EXPLORING THE ROLE OF TASK MODALITY AND TASK COMPLEXITY IN L2 PERFORMANCE IN EFL CLASSES

TUĞÇE BETÜL AKTAŞ SMITH

BOĞAZİÇİ UNIVERSITY

2023

EXPLORING THE ROLE OF TASK MODALITY AND TASK COMPLEXITY

IN L2 PERFORMANCE IN EFL CLASSES

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

English Language Education

by

Tuğçe Betül Aktaş Smith

Boğaziçi University

2023

Exploring the Role of Task Modality and Task Complexity

in L2 Performance in EFL Classes

The thesis of Tuğçe Betül Aktaş Smith

has been approved by:

Assist. Prof. Şebnem Yalçın ____________________________
(Thesis Advisor)

Assoc. Prof. Senem Yıldız ____________________________

Assoc. Prof. Hale Işık Güler ____________________________
(External Member)

May 2023

DECLARATION OF ORIGINALITY

I, Tuğçe Betül Aktaş Smith, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature...

Date ....

ABSTRACT

Exploring the Role of Task Modality and Task Complexity

in L2 Performance in EFL Classes

Task-based language learning and teaching (TBLT) research has shown that task design can be cognitively manipulated to impact second language (L2) performance and create unique opportunities for exploiting different aspects of L2 performance. There are two influential hypotheses on this matter: the Cognition Hypothesis (Robinson, 2001a, 2011a) and the Limited Attentional Capacity Hypothesis (Skehan, 1998, 2009). While the former account predicts a concurrent increase in L2 complexity and accuracy during complex tasks, the latter model suggests a competition over attentional resources leading to the prioritization of certain aspects of L2 performance. Both accounts offer a valuable contribution to cognitively driven task-based research. However, it remains an open empirical question whether these predictions can be applied to written performance and how L2 performance is affected by the interplay of task complexity and modality (speaking versus writing). Therefore, this study examines separate and combined effects of task complexity and task modality on learners' syntactic and lexical complexity, and accuracy. The results revealed significant differences between writing and speaking in the aspects of phrasal complexity and lexical complexity. Further, a significant interaction was found between modality and task complexity for unit accuracy. The findings indicate that the effects of task complexity differ in writing and speaking, providing insights into the discussion of whether current task complexity models can account for both writing and speaking. Further, combining task complexity and modality can offer learners excellent opportunities for exploiting different aspects of L2 performance.

ÖZET

İngilizce'nin Yabancı Dil Olarak Öğretiminde Görev Karmaşıklığının ve Görev Modlarının İkinci Dil Performansındaki Etkisinin İncelenmesi

Göreve dayalı öğretim araştırmaları, farklı bilişsel zorluktaki görevlerin ikinci dil performansını etkilediğini göstermiş ve bu etkinin dil performansının farklı yönlerine avantaj sağladığını ortaya koymuştur. Biliş Hipotezi (Robinson, 2001a, 2011a) ve Sınırlı Dikkat Kapasitesi Modeli (Skehan, 1998, 2009), bu alandaki iki önemli modeldir. İlk model, bilişsel zorluğu artan görevlerde dil karmaşıklığının ve doğruluğunun eş zamanlı olarak artacağını savunurken, Sınırlı Dikkat Kapasitesi Modeli sınırlı işlem kapasitesi nedeniyle karmaşık görevlerin dil performansındaki alanlarda önceliklendirmeye neden olacağını savunmaktadır. İki model de bilişsel göreve dayalı araştırmalara değerli katkılar sunmaktadır. Ancak, bu varsayımların yazılı performansa uygulanabilirliği ve bilişsel görev zorluğunun görev modlarıyla (yazılı ve sözlü) etkileşiminin ikinci dil performansına etkisi açık bir ampirik soru olmaya devam etmektedir. Bu nedenle, bu çalışma, görev karmaşıklığının ve modların öğrencilerin sözcük doğruluğu, sözdizimsel ve sözcük karmaşıklığı üzerindeki ayrı ve birleşik etkilerini incelemektedir. Sonuçlar, yazma ve konuşma modlarının öbek ve sözcük karmaşıklığını farklı şekilde etkilediğini göstermiştir. Ayrıca, sözcük doğruluğunda görev karmaşıklığı ve modlar arasında önemli bir etkileşim bulunmuştur. Bulgular, görev karmaşıklığının etkilerinin yazma ve konuşmada farklılık gösterdiğini ortaya koyarak, mevcut görev karmaşıklığı modellerinin farklı modlara uygulanabilirliğine dair tartışmalara ampirik bir ışık tutmaktadır. Ayrıca, görev karmaşıklığı ve modları birleştirmek öğrencilere ikinci dil performansının farklı yönlerinden yararlanma konusunda fırsatlar sunabilmektedir.

ACKNOWLEDGEMENTS

Finishing this thesis was a wonderful and challenging journey. I was very lucky to have wonderful people who supported me and made this journey less challenging. First and foremost, I would like to thank my supervisor Dr. Şebnem Yalçın. I am grateful for her help and generosity. I feel honored and privileged to receive her support and benefit from her knowledge.

I would also like to acknowledge the 2210-A scholarship that I received from the Scientific and Technological Research Council of Turkey. I am grateful for the financial support.

Finally, I cannot express how lucky I am to have such an amazing family and friends who supported me every step of this journey. I would like to thank my husband Hunter Smith for providing editorial and emotional support. He made this journey so much less challenging, especially emotionally. I am grateful for my mother-in-law Donna's help with the transcripts and her kindness. My math wizard sister Arya offered her sincere help with the statistical analysis. I would also like to thank Özge who is a great teacher and a friend for her support. I am truly grateful for all their time and effort.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

1.1 Background for the study 1

1.2 Goal of this study 3

1.3 Chapters and organization 6

CHAPTER 2: LITERATURE REVIEW 7

2.1 The importance of output in SLA 7

2.2 Speech models of first and second language 8

2.3 Models of writing 11

2.4 Modality in the comparison of language production models 14

2.5 L2 production measures 16

2.6 Tasks in task-based language teaching 20

2.7 Skehan’s cognitive understanding of tasks 25

2.8 Triadic Componential Framework and the Cognition Hypothesis 31

2.9 Modality in the revision of Skehan’s and Robinson’s models 37

2.10 Studies of language performance in oral and written modes 42

2.11 Task complexity studies in oral and written mode 50

2.12 Evaluation of task complexity studies in oral and written modes 61

CHAPTER 3: METHODOLOGY 67

3.1 Introduction 67

3.2 Participants 67

3.3 Research design 68

3.4 Materials 68

3.5 Procedure 71

3.6 Operationalization of task complexity 73

3.7 Measures 75

3.8 Data analysis 82

CHAPTER 4: RESULTS 87

4.1 Introduction 87

4.2 Syntactic complexity 87

4.3 Accuracy 96

4.4 Lexical complexity 104

4.5 Summary of the chapter 107

CHAPTER 5: DISCUSSION 110

5.1 Introduction 110

5.2 Research question 1: The effects of task modality 110

5.3 Research question 2: The effects of task complexity 116

5.4 Research question 3: Task complexity and modality interaction 124

CHAPTER 6: CONCLUSION 127

6.1 Summary and conclusions 127

6.2 Limitations and suggestions for future research 132

APPENDIX A: PICTURES USED IN THE STUDY 135

APPENDIX B: TASK INSTRUCTIONS (ENGLISH) 139

APPENDIX C: EXAMPLE AS-UNIT CODING 141

APPENDIX D: AS-UNIT CODING GUIDELINES 143

APPENDIX E: T-UNIT CODING GUIDELINES 148

APPENDIX F: CALCULATIONS OF MANUAL MEASURES 151

APPENDIX G: INTERCODER RELIABILITY SCORES 152

APPENDIX H: ETHICS COMMITTEE APPROVAL 153

REFERENCES.. 154

LIST OF TABLES

Table 1. Framework for Task-based Instruction 28

Table 2. Studies Investigating Language Performance in Two Modes 49

Table 3. Studies of Task Complexity Effects in Oral and Written Modes 57

Table 4. The Design of the Study 68

Table 5. Timeline of the Study... 72

Table 6. Measures in the Study 80

Table 7. Descriptive Statistics for Syntactic Complexity Measures 87

Table 8. 2x2 Repeated Measures ANOVA Results for Overall Complexity............ 89

Table 9. 2x2 Repeated Measures ANOVA Results for Subordination Complexity . 91

Table 10. Subordination Complexity Pairwise Comparisons of Task Complexity .. 92

Table 11. 2x2 Repeated Measures ANOVA Results for Phrasal Complexity.......... 94

Table 12. Phrasal Complexity Pairwise Comparisons of Modality 94

Table 13. Phrasal Complexity Pairwise Comparisons of Task Complexity 95

Table 14. Descriptive Statistics for Accuracy Measures 96

Table 15. 2x2 Repeated Measures ANOVA Results for Error-free Units................ 98

Table 16. Unit Accuracy Pairwise Comparisons of Modality 99

Table 17. Mauchly’s Test of Sphericity for Unit Accuracy................................... 100

