. (2018). Does the effect of enjoyment outweigh that of anxiety in foreign language performance? *Studies in second language learning and teaching*, *8*(1), 21-45.

Dewaele, J.-M., & Dewaele, L. (2017). The dynamic interactions in foreign language classroom anxiety and foreign language enjoyment of pupils aged 12 to 18. A pseudo-longitudinal investigation. *Journal of the European Second Language Association*, *1*(1), 12-22.

Dewaele, J.-M., & Dewaele, L. (2018). Learner-internal and learner-external predictors of willingness to communicate in the FL classroom. *Journal of the European Second Language Association*, *2*(1), 24-37.

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in second language learning and teaching*, *4*(2), 237-274.

Dewaele, J.-M., & MacIntyre, P. D. (2016). Foreign language enjoyment and foreign language classroom anxiety: The right and left feet of the language learner. *Positive psychology in SLA*, *215*(236), 9781783095360-9781783095010.

Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, *22*(6), 676-697.

Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & education*, *63*, 380-392.

Donate, Á. (2018). *Cognitive task complexity, foreign language anxiety and L2 performance in Spanish: A task-based language teaching perspective*. Georgetown University.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, *4*(3), 275-300.

East, M. (2012). Task-based language teaching from the teachers' perspective.

Ebadi, S., Rasouli, R., & Mohamadi, M. (2023). Exploring EFL learners' perspectives on using Kahoot as a game-based student response system. *Interactive Learning Environments*, *31*(4), 2338-2350.

Elahi Shirvan, M., & Talebzadeh, N. (2018). Exploring the fluctuations of foreign language enjoyment in conversation: an idiodynamic perspective. *Journal of Intercultural Communication Research*, *47*(1), 21-37.

Ellis, R. (2003). Designing a task-based syllabus. *RELC journal*, *34*(1), 64-81.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International journal of applied linguistics*, *19*(3), 221-246.

Ellis, R. (2018). *Reflections on task-based language teaching* (Vol. 125). Multilingual Matters.

Eslami, Z. R., & and 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. https://doi.org/10.1080/09571736.2016.1227219

Fanselow, M. S. (1989). The adaptive function of conditioned defensive behavior: An ecological approach to Pavlovian stimulus-substitution theory.

Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, *1*(4), 429-438. https://doi.org/10.3758/BF03210947

Fanselow, M. S., Lester, L. S., & Helmstetter, F. J. (1988). Changes in feeding and foraging patterns as an antipredator defensive strategy: a laboratory simulation using aversive stimulation in a closed economy. *Journal of the experimental analysis of behavior*, *50*(3), 361-374.

Ganta, T. G. (2015). The strengths and weaknesses of task-based learning (TBL) approach. *Scholarly research journal for interdisciplinary studies*, *3*(16), 2760-2771.

Gattegno, C. (1977). English the Silent Way (videotape series). *New York Silent Way Video Company*.

Genesee, F. (1987). Learning through two languages: Studies of immersion and bilingual education. *(No Title)*.

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.

Habermas, J. (1970). Towards a theory of communicative competence. *Inquiry*, *13*(1-4), 360-375.

Harley, B., & Swain, M. (1984). The interlanguage of immersion students and its implications for second language teaching. *Interlanguage*, *291*(1), 311.

Harmer, J. (2007). The practice of English language teaching (Fourth Edi). *UK: Pearson Education Ltd*.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. Routledge.

Horwitz, E. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, *21*, 112-126.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125-132.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, *30*(4), 461-473.

Hung, H.-T. (2018). Gamifying the flipped classroom using game-based learning materials. *ELT Journal*, *72*(3), 296-308.

Hwang, G.-J., Hsu, T.-C., Lai, C.-L., & Hsueh, C.-J. (2017). Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers & education*, *106*, 26-42.

Hymes, D. H. (1971). *Pidginization and creolization of languages*. CUP Archive.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language learning*, *63*, 87-106.

Jarvis, S., & Daller, M. (2013). *Vocabulary knowledge: Human ratings and automated measures* (Vol. 47). John Benjamins Publishing.

Jin, Y., & Zhang, L. J. (2021). The dimensions of foreign language classroom enjoyment and their effect on foreign language achievement. *International Journal of Bilingual Education and Bilingualism*.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.

Kormos, J. (2011). Chapter 2. Speech production and the cognition hypothesis. In *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 39-60). John Benjamins Publishing Company.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of second language writing*, *17*(1), 48-60.

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior*, *71*, 508-515.