Table 18. Tests of Within-Subjects Effects for Unit Accuracy 100

Table 19. Repeated Measures Pairwise Comparisons for Unit Accuracy 101

Table 20. 2x2 Repeated Measures ANOVA Results for Error-free Clauses 103

Table 21. Descriptive Statistics for Lexical Complexity 104

Table 22. 2x2 Repeated Measures ANOVA Results for Lexical Complexity........ 106

Table 23. Pairwise Comparisons of Modality for Lexical Complexity 106

Table 24. Summary of the Results 109

LIST OF FIGURES

Figure 1. Theorizing dimensions of performance (Skehan & Foster, 2001, p. 190) 31
Figure 2. The triadic componential framework (Robinson, 2011a, p. 7).................. 34
Figure 3. Overall complexity scores with %95 confidence intervals........................ 88
Figure 4. Subordination complexity scores with 95% confidence intervals............. 90
Figure 5. Modality and task complexity interaction for subordination complexity.. 92
Figure 6. Phrasal complexity scores with 95% confidence intervals........................ 93
Figure 7. Modality and task complexity interaction for phrasal complexity 96
Figure 8. Error-free units scores with 95% confidence intervals.............................. 97
Figure 9. Modality and task complexity interaction for error-free unit accuracy..... 99
Figure 10. Error-free clauses with 95% confidence intervals 102
Figure 11. Lexical complexity scores with 95% confidence intervals 105
Figure 12. Modality and task complexity interaction for lexical complexity 107

CHAPTER 1

INTRODUCTION

1.1 Background for the study

Task-based language learning and teaching (TBLT) has taken far more attention in the field of language learning and teaching from second language acquisition (L2) researchers than earlier methodologies. The implementations of tasks, their designs, and how to sequence them are much discussed in the field (Nunan, 2004). Much of this research has taken a psycholinguistic perspective. This perspective includes the investigation of the use of pedagogic tasks and takes sequencing of tasks based on their complexity “as the basis of syllabus design” and underlines that “such a sequencing decision should effectively facilitate L2 development: the acquisition of new L2 knowledge, and restructuring of existing L2 presentations” (Robinson, 2001b, p. 34). The basis of this argument can be traced to the theory that tasks with different complexity levels are predicted to affect learners’ L2 performances since cognitive processing imposed by tasks has an impact on the use of attentional resources leading to varieties in areas of language production (Robinson, 2001a, 2011a, 2011b; Skehan, 1996, 1998). According to Housen and Kuiken (2009), studies investigating these predictions have researched and analyzed areas of L2 performance through three main constructs: complexity, accuracy, and fluency (CAF). A multi-componential perspective is offered rather than a unitary one for L2 performance.

L2 complexity, briefly, refers to how much learners elaborate on their L2 system and elements in the interlanguage (IL) system (Housen & Kuiken, 2009). Opportunities for stretching existing interlanguage system and the acquisition of new

L2 components can greatly benefit from the elaboration on language production. Skehan (1996) argues that accuracy is mainly related to handling interlanguage complexity and the changes in this system. Finally, fluency is mainly concerned with how language is produced in real-time in reference to pausing or hesitation (Ellis & Barkhuizen, 2005).

Two hypotheses have been influential in analyzing and predicting aforementioned L2 performances in relation to task demands: the Limited Attentional Capacity (Skehan, 2009) and the Cognition Hypothesis (Robinson, 2001b). The main difference between these two hypotheses comes from the way they view learners' attentional resources. Skehan (2009, 2014) argues that learners have limited attentional capacity, resulting in competition over attentional resources during demanding tasks. Consequently, when learners encounter cognitively demanding tasks, they prioritize meaning over form and fluency over accuracy and complexity (Skehan, 2009, 2014). This is called Trade-off Hypothesis as learners' limited attentional capacity might result in choosing between accuracy and complexity (Skehan, 2014).

Unlike Limited Attentional Capacity, Robinson (2001b, 2011a) adopts a multi-resource perspective towards attentional capacity and argues that learners can expand their attention to meet the demands of a task. In fact, according to Robinson (2011b), more cognitively complex tasks can lead to more complex and accurate language as they push learners' attentional resources. Consequently, higher accuracy and syntactic complexity are possible in more complex tasks, but fluency is expected to decrease.

Studies concerning the relationship between task demands and L2 performances have offered valuable insights and discussions into the development of

TBLT research, task design, and sequence. However, there seems to be a need to address task modality in task complexity research. Speaking has been widely researched in relation to task complexity along with some studies focusing on writing, but the relationship between task complexity and modality has been understudied and needs to be addressed and investigated further (Gilabert, Manchón & Vasylets, 2016; Kormos & Trebits, 2012; Kormos, 2014; Kuiken & Vedder, 2011; Kuiken & Vedder, 2012; Tavakoli, 2014; Vasylets, Gilabert & Manchón, 2017; Zalbidea, 2017). Kormos (2014) argues that modality needs to be addressed and further considered as an element of task complexity since different cognitive processes are involved in writing and speaking. Some studies revealed that written and oral language production can lead to different outcomes in certain areas of L2 language production (Bulté & Housen, 2009, as cited in Vasylets, Gilabert & Manchón, 2019; Ellis, 1987; Ellis & Yuan, 2005; Ferrari & Nuzzo, 2009, as cited in Vasylets et al., 2019; Gilabert et al., 2016; Granfeldt, 2008; Kormos, 2014; Vasylets et al., 2017). Considering the importance of what each mode can offer to language learning, mode-sensitive research integrating task complexity is necessary to shed further light on how manipulation of task demands affects L2 performances in writing speaking.

1.2 Goal of this study

The current study aims to investigate how language performance is affected by task complexity and modality. The cognitive demands of a task in the conceptualization stage can be considered a part of task complexity (Robinson, 2001b; Skehan, 1998). When the tasks chosen for this study are considered, the conceptualization stage is manipulated to impose different levels of task complexity. The tasks that are defined as simple in this study include telling a story based on related pictures. These tasks are considered to put less demand on the conceptualization of the story as the pictures in the tasks are presented in the order of the story. The tasks that are considered complex include telling a story based on an unrelated picture sequence. These tasks require participants to design a story plot by interweaving the pictures, conceptualizing the stages in the story on their own, and finding the linguistic resources available to them for articulation. This process results in higher demands on the conceptualization stage (Kormos & Trebits, 2012; Tavakoli, 2014). Although these tasks have been used before for task complexity and modality studies, task modality studies constitute a very small amount of task complexity research and it needs further research to be able to make stronger claims and increase the reliability of the results (Gilabert et al., 2016; Kormos & Trebits, 2012; Kormos, 2014; Kuiken & Vedder, 2011; Kuiken & Vedder, 2012; Tavakoli, 2014; Vasylets et al., 2017; Zalbidea, 2017). Therefore, this study's theoretical and empirical contribution is to respond to the need for mode-sensitive research in TBLT. By doing so, it investigates how learners' accuracy and syntactic and lexical complexity vary in written and oral modes in tasks with different cognitive demands. This study conducts an interactional analysis of task complexity and modality to have a more in-depth analysis of potential differences between the independent variables. Interaction analysis is missing in other studies, except for one (Vasylets et al., 2017). As a methodological contribution, this study offers results from an understudied group in the field of mode-integrative research: younger learners (ages between 11-13). Another methodological contribution comes from the research design. This study implemented a within-subject research design where the same individuals were measured over time. This was considered a necessary design as the study investigates the role of cognitive demands and the mode of language production on language performance. Other studies implemented a between-subject design assigning participants to different conditions such as writing and speaking (Kuiken & Vedder, 2011; Tavakoli, 2014; Vasylets et al., 2017; Zalbidea, 2017). Comparison of different individuals for task modality can constitute a limitation for the interpretations of the results. For this very reason, Tavakoli (2014) refrained from making any statistical analyses between the writing and speaking groups. Other potential methodological problems stood out from previous studies that the current study attempted to fix. Counterbalancing of modality was absent in Kormos and Trebits (2012) which is the only study employing a within-subject design. Data from different years and/or different participants were used in the comparison of oral and written data in Kuiken and Vedder (2011), Tavakoli (2014), Vasylets et al. (2017), and Zalbidea (2017). To address these issues, counterbalancing for both modality and tasks was implemented in this study and data were collected from the same individuals over two months in their school.

This study provides pedagogical and practical implications as it attempts to uncover how different modes affect learners' language performance and interact with task complexity. The results will be discussed in terms of their implications for syllabus design and implementations of tasks in language classrooms to further foster interlanguage development. Thus, the following research questions are addressed:

i. How does task modality affect L2 complexity and accuracy?

ii. How does task complexity affect L2 complexity and accuracy?

iii. How do the effects of task complexity and task modality on L2 complexity and accuracy interact with each other?

1.3 Chapters and organization

This study consists of 6 chapters. It starts with an introductory chapter where it gives information about the study and states the aims. Chapter 2 starts with the importance of output in SLA and the theoretical background regarding the issue. It later covers oral and written language production models, a comparison of oral and written production models, L2 production measures, task complexity, and the previous studies about task complexity in relation to task modality. Chapter 3 discusses the methodology, materials used in the study, operationalization of terminologies, and data analysis. Chapter 4 reports the results of accuracy, syntactic, and lexical complexity measures in relation to task complexity, modality, and the interaction of the two. Chapter 5 discusses the results from the previous chapter considering relevant task complexity hypotheses, oral and written language production models, and previous studies. The last chapter, Chapter 6, offers a summary of the study and discusses its limitations.

CHAPTER 2

LITERATURE REVIEW

2.1 The importance of output in SLA

Input hypothesis (Krashen, 1981, 1982) as a dominant theory at its time argued that L2 acquisition is triggered by comprehensible input that is beyond learners' interlanguage, in other words, *i* + 1. The input hypothesis assumed that the comprehension of language input that is slightly beyond learners' current interlanguage system would require them to make use of their general knowledge and the context along with their linguistic knowledge. At that time of SLA research, the focus was to study comprehensible input and its implementations while output was evaluated as the mere product of the acquisition which relied on the comprehensible input (Swain, 1993).

In this historical environment, Swain (1985) started to argue that language production should be evaluated as part of the learning process. Swain (1985) elaborated on Krashen's (1982) argument that learners might not use or process syntactic information in the comprehensible input, instead, they might rely on lexical or extra-linguistic information to process the input and claimed that language output might be the answer to this problem. Output can lead learners to shift from semantic processing to syntactic processing (Swain, 1985). She explained that engaging in output can result in testing out hypotheses about the means of expression and might help learners shift their attention from semantic processing to syntactic processing.

Swain's (1985) arguments and propositions relied on French immersion programs in Canada. Her observational study with Grade 6 French students revealed that students' level of writing and speaking was behind the level of proficiency that was expected although they had been exposed to comprehensible input for 7 years (Swain, 1985). Therefore, she offered the output hypothesis as the potential answer to her findings from immersion programs and concluded that learners did not have enough opportunities to utilize the language and to be engaged in accurate and precise output (Swain, 1985).

Swain and Lapkin (1995) elaborated on the cognitive processes that output generates through analysis of learners' dialogues. In their study, they analyzed language-related episodes (LREs). LREs were originally referred to as dialogues where learners comment on, question, or correct the language they are producing (Swain & Lapkin, 1995). They argued in their study that language production is not merely a product but a trigger for L2 development as it is part of the L2 acquisition process. For this very reason, it is important to explore the cognitive processes and mechanisms underlying language production and how different modes of language production affect L2 performance.

2.2 Speech models of first and second language

The dominant psycholinguistic model of speech in the field was offered by Levelt (1989) for first language speaking performance. Levelt's (1989) speech model consists of autonomous components which are in charge of different aspects of language production. These basic components are conceptualization, formulation, articulation, and monitoring. Conceptualization includes the stage of generating ideas, thinking about the content of speech, and macroplanning. A preverbal message is considered the output of this stage. To be able to generate a preverbal message, macro, and microplanning stages are used (de Bot, 1992). Macroplanning involves thinking about the communicative goals and being engaged in the retrieval of the information that can explain these goals. Microplanning involves selecting information that expresses the communicative intentions the best (Levelt, 1989). The formulation stage involves encoding the preverbal message formed in the previous stage. The formulation stage includes the selection of lexis and necessary syntactic information to form the linguistic output (Levelt, 1993). Finally, monitoring is concerned with checking the message in both the formulation and articulation stages. The comprehension system is utilized in this stage to check the language output. Levelt (1989) states that this process is parallel and automatic in L1 speech which means that while the Conceptualizer is transferring macroplanning to the Formulator, it starts being occupied with the next information. While this automatic process works for L1 speakers and L2 learners with high proficiency levels, the Formulator and the Articulator might ask for additional attentional resources for less proficient L2 learners because of their low L2 linguistic knowledge (Kormos, 2006). Deficiencies in speech processing such as lexical, grammatical, and phonological encoding can result in problems in speech production because of limited attentional resources and having to process the message within time constraints (Kormos, 2006). Consequently, the parallel process is disrupted and becomes serial for L2 speakers with lower proficiency levels (Skehan, 2014).

de Bot (1992) argued that "many aspects of speaking are the same for monolingual and bilingual speakers, and a single model to describe both types of speakers is to be preferred over two separate models" (p. 2). Similarly, Kormos (2006) argued for Levelt's (1989) three stages of speech production: conceptualization, formulation, articulation, and monitoring for bilingual speakers. Consequently, both de Bot (1992) and Kormos (2006) took Levelt's (1989) speech model as the basis of their arguments for the bilingual speech model but addressed a couple of issues. de Bot (1992) identified additional requirements for the bilingual version of the model. The model should address the two language systems and how they can be used separately or mixed based on the situation such as code-switching, and cross-linguistic influences because the assumption that the speaker uses two systems has an effect on the model's organization (de Bot, 1992). Another issue addressed in the Levelt's (1989) speech model was the role of using attention in two systems. Kormos (2011a) argued that, unlike L1 speakers, L2 speakers may need to divert their attentional resources when encountering a problem in other stages. Depending on L2 speakers' level of proficiency, automaticity in linguistic coding can be disrupted and additional attentional resources can be needed to repair the problem (Kormos, 2011a). A conscious search for the appropriate lemma or morphological coding might be required therefore parallel or automatic speech production can be exposed to disruption (Kormos, 2011a). Consequently, as proficiency increases, formulation and articulation stages become less disrupted and more automatized.

To summarize, L1 and L2 speech productions are very similar in the sense that they both include stages of conceptualization, formulation, articulation, and monitoring. However, the allocation of attentional resources to realize these stages may vary depending on the level of L2 acquisition. A speaker from the early stages of L2 acquisition needs to attend these stages consciously resulting in the disruption of the automaticity of speech production.

After reviewing the dominant monolingual and bilingual speech production models, the following section will focus on writing models.

2.3 Models of writing

Models of writing are influenced by Hayes and Flower’s (1980) protocol analysis. They conducted a study where they described the technique of protocol analysis and used writing processes to identify them (Hayes & Flower, 1980). The protocols they collected involved solving problems. They analyzed the sequences that participants used to solve the given problem such as drawing a diagram or making a computation. Verbal protocols included what participants said while performing the given task. They underlined the psychological processes that participants used to perform a task. As they put it “between surfacings, the mental process, like the porpoise, runs deep and silent. Our task is to infer the course of the process from these brief traces” (Hayes & Flower, 1980, p. 10). As a result, they proposed a writing model involving the task environment, the writer’s long-term memory, and the writing processes. The task environment refers to the factors that are outside the writer such as the task itself, the description of the task as well as the intended audience. The task environment also includes the text produced so far because the writer can utilize this text as a reference for future decisions (Hayes & Flower, 1980). Long-term memory involves the knowledge that the writer possesses about the topic, the audience, or the writing plans required for the task. The writing processes include the stages of planning, translating, and reviewing. In the planning stage, the writer goes through generating, organizing, and goal-setting subprocesses. The writer uses the planning stage to take information from the task environment and long-term memory so that goals and plans can be set. Translating is responsible for producing the language in accordance with the planning stage. Reviewing stage consists of reading and editing subprocesses to improve the quality of the written product.

Hayes (1996, 2012) proposed a new framework based on the 1980 model considering the research that has been done in writing. This model included two major components: the task environment and the individual. The task environment here includes a social element along with physical factors. The new model stresses how writing is a social activity considering what we write is shaped by our social past, history, and interaction. In the individual part, working memory was emphasized compared to the 1980 model because the new model argued that “all the processes have access to working memory” and was added to underline its central importance (Hayes, 1996, p. 8). Another important addition to the 1980 model was motivation (Hayes, 1996, 2012). Hayes (1996) claimed that motivation is closely related to setting goals in writing.

Subprocesses were undergone a major change in the proposal of the new model. Hayes (2012) added and defined the transcription stage in the new model. While in the previous model, transcription was not considered as a significant impact on other writing processes, the new model recognizes that transcription can affect cognitive sources and therefore other writing processes (Hayes, 2012). While the new model emphasizes transcription, it removes monitoring, planning, and revision stages from subprocesses. Although writers go through these stages, Hayes (2012) argues that there is no need to consider planning and revision as subprocesses of writing because these subprocesses are part of translating and transcription rather than being separate writing processes. Translating and transcribing necessarily involve revising and planning so they are more like parallel processes than separate (Hayes, 2012). The new model also removed monitoring from subprocesses. The reason for removing monitoring stemmed from empirical studies showing that monitoring can be modified by instruction. Therefore, Hayes (2012) argues that monitoring can be better understood as a task schema stored as declarative knowledge. Although an update on the 1980 model was needed after much research was conducted, considering planning and revision as part of translation and transcription may lead to overlooking specific roles and characteristics of these subprocesses.

Hayes' (1980) model became a basis for writing models including Kellogg's (1996) model, one of the prominent models in the field. This model has also been widely used as a framework in writing studies. Kellogg's (1996) writing model includes three main stages: formulation, execution, and monitoring. The formulation stage includes planning and translating. Kellogg (1996) defines planning as conceptualizing the message as well as organizing it. The translation includes encoding the message linguistically. Here, writers convert their plans for the written product into linguistic structures and go through operations such as lexical retrieval and syntactic encoding. This process can be seen in other models as translating. The execution stage is the action of writing and creating the text which includes graphomotor movements. Finally, monitoring includes reading the written product and editing it. Editing can lead to feedback if any mismatch is detected and to restarting parts of the written product. These processes in writing can be interactive and cyclical as well as serial (Kellog, 1996; Vasylets et al., 2017). Kellogg (1996) in his model adopted Baddeley's (1986) view on working memory to account for interactions between processes described above. Kellog (1996) claims that working memory plays a crucial role in each subprocess. Baddeley (1986) argued that working memory consists of a central executive that deals with complex processes together with the visuospatial sketchpad and the articulary loop. While the visuo-sketchpad is responsible for visual representations, the articulary loop deals with phonological representations. Kellogg (1996) argues that formulation is a burdening process that needs all the sub-components of working memory such as executive function, visual short-term memory, and phonological short-term memory. Monitoring is also demanding as it requires central executive and phonological short-term memory while execution needs only the central executive (Kellogg, 1996). In the model, Kellogg (1996) offers a theoretical relationship between working memory and writing as well as the limitation of cognitive resources during writing.

Having mentioned prominent models in both written and oral language production, it is necessary to compare them and offer an overview to address the issue of task modality in TBLT.

2.4 Modality in the comparison of language production models

Comparing oral and written production models are not only in researchers' interest in the field but also practitioners' because such a comparison offers theoretical ground on how these two modes can lead to different language outcomes and how they can potentially attract particular linguistic forms in the production. These differences might help gain perspectives and insights on learners' performances in oral and written modes. To compare the two modes, the psycholinguistic perspective of oral and written production will be addressed.