Lee, Y., & Davis, M. (1997). Role of the hippocampus, the bed nucleus of the stria terminalis, and the amygdala in the excitatory effect of corticotropin-releasing hormone on the acoustic startle reflex. *Journal of Neuroscience*, *17*(16), 6434-6446.

Levkina, M., Gilabert, R., Baralt, M., Gilabert, R., & Robinson, P. (2014). Task sequencing in the L2 development of spatial expressions. *Task sequencing and instructed second language learning*, 37-70.

Ling, X., Wang, H., & Wang, Z.-h. (2019). A study on the effectiveness of mobile game-based learning for vocabulary acquisition among college learners. *Technol. Enhanc. Foreign Lang. Educ*, *6*, 9-15.

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *International review of research in open and distributed learning*, *14*(3), 202-227.

Long, M. H. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics*, *36*, 5-33.

Long, M. H., Hyltenstam, K., & Pienemann, M. (1985). Modelling and assessing second language acquisition. *Clevedon Avon: Multilingual Matters Ltd*.

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere*, *24*(1), 34-70.

MacIntyre, P. D. (2016). *So far so good: An overview of positive psychology and its contributions to SLA*. Springer.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language learning*, *39*(2), 251-275.

Malicka, A. (2020). The role of task sequencing in fluency, accuracy, and complexity: Investigating the SSARC model of pedagogic task sequencing. *Language Teaching Research*, *24*(5), 642-665.

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. *British Studies in Applied Linguistics*, *12*, 58-71.

Michel, M. (2011). Effects of task complexity and interaction on L2 performance. InP. Robinson.(Ed.), Second language task complexity: Researching the Cognition Hypothesis of language learning and performance (pp. 141–174). In: John Benjamins.

Mierzwa, E. (2018). The relationship between foreign language enjoyment and gender among secondary grammar school students. *The Journal of Education, Culture, and Society*, *9*(2), 117-135.

Natsir, M., & Sanjaya, D. (2014). Grammar translation method (GTM) versus communicative language teaching (CLT); A review of literature. *International Journal of Education and Literacy Studies*, *2*(1), 58-62.

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.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning*, *50*(3), 417-528.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge university press.

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.

Nunan, D. (2004). *Task-based language teaching*. Cambridge university press.

Ortega, L., & González-Lloret, M. (2014). Technology-mediated TBLT.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In *Emotion in education* (pp. 13-36). Elsevier.

Perusini, J. N., & Fanselow, M. S. (2015). Neurobehavioral perspectives on the distinction between fear and anxiety. *Learning & Memory*, *22*(9), 417-425.

Peterson, M. (2012). Towards a research agenda for the use of three-dimensional virtual worlds in language learning. *Calico Journal*, *29*(1), 67-80.

Pinel, J. P., Mumby, D. G., Dastur, F. N., & Pinel, J. G. (1994). Rat (Rattus norvegicus) defensive behavior in total darkness: risk-assessment function of defensive burying. *Journal of comparative psychology*, *108*(2), 140.

Prabhu, N. S. (1987). *Second language pedagogy* (Vol. 20). Oxford university press Oxford.

Resnik, P., & Dewaele, J.-M. (2020). Trait emotional intelligence, positive and negative emotions in first and foreign language classes: A mixed-methods approach. *System*, *94*, 102324.

Reynolds, E. D., & Taylor, B. (2020). Kahoot! in Korean university EFL: Is it effective for vocabulary learning? *Computer-Assisted Language Learning Electronic Journal*, *21*(2), 70-92.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge university press.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge.

Robinson, P. (2001a). *Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA*. na.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, *22*(1), 27-57.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design.

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243-268). John Benjamins Publishing Company.

Robinson, P. (2011). Task-based language learning: A review of issues. *Language learning*, *61*, 1-36.

Robinson, P. (2022). The cognition hypothesis, the triadic componential framework and the SSARC model: An instructional design theory of pedagogic task sequencing. *The Cambridge handbook of task-based language teaching*, 205-225.

S. Fanselow, M. (1984). What is conditioned fear? *Trends in Neurosciences*, *7*(12), 460-462.

Šafranj, J., & Zivlak, J. (2019). Effects of big five personality traits and fear of negative evaluation on foreign language anxiety. *Croatian Journal Educational/Hrvatski Casopis za Odgoj I Obrazovanje*, *21*(1).

Samuda, V., Bygate, M., Samuda, V., & Bygate, M. (2008). Task research from a pedagogical perspective. *Tasks in Second Language Learning*, 133-191.

Sasayama, S. (2016). Is a ‘complex’task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal*, *100*(1), 231-254.