Ravid and Tolchinsky (2002) addressed the differences in online and off-line processing nature of these two modalities together with many researchers in the field (Gilabert et al., 2016; Kormos & Trebits, 2012; Kormos, 2014; Kuiken & Vedder, 2012; Tavakoli, 2014; Vasylets et al., 2017; Williams, 2012; Zalbidea, 2017). When speech and writing models are analyzed, there is a clear difference in terms of online and offline planning. Writing offers self-paced production thanks to offline planning opportunities which can lead to careful linguistic encoding (Gilabert et al., 2016). Writers are comparatively under less pressure as content planning is integrated into the actual writing (Kormos, 2014). Vasylets et al. (2017) further argue that the self-paced nature of writing provides opportunities for more careful planning and linguistic encoding as well as monitoring compared to speaking. According to William (2012), this can lead to more control over using attentional resources and more opportunities to attend language production both during and after production. Information in learners' memory has already been put in writing, therefore it can be revisited while learners must keep the information already produced in their memory in speaking (Kuiken & Vedder, 2012).

Another critical difference related to making use of cognitive resources comes from time constraints as writing inherently offers more time than speaking. Learners can benefit from the time available to them to be engaged in the task and can make use of their cognitive resources for a longer period (Kuiken & Vedder, 2012; Williams, 2012). Grabowski (2007) conducted a study regarding recall performance from working memory and long-term memory and revealed the superior effect of recall performance in written mode among adult learners. When the use of cognitive resources is considered, writing can favor formulation and monitoring and can increase the quality of production compared to speaking. Additionally, Kormos (2014) points out that the monitoring process benefits from the visibility of output in writing compared to speaking. Ravid and Tolchinsky (2002) have also recognized this:

> Control over linguistic output is potentially higher when writing, both because writing usually takes longer than talking, and because the stable and visually accessible written text permits writers to view the text as a whole, while the ephemeral nature of spoken language leaves a tight window for processing (p. 427).

The offline planning nature of writing allows language production to be self-determined which means that learners can stop the writing process and focus on retrieving information or planning (Grabowski, 2007). Learners can also focus more on the planning process by stopping the actual writing process if they need to be engaged in a new conceptual message (Vasylets et al., 2017). Granfeldt (2008) conducted a study to examine the effect of modality among L2 French learners and hypothesized that the written mode leads to higher accuracy and complexity than the oral mode because of self-determined planning and monitoring opportunities. However, no significant difference was found in grammar accuracy between writing and speaking (Granfeldt, 2008). More research is needed to confirm and offer a research-based interpretation of these differences. Yet, when all these different critical processes are considered between writing and speaking, Kormos (2014) argues that modality should be considered as a variable in task complexity based on the argument that the recursive and interactive nature of writing allows students to direct their attention to linguistic features and allocate more time on execution and monitoring stages.

After mentioning critical differences in writing and speaking, it is necessary to cover L2 measures of language production to understand how these differences have been measured and discussed in the literature.

2.5 L2 production measures

L2 performance and L2 proficiency are considered multi-componential and to be able to capture this nature of L2 production comprehensively, three main constructs have been offered by researchers: complexity, accuracy, and fluency (Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Skehan, 1998). These dimensions have been investigated as dependent variables to analyze independent variables such as task features or level of acquisition (Pallotti, 2009). They can be in form of ratios, frequencies, or formulas (Norris & Ortega, 2009).

Defining these constructs has been controversial in the field. Accuracy, however, can be considered the oldest, the most consistent, and the most transparent one since it mainly refers to the deviation from a particular norm (Housen & Kuiken, 2009). These deviations from the norm are considered errors. Accuracy involves a few measures. One way of measuring accuracy is to focus on the accuracy of language production as in the ratio of error-free clauses or the ratio of error-free units. Another way is to focus on the errors by calculating errors per 100 words or by calculating error gravity which does not treat all errors equally (Ellis, Skehan, Li, Shintani & Lambert, 2020). Ellis et al. (2020) state that the measure of error gravity gives errors different levels of weight since some of them pose more importance to communication than others. The ratio of error-free units and clauses has been used in the existing body of research greatly and offers reliable results (Ellis et al., 2020).

Fluency has been used to refer to “the ease, eloquence, ‘smoothness’ and native-likeness of speech or writing” (Housen, Kuiken & Vedder, 2012, p. 4). However, researchers in the field argue that fluency is multidimensional and there are subdimensions. Tavakoli and Skehan (2005) divided fluency into breakdown fluency, repair fluency, and speed. Breakdown fluency is indexed by pausing which includes the analysis of the number, length, and location of pauses (Skehan, 2009). Repair fluency involves the measures of false starts, reformulation as well as repetitions (Skehan, 2009). Finally, the speed of language production involves the measures of how many syllables per unit are produced (Housen et al., 2012).

Complexity is considered to be “the most complex, ambiguous, and least understood dimension of the CAF triad” (Housen & Kuiken, 2009, p. 463). The term complexity has been used for both task complexity referring to tasks’ language properties and for L2 complexity referring to language performance and proficiency (Housen & Kuiken, 2009). L2 complexity will be covered as it is relevant to this section.

Housen and Kuiken (2009) argue that L2 complexity has two subdimensions studied in the field: cognitive complexity and linguistic complexity. Cognitive complexity can depend on subjective variables such as learner factors while linguistic complexity depends on the L2 system or the L2 features (Housen & Kuiken, 2009). Focusing on linguistic complexity as part of CAF, it consists of the learner’s interlanguage system as well as the linguistic elements that make up their interlanguage system (Housen & Kuiken, 2009). Therefore, linguistic complexity at the level of the interlanguage system can be explained as “the size, elaborateness, richness, and diversity of the learner’s linguistic L2 system” while at the level of individual linguistic elements in the interlanguage system, it can be explained as “structural complexity, which itself can be further broken down into formal and the functional complexity of an L2 feature” (Housen & Kuiken, 2009, p. 464). When syntactic complexity measures are concerned, the mean length of clauses per a chosen unit of measure can be considered a global measure of linguistic complexity (Norris & Ortega, 2009). However, Norris and Ortega (2009) argue that syntactic complexity has subconstructs that need to be measured for a better analysis. One of them is complexity via subordination which involves the amount of subordinate or dependent clauses. The other one includes phrasal elaboration or complexity which can be measured by the mean length of clauses.

Skehan (2009) included lexis as part of the CAF measure. Two measures are generally used for lexical complexity in the field. One is measuring lexical diversity which uses the type-token ratio measure. It calculates the ratio of word type to all word tokens (Ellis et al., 2020). However, as Skehan (2009) argues this measure is sensitive to text length and measures compensating for text length are needed. Mean segmental type-token ratio or D-value are popular alternatives (Ellis et al., 2020). Skehan (2009) further argues to include the measure of lexical sophistication since it targets lexical richness in terms of using difficult words. The term difficulty here is defined as the frequency of that word. The presence of more difficult words in a text is the reflection of a more elaborate mental lexicon (Ellis et al., 2020).

Having mentioned CAF and lexical complexity measures, it is necessary to address one last issue. There are general and specific measures of CAF. Whether to use general or specific measures can become an issue. Skehan and Foster (1997) argue that the research goals and hypotheses are key to deciding whether to implement specific or general measures. Foster and Skehan (1996) further argue for generalized measures on the basis that they have more sensitive indices of task performance and are more sensitive to differentiating between levels of proficiency as well as analyzing treatment effects between participant groups.

This section offered a brief overview of measures used for assessing L2 performances. The definition of a task and how to determine its complexity is crucial to measuring L2 performances. It becomes necessary to cover tasks in task-based language teaching, task complexity, and underlying theories.

2.6 Tasks in task-based language teaching

Task-based language teaching (TBLT) has been an influential methodology since the 1980s starting as an approach and developing into a methodology. During this journey, TBLT has undergone diversification through the underlying theories of second language acquisition and the intertwining of task-based pedagogy and research. TBLT research has benefited from combinations of complementary theoretical stances from the Input Hypothesis (Krashen, 1981, 1982) to the Output Hypothesis (Swain, 1985, 1993, 1995), the Interaction Hypothesis (Long, 1996; Swain & Lapkin, 1995, 2002), and the Noticing Hypothesis, (Schmidt, 1990, 2001). These theories and ever-lasting research have opened an extensive discussion on what a task is and led to the development of Task-Based Language Teaching (TBLT). Taking a look at the hypotheses influencing TBLT provides a start to the discussion of what a task is.

Based on his series of morpheme studies, Krashen (1981, 1982) formulated four hypotheses: the acquisition-learning hypothesis, the natural order hypothesis, the monitor hypothesis, and the input hypothesis. The acquisition-learning hypothesis suggests that conscious learning and subconscious acquisition are separate processes so conscious learning “does not turn into” subconscious acquisition (Krashen, 1982, p. 83). It further makes the claim that “in intake-rich informal environments, acquisition occurs, and in intake-poor classrooms, acquisition suffers” (Krashen, 1981, p. 116). This hypothesis implements itself in task-based learning by ensuring that students engage in meaning-focused, communicative tasks instead of form-focused exercises (Nunan, 2004). Therefore, tasks should be meaning-focused to give learners opportunities to acquire language subconsciously (Nunan, 2004). The natural order hypothesis comes from the morpheme order findings and claims that there is a natural order in which learners acquire key grammatical features regardless of their first language and the order of presentation in which these features are presented in the classroom (Krashen, 1982). Although this hypothesis does not have an immediate implication in task-based language learning, it leads to the argument that sequencing grammatical features has very little point. Combining this hypothesis with the input hypothesis, Krashen (1982) claims that "unsequenced but natural input, it is hypothesized, will contain a rich variety of structure--if it is comprehensible, there will be i + 1 for everyone as long as there is enough input" (p. 68). The input hypothesis' focus on comprehensible input which is a little beyond the level of competence emphasizes the importance of receptive skills such as listening and reading rather than productive skills such as speaking and writing (Nunan, 2004). Therefore, it advocates tasks rich in providing input.

Swain (1985) challenged Krashen's (1981) proposal on the very limited role of output arguing that language production is as crucial as receiving input in language acquisition because comprehension and production tap into distinct processes with different and intertwined importance. Long (1985) supports the Output Hypothesis by arguing for the necessity of negotiation of meaning in language acquisition. The Interaction Hypothesis suggests that conversational adjustments lead to comprehensible input and these adjustments require revision and reformulation of utterances that promote acquisition (Long, 1996). Ellis (1995) showed in his study that learners who were given opportunities to negotiate for comprehensible input acquired more words than those who were just exposed to pre-modified input. These two hypotheses were influential and informative in designing tasks that promote language output and negotiation of meaning such as jigsaw tasks, information-gap activities, problem-solving tasks, decision-making, or opinion-exchanging tasks. Similarly, Willis and Willis (2001) emphasize the importance of the exchange of meanings in communicative tasks by viewing it as a generator for a language outcome.

Prabhu (1987) along with Long and Crookes (1991) proposed an approach to how communicative use of language can be achieved in language teaching. They suggested putting tasks in the unit of syllabus design which means that teaching is organized around tasks and there would be no addition of linguistic structures to the tasks. This approach views tasks as the primary way for using language in a meaningful manner and leading to interlanguage development. Learners are in a learning environment where they are engaged with meaning (Skehan & Foster, 1997). Willis and Willis (2001) argue that when learners are engaged in tasks requiring interpretation of language and exchanging of meaning, they encounter new forms and are encouraged to meet the language demands of the tasks and rearrange their interlanguage system. Skehan (1998) further argues that tasks should be relevant to real-world activities where the outcome is a priority, and the achievement of a task should be assessed in terms of its outcome. Tasks should provide activities where the target language is used to be able to reach an outcome (Prabhu, 1987). Previous theoretical frameworks and research imply that presenting learners with tasks concerning focus on meaning and real-world activities and involving both language comprehension and production is a way of promoting language development.

Lastly, the Noticing Hypothesis highlights the importance of tasks requiring the allocation of attention in facilitating language acquisition. Schmidt (1990) argues that intake occurs depending on noticing, subliminal perception, and the conscious control that learners have over the process of intake. Schmidt (1990) further highlights that “task demands are a powerful determinant of what is noticed” and argue that “what is learned is what is noticed” (p. 143). Skehan (2009) also elaborates on task conditions and characteristics that affect what is produced. Abstract or concrete task demands and familiarity with information affect language performance and the chances of noticing (Schmidt, 1990; Skehan, 2009). Additionally, interactive tasks can be advantageous for language performance (Skehan, 2009). As an early study on this issue, Schmidt and Frota (1986) revealed the importance of interaction and input by comparing Schmidt’s class notes of what he had been taught with his taped conversations in the target language. The comparison revealed that he used the forms he noticed during his interactions in Portuguese with Brazilians, and he did not use the forms that never appeared in the input during these interactions (Schmidt, 1990). Lastly, tasks that demand manipulation of information and organization will probably impose high task demands and ultimately affect learners’ allocation of attention, noticing, and language performance (Schmidt, 1990; Skehan, 2009).

TBLT focuses on tasks that are meaningful in the real world and include activities corresponding with what learners need to be able to do in the target language. These target tasks that are identified through a needs analysis of a group of learners sometimes are not suitable for classroom implementation and a need for pedagogical adjustments can be required. Consequently, pedagogic tasks become a necessary part of a task syllabus where tasks serve as the content and a series of tasks are sequenced to become gradually more complex (Long, 2015). This study refers to pedagogic tasks when the term “task” is used.

This section has provided a general review of influential theories, hypotheses, and task-based pedagogy and the research affecting the development of TBLT and the definitions of a task. In light of the previous discussion, tasks can impose different demands and complexity which affect learners' cognitive engagement, allocation of attention, noticing, and pedagogical outcomes. Therefore, it is important to cover in the next section the psycholinguistic perspective on tasks and reveal how tasks can be perceived as a tool affecting learners' language performance and cognitive processes. Most studies of tasks from a psycholinguistic perspective have been influenced by two main theoretical stances: The Cognition Hypothesis (Robinson, 2001b, 2011a) and the Trade-Off Hypothesis (Skehan, 1998, 2009). Both hypotheses follow the TBLT approach by viewing tasks as a central tool for affecting how learners use and acquire language, but they differ in how they see the capacity of attentional resources used while performing tasks with different demands and how the allocation of these resources influence parts of learners' language production. The Cognition Hypothesis (Robinson, 2001b, 2011a) argues for a multiple-resource view and that attention can expand to meet the demands of the tasks while the Trade-Off Hypothesis (Skehan, 2009) argues for limited attentional resources and that more complex tasks requiring attentional demands can cause competition and result in prioritizing some parts of language production at the expense of the others (Skehan, 2014). Both approaches offer a valuable contribution to cognitively driven task-based research. Therefore, the next section offers an investigation of Skehan's and Robinson's perspectives in detail.

2.7 Skehan's cognitive understanding of tasks

Skehan (1996, 1998, 2009) has developed his cognitive approach towards tasks and task-based instruction by analyzing the use of linguistic resources and the subsequent problems posed by the development of the interlanguage system in meaning-driven communication and tasks. By doing so, he makes a distinction between a rule-based and exemplar-based system. The rule-based system is generative which allows the replacement of old rules with new ones through linguistic materials and can be creatively modified to meet the meaning demands of the communication (Skehan, 1996). Stretching and allowing the growth of the interlanguage system is necessary for language acquisition. However, it comes with a high cognitive cost, and its implementation during speech with the concern of highly accurate utterances is prone to disrupt maxims of conversation (Skehan, 1996). The exemplar system, on the other hand, is less cognitively demanding as it consists of useful lexical chunks or formulaic items. These items are most likely to be learned in the context of communication through strategies of comprehension which sometimes leads to the neglect of form and linguistic strategies (Skehan, 1996). Although the use of the exemplar system can be very handy under the pressure of processing and conveying meaning, it can ultimately lead to over-reliance on comprehension and communication strategies and therefore the removal of the necessary engagement with the linguistic material to stretch interlanguage (Skehan, 1996). Besides its limitation, the exemplar system allows learners to spare attentional resources as the retrieval of ready-made lexical chunks is cognitively less demanding and avoids excessive computation during meaningful communication.

Skehan argues that both systems have their values when working together to achieve the communicative conditions and goals imposed on learners during task performance. When learners are under pressure for accessibility, they can rely on a well-organized exemplar system and benefit from an easily accessible memory system. If communicative conditions impose an analytical approach and a focus on planning and form, a rule-based system can be utilized. This dual system allows language learners to switch based on the prioritization of processing demands and ultimately to organize their capacity and processing resources for successful communication (Skehan, 1998). Skehan (1996) emphasizes the importance of both systems for language development because they supplement each other in terms of interlanguage development. Combining both systems inducing both rule-based and exemplar-based learning can result in a "relexicalized repertoire of language" (Skehan, 1996, p. 43). What learners create as exemplars might not always be correct exemplars. In a system where both rule-based and exemplar-based systems work together, the control of a more evolved underlying rule-based system can "relexicalize" potentially erroneous exemplar which was previously found to be communicatively effective and avoid possible syntactic fossilization (Skehan, 1996). Consequently, Skehan proposes task-based instruction where both rule-based and exemplar-based systems are involved in communicative conditions and goals.

2.7.1 Task-based instruction from Skehan's framework

Skehan (1996, 1998) underlines the potential problems of a task-based approach in interlanguage development stemming from the unbalanced prioritization of meaning over form. The suggested solution is to recognize the pertinence of scarcity of attentional resources leading to the prioritization of meaning and to recognize the role of the dual-mode system in fostering the development of the interlanguage system (Skehan, 1996). Accordingly, Skehan (1996) developed a framework for task-based instruction by providing three major task sequencing features: code complexity, cognitive complexity, and communicative stress. Code complexity refers to linguistic complexity and variety that some tasks might induce at different degrees during language production and completion of a task. Cognitive complexity is related to the content of what is produced during communication and takes the conceptualization stage in Levelt's (1989) speech production model as a reference point. Cognitive complexity further distinguishes two areas concerning processing burdens: cognitive processing and cognitive familiarity. Following the conceptualization stage in Levelt's (1989) model, Skehan (1996) relates cognitive processing to how much the learner relies on online computation during task performance and to the need for active cognitive engagement with the task content. Cognitive familiarity, on the other hand, refers to the reliance on already organized and ready-made solutions and materials to complete the task. Finally, communicative stress includes characteristics related to performance conditions such as time pressure, and the number of participants (see Table 1).

Cognitive familiarity includes familiarity with the topics and how predictable the topic is for the task performer which depends on the available background knowledge. Familiarity with discourse genre is included in cognitive familiarity as it is concerned with the use of ready-made available genre schemas. Familiarity with the task under cognitive familiarity refers to the learner's knowledge about the task type. This is concerned with whether the learner applies previous successful communicative strategies or comes up with new strategies with an unfamiliar task. Cognitive processing as the other sub-dimension under cognitive complexity is concerned with the processing load required during performing a task. Organization of information is a factor affecting cognitive processing as it includes the organization of logic and the structure of information related to the task. The amount of computation under this dimension is related to the transformation of information and computation effort that is required for the completion of a task. The clarity and sufficiency of information given to learners during task performance are concerned with how directly the information is presented to the learners as well as the extent to which learners need to infer from the information given during task performance.