Sasayama, S., & Izumi, S. (2012). Chapter 2. Effects of task complexity and pre-task planning on Japanese EFL learners’ oral production. In *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 23-42). John Benjamins Publishing Company.

Savignon, S. J. (2005). Communicative language teaching: Strategies and goals. In *Handbook of research in second language teaching and learning* (pp. 635-651). Routledge.

Setiawan, M. R., & Wiedarti, P. (2020). The effectiveness of Quizlet application towards students' motivation in learning vocabulary. *Studies in English Language and Education*, *7*(1), 83-95.

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, *36*(3), 517-554.

Skehan, P. (1989). Language testing part II. *Language Teaching*, *22*(1), 1-13.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38-62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Smith, B., & González-Lloret, M. (2021). Technology-mediated task-based language teaching: A research agenda. *Language Teaching*, *54*(4), 518-534.

Spada, N. (2007). Communicative language teaching: Current status and future prospects. *International handbook of English language teaching*, 271-288.

Spada, N., & Lightbown, P. M. (1989). Intensive ESL programmes in Quebec primary schools. *TESL Canada Journal*, 11-32.

Spielmann, G., & Radnofsky, M. L. (2001). Learning language under tension: New directions from a qualitative study. *The Modern Language Journal*, *85*(2), 259-278.

Su, F., Zou, D., Xie, H., & Wang, F. L. (2021). A comparative review of mobile and non-mobile games for language learning. *Sage Open*, *11*(4), 21582440211067247.

Sumarsono, D., Muliani, M., & Bagis, A. K. (2020). The forecasting power of task-based language teaching and self-efficacy on students' speaking performance. *Journal of Languages and Language Teaching*, *8*(4), 412-421.

Sunarwan, A. (2016). INCREASING STUDENTS'NARRATIVE PARAGRAPH WRITING ABILITY THROUGH THE USE OF PICTURE SEQUENCE. *Pedagogy: Journal of English Language Teaching*, *4*(1), 29-39.

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*(2), 363-387.

Thornbury, S. (2018). Learning grammar. *The Cambridge guide to learning a second language*, 183-192.

Timberlake, W., & Lucas, G. A. (2019). Behavior systems and learning: From misbehavior to general principles. *Contemporary learning theories*, 237-276.

Torres, J. R. (2013). *Heritage and second language learners of Spanish: The roles of task complexity and inhibitory control*. Georgetown University.

Tzoannopoulou, M. (2016). Foreign language anxiety and fear of negative evaluation in the Greek university classroom. *Selected papers on theoretical and applied linguistics*, *21*, 823-838.

Van den Branden, K. (2006). Task-based language education: From theory to practice. *Language*, *53*(82), T35.

Van den Branden, K., Bygate, M., & Norris, J. (2009). Task-based language teaching: A reader.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language testing*, *17*(1), 65-83.

Walker, D. L., & Davis, M. (1997). Double dissociation between the involvement of the bed nucleus of the stria terminalis and the central nucleus of the amygdala in startle increases produced by conditioned versus unconditioned fear. *Journal of Neuroscience*, *17*(23), 9375-9383.

Wardhani, A. (2019). Exploring fear of negative Evaluation in foreign language anxiety: Photovoice of undergraduates in speaking class. *JSSH (Jurnal Sains Sosial dan Humaniora)*, *3*(1), 1-10.

Werbach, K. (2014). (Re) defining gamification: A process approach. International conference on persuasive technology,

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.

Xue, S. (2022). A conceptual model for integrating affordances of mobile technologies into task-based language teaching. *Interactive Learning Environments*, *30*(6), 1131-

1144.

Yan, J. X., & Liang, J. (2022). Foreign language anxiety and dependency distance in English–Chinese interpretation classrooms. *Frontiers in Psychology*, *13*, 952664.

Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, *13*, 1030790.

Zohud, N. W. (2019). Exploring Palestinian and Spanish teachers' perspectives on using online computer games in learning English vocabulary.

Zou, D. (2020). Gamified flipped EFL classroom for primary education: Student and teacher perceptions. *Journal of Computers in Education*, *7*(2), 213-228.

APPENDIX

Appendix 1

The FLF Scale

The Foreign Language Fear scales

To what extent do you agree with the following statements?

[Strongly disagree; Disagree; Undecided; Agree; Strongly agree]

1. I feel uncomfortable BEFORE doing the task.

2. I feel uncomfortable WHILE doing the task.

3. I feel uncomfortable AFTER doing the task.

4. I want to escape from doing the task WHILE doing it.

5. I feel relaxed when I finish the task.

6. Please briefly tell us how you felt **during** the task.

____

Appendix 2

The FLE Scale

The Foreign Language Enjoyment Scales

To what extent do you agree with the following statements?