Table 1. Framework for Task-based Instruction

Code Complexity	Cognitive Complexity	Communicative Stress
-Linguistic complexity, range, and variety -Lexical variety and vocabulary load -Density and redundancy in language production	-Cognitive Familiarity - Being familiar with topics and topic's predictability. - Familiarity of the discourse genre - Familiarity of task -Cognitive Processing - Organizing information - Amount of 'computation' - Sufficiency and clarity of information given	-Time limitations and pressure -Speed of presentation -Number of participants -Length of texts used -Type of response (modality) -Opportunities for controlling the interaction

Communicative stress, as the final dimension, is concerned with the conditions of pressure related to performance. This includes how urgently the task needs to be completed and how learners perceive this pressure. The speed at which materials are presented can lead to more processing demands especially when the materials are lengthy. Controlling interaction refers to the extent to which the control of implementing the task is given to learners such as being able to decrease the presentation speed of information or being free to ask for clarification.

By proposing such a framework, Skehan (1996, 1998) aims to analyze, compare, and sequence tasks in a principled way to tap into a balanced use of attention to meaning and form and to promote high learning benefits. By doing so, Skehan (1996, 1998, 2009, 2014) proposes the Limited Attentional Capacity Hypothesis discussing the bi-directional effect of task characteristics, and language performance and development, and lays out his predictions of these effects on complexity, accuracy, and fluency (CAF).

2.7.2 The Limited Attentional Capacity and Trade-off Hypothesis

Skehan's (1996, 1998, 2009, 2014) predictions of manipulating tasks along the three dimensions are based on the effect of the competition of attentional resources on CAF performances. The hypothesis assumes that attentional resources are limited which results in failure of paying concurrent attention and therefore competition over areas of L2 performances during demanding tasks. Skehan (2009, 2014) argues that when learners encounter cognitively demanding tasks, they prioritize meaning over form and fluency over accuracy and complexity. This linguistic compromise is explained as the "Trade-off Hypothesis" since learners' limited attentional capacity results in choosing between accuracy and complexity (Skehan, 2014). Skehan built upon VanPatten's (1990) study about the effects of limited attentional capacity on what to pay attention to during meaning-driven communication. In the study, Spanish L2 learners were randomly assigned to four listening task conditions in which participants were asked to pay attention to meaning alone (control group); concurrent attention to meaning and key lexical item; concurrent attention to meaning and grammar (definite article); concurrent attention to meaning and a specific verb form. The study revealed the difficulty in paying simultaneous conscious attention to both meaning and form in the input. In the study, lower comprehension was observed when paying simultaneous conscious attention and learners prioritized meaning over form if there was no specific instruction given to them to pay attention to form (VanPatten, 1990). Based on the findings, Skehan and Foster (2001) made the following argument:

> In this view, tasks which are cognitively demanding in their content are likely to draw attentional resources away from language forms, encouraging learners to avoid more attention-demanding structures in favour of simpler language for which they have already developed automatic processing (the 'safety first approach'). Conversely, very cognitively demanding content might result in learners paying insufficient attention to language forms which still require controlled processing and which are therefore likely to be poorly done (the 'accuracy last' approach) (p. 189).

Figure 1 demonstrates the tension between the performance areas by relating fluency to the meaning or "getting a task done" and relating form (accuracy and complexity) to "language focus and development" (Skehan & Foster, 2001, p. 190).

Foster and Skehan (1996) support this hypothesis of tension in their study where participants were given tasks with different cognitive demands. They found different performance levels in linguistic aspects. While the personal task resulted in greater accuracy and lower complexity, the narrative task brought about higher complexity, but the accuracy level was found to be low in this case. These results support the argument of the trade-off effect between accuracy and complexity. Skehan and Foster (2001) further argue that attempting to achieve greater complexity is related to being experimental with the interlanguage system as it includes extending aspects of this system. Accuracy, on the other hand, taps into achieving greater control over more stable elements in the interlanguage system, therefore its emphasis becomes the achievement of communication by using an IL repertoire. Skehan and Foster (2001) recognize the role of individual differences in this regard along with the manipulation of task conditions and characteristics.

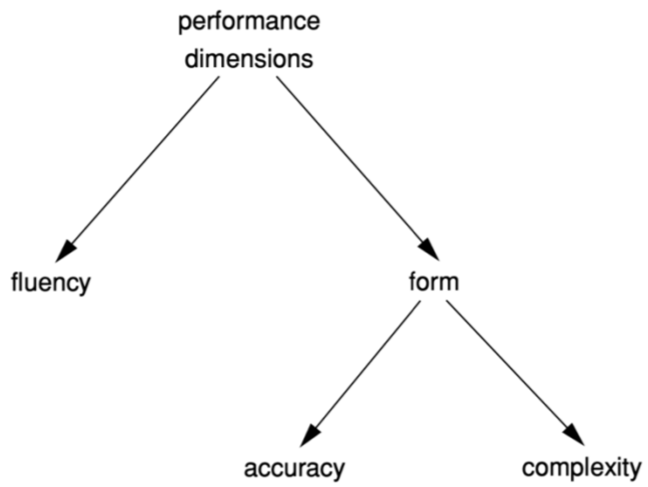


Figure 1. Theorizing dimensions of performance
(Skehan & Foster, 2001, p. 190)

Considering Limited Attentional Capacity and Trade-off Hypothesis, Skehan (1998) argues for achieving a balance in language teaching and task design by avoiding over-prioritization of any of the aforementioned dimensions in the long term. Therefore, sequencing tasks and instructional activities in a way that encourages development in fluency, complexity, and accuracy becomes crucial.

The following section provides Robinson's (2001a) framework which holds an opposite understanding of allocating attentional resources and therefore proposes different predictions of the effects of task complexity on L2 performance.

2.8 Triadic Componential Framework and the Cognition Hypothesis

An alternative account of explaining how task demands affect L2 performance and development was developed by Robinson (2001b, 2011a). Skehan (1998, 2009) proposes a competition over complexity and accuracy and that increase in both constructs simultaneously is not possible due to trade-off effects while Robinson (2011a) argues for the possibility of an increase in performance in these two constructs thanks to the multiple attentional capacity.

Robinson (2011a) takes the Multiple Attentional Capacity as the basis for the use of attentional resources and proposes that people are not limited to directing their attention to only one aspect of their performances. He was inspired by Wickens' (2002, 2007) theories of attention. Wickens (2002) argues that two tasks requiring demands on the same dimensions (e.g., demanding visual perception for two tasks) will result in more interference in task performance than two tasks posing demands on separate dimensions (e.g., one task posing visual demand and another one posing auditory demand). Therefore, the multiple resource model suggests that there are three dimensions with different physiological mechanisms that affect time-sharing performance: processing stages, processing modalities, and processing codes (Wickens, 2007).

Processing stages include perception, cognition, and responding. While resources spent for perception and cognition appear to be the same, resources used for responding are functionally separate from perception and cognition. Responding is responsible for selecting and executing. For example, tasks demanding speech recognition (perception) and production (responding) underline different processes and can be associated with different brain structures. Therefore, these tasks targeting separate dimensions are predicted to pose less dual-task interference compared to tasks targeting the same dimension such as perception and cognition (e.g., speech comprehension demand along with verbal rehearsal) (Wickens, 2002).

Processing modalities make a distinction between visual and auditory demands arguing that tasks demanding attention from the same modality result in more interference compared to tasks demanding attention from the opposite modalities (Wickens, 2007). For example, listening to a lecture (auditory) and listening to a friend simultaneously (auditory) is more taxing on attentional resources than listening to a lecture (auditory) and looking at a map and trying to understand (visual). Wickens (2002) argues that "cross-modal time-sharing is better than intra-modal time-sharing" (p. 164).

Processing codes make a distinction between verbal/linguistic material and spatial/analog which is non-verbal material. Wickens (2007) argues that dividing attention across codes is less taxing and causes less interference than dividing it within codes. For example, driving (spatial/analog) and listening to a song (verbal/linguistic) is easier on dividing attentional resources than listening to a song (verbal/linguistic) and listening to a friend (verbal/linguistic).

Drawing on Wickens' (2002, 2007) multiple attention theory, Robinson (2011a) argues that concurrent attention to different aspects of the language is possible and certain manipulation of task demands can lead to multiple uses of attentional resources rather than competition over these resources.

Robinson (2001a, 2007a, 2007b, 2011a) offers a taxonomy for classifying pedagogic tasks based on their complexity levels. Predictions of L2 performance and development following the taxonomy of tasks are explained and elaborated in the Cognition Hypothesis (Robinson, 2001a, 2001b, 2011a). Robinson's classification of tasks differs from Skehan's three-way distinction of code complexity, cognitive complexity, and communicative stress. Robinson (2001a) makes a distinction between task complexity and task difficulty in his Triadic Componential Framework (TCF) (See Figure 2). He also makes a distinction between resource-directing and resource-dispersing demands under task complexity. The reason for such a distinction can arguably be traced to Wickens' (2002, 2007) arguments regarding the interference effects of task demands on the same dimensions as opposed to separate dimensions. He manages to offer a detailed framework of task classification including a relationship between task and learner factors, and linguistic performances. His framework offers three aspects of task characteristics: task complexity, task condition, and task difficulty.

Task complexity (Cognitive factors)	**Task condition** (Interactive factors)	**Task difficulty** (Learner factors)
(Classification criteria: cognitive demands)	(Classification criteria: interactional demands)	(Classification criteria: ability requirements)
(Classification procedure: information-theoretic analyses)	(Classification procedure: behavior descriptive analyses)	(Classification procedure: ability assessment analyses)
Sub categories: a. resource-directing variables making cognitive/conceptual demands	Sub categories: a. participation variables making interactional demands	Sub categories: a. ability variables and task relevant resource differentials
± here and now (Gilabert, 2007) ± few elements (Kuiken et al., 2005) ± spatial reasoning (Becker & Carroll, 1997) ± causal reasoning (Robinson, 2005a) ± intentional reasoning (Ishikawa, 2008) ± perspective-taking (MacWhinney, 1999)	± open solution (Lambert & Engler, 2007) ± one way flow (Pica et al., 1993) ± convergent solution (Duff, 1986) ± few participants (Crookes, 1986) ± few contributions needed (McGrath, 1984) ± negotiation not needed (Gass & Varonis, 1985)	h/l working memory (Mackey et al., 2002) h/l reasoning (Stanovitch, 1999) h/l task-switching (Monsell, 2003) h/l aptitude (Robinson, 2005b) h/l field independence (Skehan, 1998) h/l mind-reading (Langston et al., 2002)
b. resource-dispersing variables making performative/procedural demands	b. participant variables making interactant demands	b. affective variables and task relevant state-trait differentials
± planning time (Skehan, 1998) ± prior knowledge (Urwin, 1999) ± single task (Robinsonet al., 1995) ± task structure (Skehan & Foster, 1999) ± few steps (Fleishman & Quaintance, 1984) ± independency of steps (Romiszowski, 2004)	± same proficiency (Yule & MacDonald, 1990) ± same gender (Pica et al., 1991) ± familiar (Plough & Gass, 1993) ± shared content knowledge (Pica et al., 1993) ± equal status and role (Yule & MacDonald, 1990) ± shared cultural knowledge (Brindley, 1987)	h/l openess (Costa & Macrae, 1985) h/l control of emotion (Mayer et al., 2000) h/l task motivation (Dörnyei, 2002) l/h anxiety (MacIntyre & Gardner, 1994) h/l willingness to communicate (MacIntyre, 2002) h/l self-efficacy (Bandura, 1997)

Figure 2. The triadic componential framework
(Robinson, 2011a, p. 7)

Task Complexity is related to cognitive demands posed on learners by tasks and is concerned with attention and memory demands. TCF makes a theoretically based distinction between the resource-directing variable which refers to cognitive demands of tasks and the resource-dispersing variable which is related to performative or procedural demands on the task (Robinson, 2001b, 2003). Resource-directing dimension affects the allocation of cognitive resources while also potentially directing the attention to L2 codes and structures to meet the conceptual demands of a task (Robinson, 2001a, 2001b, 2005, 2007a, 2011a). When there is an increase in resource-directing dimension such as intentional reasoning demands in a task, this increase results in finding linguistic resources such as believe and wonder indicating the mental states of others to meet the reasoning demands (Robinson, 2011b). Manipulation of this dimension allows learners to be engaged in different levels of conceptualization while also directing the use of attention and memory with the objective of meeting linguistic demands posed by the conceptualization (Robinson, 2011b). "Complex task demands lead to greater effort at conceptualization and elicit the morphologically richer and structurally more complex syntactic mode" (Robinson, 2011b, p. 14). Some of the resource-directing variables include +- here-and-now (time reference to events happening now versus events that happened in the past), +- few elements (low or high number elements to be considered in a task), +- intentional reasoning (simply transforming information versus reasoning about others' intentions), +- causal reasoning (simply transforming information versus reasoning about cause-effect in events and relationships) or +- perspective taking (first person perspective telling on event versus thinking from other people's perspective (Robinson & Gilabert, 2007). Resource-dispersing variable, on the other hand, makes performative or procedural demands on the task such as not giving planning time. Consequently, an increase in this variable leads to distraction from language-related concerns and analyzing linguistic items. For example, removing planning time from a task, therefore making it more complex, prevents learners from directing their attention to linguistic items. The increase in this variable results in the gradual removal of processing support for extending interlanguage while it leads to practicing automatic L2 access and use (Robinson, 2011b). This dimension allows manipulation in +- planning time, +- prior knowledge, +-single task, +- task structure, +- few steps, +- independency of steps.

Robinson (2005) argues that increasing complexity in this dimension is also important because it provides conditions similar to real-time language use. “Practice along them could therefore be argued to facilitate real-time access to, and control of, an already established and developing repertoire of language, rather than to facilitate new form-function mappings in the L2” (Robinson, 2005, p. 7).

Task condition is concerned with the interactional demands and the grouping of the participants (Robinson, 2005). These demands can include manipulation of the flow of information whether it goes one way or two ways and the nature of grouping such as pairs and groups, the familiarity with each other in the group, and gender distribution (Robinson, 2001a). Interactional demands are differentiated as participation variables and participant variables. Participation variables are concerned with whether tasks require multiple answers (open) or a limited number of answers (closed), whether they are individual (one-way) or pair/group work (two-way), and whether they are convergent (having to decide on one answer) or divergent (does not necessarily require one answer). Participant variables are concerned with proficiency, gender, shared content, cultural knowledge, and familiarity in the group.

Finally, task difficulty recognizes the role of individual differences and is concerned with the learner factors. Robinson (2007a) makes a distinction between ability variables and affective variables in task difficulty. Ability variables include aptitude, working memory reasoning, task-switching, and field independence while affective variables include openness, control of emotion, task motivation, anxiety, and willingness to communicate. Robinson and Gilabert (2007) argue that when ability and affective variables factor into meeting complex task demands, individuals with high levels of these variables are assumed to be more successful in complex task performance than those with low levels of ability and affective variables.

Robinson (2005) argues that out of these three main factors, task complexity should be the foundation in sequencing tasks pedagogically in a task-based syllabus. Additionally, these pedagogic tasks should be ordered from cognitively simple to more complex levels. A reason behind this increasing complexity is that such a sequence should promote L2 development by encouraging the acquisition of new L2 knowledge and restructuring the interlanguage system (Robinson, 2001b).

Robinson (2011a) further highlights the enhancing effects of an increase in resource-dispersing dimension on L2 performance. Resource-dispersing dimension can promote access and control of interlanguage when the complexity is practiced first and then is increased again to promote further interlanguage development. After covering the cognitive understanding of the Limited Attentional Capacity Hypothesis and the Cognition Hypothesis, it is crucial to analyze the implications of these theories on writing to touch upon the issue of modality in cognitive TBLT models.

2.9 Modality in the revision of Skehan's and Robinson's models

The review of Skehan's and Robinson's frameworks suggests that the oral mode was prioritized for the development of these theories. Although neither Skehan nor Robinson explicitly states that their frameworks solely account for oral production, they do not provide any discussions for how the conditions they describe affect written task performance. Ortega (2012) discusses several reasons for the neglect of the L2 written mode. One reason for this neglect is the perception of literacy as a culture-dependent development since it comes as a secondary manifestation of primary oral capacity. Another reason can be attributed to the offline nature of writing that allows learners to have self-paced and monitored language production. This has been argued to be "an impure reflection of tacit linguistic knowledge (for those working within a formal linguistic perspective) or of online ability for use (for those motivated by a functional perspective" (Ortega, 2012, p. 405). Although spontaneous and unmonitored oral data have been preferred for analyses, in recent years, writing has been investigated in terms of its potential role for L2 learning and the alignment of this role with SLA theories through analyses of individual and collaborative writing, and analyses of task performances in speaking and writing (Manchon & Williams, 2016). However, Byrnes and Manchon (2014) argue that the idiosyncrasy of written task performance should be cautiously and critically analyzed considering the potential problems of researching writing tasks under the influence of speech-originated theories. Consequently, Byrnes and Manchon (2014) discuss the necessity of reconsidering central constructs such as task complexity and task conditions for a better understanding of written performances.

Manchon (2014) concludes in her analysis of Robinson's Triadic Componential Framework that revision is necessary for the written mode in dimensions of task complexity, task condition, and task difficulty. In the task complexity dimension, the resource-directing variable fails to recognize the recursive nature of writing processes which gives a flexible division of attention between the conceptualization of the message and linguistic encoding. Manchon (2014) further underlines the dual task repetition effect (i.e., external and internal task repetition) of writing stemming from unique planning opportunities. While external task repetition is the result of processing feedback on the written product, internal task repetition derives from two-way interaction between reflection and text production processes. This two-way interaction benefits from the availability of time in writing and provides learners with opportunities of matching meaning to language. The recursive nature of writing becomes a crucial factor in internal task repetition since the writing process is dynamic and cyclical between stages of planning, transcribing, evaluating, and revising. It allows modification and evolvement of emerging text through constant decision-making between meaning-making and linguistic choices (Manchon, 2014). Kormos (2014) also contributes to the discussion by underlying different resource-directing demands in writing compared to speaking and argues that written mode can direct learners' attention to linguistic form due to the greater availability of time and cyclical nature of writing. Manchon (2014) provides an additional layer to Kormos' (2014) argument emphasizing the issue of self-imposed goals in directing attention to L2 forms in writing which are not included in the current TCF. In this respect, planning is not only the question of availability of time but also the question of goal setting and pursuing these goals as a writer. The distribution of attentional resources is set based on the decision of achieving or prioritizing various goals in the composing process (Manchon, 2014). Consequently, the planning time for the written mode can fall into a resource-directing or resource-dispersing variable depending on learners' goal-setting. Planning as it was modeled in the TCF occurs as an optional variable that can be manipulated externally [+/-planning] which leads to the inference of it being limited to the online pressures and processes in speaking. However, planning is much more complicated and unique in writing. It is, in fact, an intrinsic component of the writing process as it involves both offline and online planning (Manchon, 2014). Therefore, Kormos (2014) argues that writers are comparatively under less pressure as content planning is integrated into the actual writing. If learners feel the need, they can focus more on the planning process by stopping the actual writing process which can lead to deeper planning opportunities during a writing task compared to a speaking task (Vasylets et al., 2017). The planning variable seems to require revisions to account for the unique opportunities in writing.