[Strongly disagree; Disagree; Undecided; Agree; Strongly agree]

1. I enjoy the task

2. It’s fun to do the task

3. I can be creative with the task

4. I don’t get bored with the task

5. I’ve learnt interesting things from the task

6. Please briefly tell us how you felt with the task.

____

Appendix 3

The Form of Information for Simple, Middle and Complex Tasks

Simple Task

Dimensions	Contents
Who	The thief was a middle-aged woman with dark skin, brown or black hair. She was dressed in a pink sleeveless top and black over-knee pants, wearing a pink face mask, and carrying a cream-colored canvas bag.
Where	The theft occurred in an unattended appliance and general goods store. The store displayed various electronic items, including electric fans, air conditioners, refrigerators, and microwaves.
When	The incident took place during the daytime.
What	The woman stole a new television, still in its original packaging, and dragged it out of the store. Although several pedestrians passed by outside during the incident, none of them noticed the theft.

Middle Task

Dimensions	Contents
Who	The thief was a tall and slim young-to-middle-aged man with a lean and elongated appearance. He was dressed in a black-and-orange padded jacket or thick coat, black trousers, and a knitted hat.
Where	The incident took place at a motorcycle and bicycle shop.
When	The theft occurred late at night.
What	The man arrived in a construction vehicle, using its mechanical arm to smash the shop's display window. He then exited the vehicle, tied two green bicycles to the mechanical arm with a rope, and drove away with the stolen items. During the incident, the street was mostly deserted, with few or no people or vehicles in sight.

Complex Task

Dimensions	Contents
Who	Motorcycle Driver: A person wearing a black helmet, green jacket, and black pants. After the collision, they got up and immediately ran away. Individuals from the White Car: Three young men wearing T-shirts, shorts, and baseball caps in various colors.
Where	The incident occurred on a road with greenery on both sides, situated in a residential area.
When	The event took place during the daytime.
What	A white sedan collided with a motorcycle, causing it to fall. The motorcycle driver fled the scene, leaving behind belongings such as a phone and wallet. Shortly after, the three men from the white car collected the valuables left on the ground and pushed the fallen motorcycle away.

Appendix 4

Glossary of Abbreviations

Abbreviation	Full Term / Definition

ALM	Audio-Lingual Method
APA	American Psychological Association
AS	Analysis of Speech unit (used in syntactic complexity measures)
CA	Communication Apprehension
CAF	Complexity, Accuracy, Fluency
CALF	Complexity, Accuracy, Lexical Complexity, Fluency
CALL	Computer-Assisted Language Learning
CH	Cognition Hypothesis
CLL	Community Language Learning
CLT	Communicative Language Teaching
CMC	Computer-Mediated Communication
CVT	Control-Value Theory
EFCR	Error-Free Clause Ratio
EFL	English as a Foreign Language
ESL	English as a Second Language
FLA	Foreign Language Anxiety
FLCA	Foreign Language Classroom Anxiety
FLCAS	Foreign Language Classroom Anxiety Scale
FLE	Foreign Language Enjoyment
FLF	Foreign Language Fear
FNE	Fear of Negative Evaluation
GTM	Grammar-Translation Method
ICALL	Intelligent Computer-Assisted Language Learning
IELTS	International English Language Testing System
ISLA	Instructed Second Language Acquisition
L2	Second Language
MLAS	Mean Length of AS-unit
MLC	Mean Length of Clause
MOOC	Massive Open Online Course
MGBLL	Mobile Game-Based Language Learning
NMGBLL	Non-Mobile Game-Based Language Learning
PE	Positive Education
PP	Positive Psychology
SLA	Second Language Acquisition
SPS	Syllables Per Second
SSARC	Simplify, Stabilize / Automatize / Restructure, Complexify
SW	Silent Way
TA	Test Anxiety

TBLT	Task-Based Language Teaching
TMTBLT	Technology-Mediated Task-Based Language Teaching
TNW	Total Number of Words
TNS	Total Number of Syllables
TPR	Total Physical Response
TTR	Type-Token Ratio
WPM	Words Per Minute

VITA

NAME	ZHANG WANYI
DATE OF BIRTH	7 August 1995
PLACE OF BIRTH	YUNNAN,CHINA
HOME ADDRESS	taoli jiayuan, yihua Road, KUNMING, YUNNAN, CHINA