Task conditions are another dimension that requires rethinking to account for the written mode. Considering the range of task implementations such as individual or collaborative performance of meaning-making and environments (pen and paper, electronic), task conditions need a more comprehensive account. In the discussion of the [+/- interaction] variable, Manchon (2014) brings about a different perspective on analyzing interaction in writing and argues that provision and processing of feedback should be considered as a factor in task condition. Feedback constitutes a crucial part of learning and teaching writing, and the process of feedback is part of the practice of writing in academia. (Manchon, 2014).

The task difficulty dimension, similarly, calls for an expansion and a revision to account for additional demands of individual factors in writing. Ability factors should provide accounts for genre knowledge and L2 writing expertise. What constitutes L2 writing expertise is the interaction and combination of writing knowledge and writing skills that form L2 writers' linguistic repertoire (Manchon, 2014). Manchon (2014) argues that L2 writing expertise can lead to a wider range of variation than oral language expertise. Additionally, motivation to write and goals for writing should be included in the affective factors because learners' motivation and goals for writing tasks are related to how much attention they aim to give to the quality of their written product (Manchon, 2014).

Another evaluation of modality and place of writing in cognitive TBLT models was offered by Tavakoli (2014). She highlights the scarcity of research on written task performances and comparative data on speaking and writing performances. She further questions whether the theorization of task complexity by

Robinson (2001a, 2001b, 2007a, 2007b, 2011a) and Skehan (1996, 1998, 2009) account for the differences in written and oral tasks and questions how language production processes and outcomes in writing and speaking performances are affected by the different conditions. Her initial analysis of these models suggests that the construct of task and task complexity is defined in the oral mode (Tavakoli, 2014). This understanding can be seen in what constitutes communicative pressure (Tavakoli, 2014). Skehan (1998) defines communicative stress in terms of its implications for the speed of presenting information, time constraints for processing, and being able to slow down interaction. This definition suggests that communicative stress is defined in terms of processes of spoken production, and little evidence is offered in the model on how to determine communicative stress in writing (Tavakoli, 2014). A similar problem can be found in the lack of discussion on whether the inherent characteristic of available online planning time in writing decreases the complexity of the task (Tavakoli, 2014). In light of these questions and the call for an expansion of current models, Kormos (2014) offers to incorporate mode as a task complexity factor to provide a more comprehensive account of the theoretical framework. She argues that modality poses both resource-dispersing and resource-directing demands, therefore it should be considered as a task complexity dimension. The resource-dispersing variable of modality stems from the fact that speaking is a time-constrained activity while writing is self-paced which can offer more planning and strategic changes in the allocation of attention. When the resource-directing variable is concerned, more relaxed conditions in writing and the cyclical nature of writing processes can result in more attention to linguistic form both during and after production (Kormos, 2014).

This critical evaluation of influential cognitive TBLT models reveals that Robinson (2001a, 2001b, 2005, 2007a, 2007b) and Skehan (1996, 1998, 2009) may not have considered possible theorizing effects of writing on task complexity and task demands, and on L2 learning and performance. However, this theoretical gap in covering writing should not lead to discarding these frameworks from applying to writing. Instead, the adjustment of these models accounting for the idiosyncrasy of writing conditions can be offered through mode-sensitive research. Such adjustments and changes in the theory should be based on relevant empirical data by investigating these theoretical constructs in the written mode. Findings from such studies would shed light on how these models can address and consider task modality and would lead to a mode-sensitive rethinking in cognitive TBLT research. Consequently, this study analyzes the effects of writing and speaking on accuracy and, syntactic and lexical complexity to provide empirical data, also investigates a possible interaction between task complexity and modality to contribute to the mode-sensitive rethinking of cognitive TBLT models.

Having covered critical views on two very influential cognitive models in TBLT, the next section presents a review of relevant empirical studies to the current study concerning the comparison of L2 oral and written performances. Then, it focuses on the studies investigating task complexity effects in L2 oral and written performances. Studies focusing on monologic L2 oral and written production are covered in the next sections.

2.10 Studies of language performance in oral and written modes

One of the early studies related to L2 oral and written production is found in Ellis' (1987) comparison of accuracy in the two modes. 17 participants completed three tasks about storytelling in three conditions; planned writing, planned speaking, and unplanned speaking. Participants were L2 English learners with different L1 backgrounds, and their level of proficiency was early intermediate corresponding to level two of Cambridge proficiency. Linguistic forms under investigation for accuracy involved regular past, irregular past, and past copula. His analysis revealed that planned writing led to the greatest accuracy in the regular past while the least accurate result for the regular past was found in unplanned speech. The planned speech was reported to have an intermediate level of accuracy compared to the other two conditions. Overall, the more accurate use of all the past tense forms under investigation was found in the written performance compared to the oral. Later, Ellis and Yuan (2005) broadened Ellis' (1987) study aiming to isolate whether greater accuracy in writing stemmed from writing as a modality or from the opportunity of careful online planning. In Ellis (1987), participants were given an hour to complete their writing task which can be considered as 'writing/careful within-task planning' while they were given specifically two minutes to plan before starting to tell the story in the oral performance which can be considered as 'speaking/pressured within-task planning' (Ellis & Yuan, 2005, p. 168). Ellis and Yuan (2005) also employed broader accuracy measures in the new study. Ellis and Yuan (2005) mainly focused on different planning conditions' (careful and pressured planning) effects in oral and written performances, but they also reported overall performance differences between the two modes. 42 L2 English learners with L1 Chinese background completed two monologic tasks consisting of picture sets. Picture prompts were argued to be similar in terms of the number of pictures and situations presented. Their measure included fluency calculated by syllables per minute and percentage of disfluencies, global syntactic complexity, syntactic variety calculated by the total number of different grammatical forms, lexical complexity calculated by type/token ratio, and accuracy calculated by error-free clauses and correct verb forms. The fluency measure revealed more syllables per minute and more disfluency in speaking than writing. Written production was found to be both syntactically and lexically more complex and varied than oral production. Finally, writing revealed more accurate language production in both the clausal and correct use of verb measures than speaking. Ellis and Yuan’s (2005) findings for accuracy in correct verb use were aligned with Ellis’ (1987) earlier study.

Corroborating Ellis and Yuan’s (2005) study, Granfeldt (2008) found that writing resulted in more lexical complexity than speaking. L2 French participants completed expository and narrative tasks in different sessions in writing and speaking. Half of the subjects started with speaking first while the other half started with writing first. Although lexical complexity measured by D calculation revealed the same results as Ellis and Yuan (2005), writers produced significantly more errors in lexis than speakers. No significant difference, on the other hand, was found in terms of grammatical errors between the two modes. Contradictory to Ellis and Yuan’s (2005) results, grammatical complexity, measured by a subclause ratio, was found to be lower in writing than in speaking, but the difference was not statistically significant.

Ferrari and Nuzzo (2009, as cited in Vasylets et al., 2019) contributed to the modality studies by including Italian native speakers and comparing their modality differences to that of L2 Italian learners. Their participants consisted of young learners. They measured accuracy and grammatical complexity by the length of clause and ratio of dependent clauses per AS-unit or T-unit and included specific measures by calculating the type of dependent clauses. Their results revealed that native speakers produced significantly more complex syntactic structures in writing than in speaking. However, only a tendency toward higher syntactic complexity was found in writing for L2 learners. While L2 learners' accuracy results corroborated with Granfeldt (2008) showing more accuracy in speech, native speakers were more accurate in writing. Different linguistic behaviors between native speakers and L2 speakers led to the conclusion that the level of linguistic proficiency might be a moderating effect in investigating the superiority of the written or the oral mode (Ferrari & Nuzzo, 2009, as cited in Vasylets et al., 2019).

Some modality studies focused on the latest member of the CAF measures. Bulté and Housen (2009, as cited in Vasylets et al., 2019) and Yu (2009) investigated lexical complexity in written and oral modes. Bulté and Housen (2009, as cited in Vasylets et al., 2019) employed a variety of measures for lexical complexity. Their study with L2 French learners with L1 Dutch revealed similar results to Ellis and Yuan (2005) and Granfeldt (2008). Lexical complexity scores were generally found to be higher in writing than speaking, except for lexical profile measures (Bulté & Housen, 2009, as cited in Vasylets et al., 2019). However, Yu's (2009) study showed different results. He employed D-value to measure lexical complexity between two modes. 25 L2 English participants completed written compositions for writing and interactive interviews for speaking. Yu (2009) found no significant difference in lexical complexity between writing and speaking. He concluded that time pressure and pre-task planning might have a more effective role in lexical diversity than task modality. Another explanation for Yu's (2009) contradictory results to other studies can stem from the fact that tasks had different genres and topics. Participants were given writing compositions based on a topic which was monological while they completed interviews that were interactive and had no particular topic for the oral mode.

A more recent study is from Kormos (2014) which investigated the linguistic differences and cohesive devices between writing and speaking. Although Kormos (2014) included task complexity in her study by using a cartoon description task with pictures forming a coherent storyline and a picture narration task with unrelated pictures, only differences across modalities were discussed in the study. She did not discuss the differences within modality performances in relation to task complexity because these differences were reported in her other studies (Kormos 2011b; Kormos & Trebits, 2012). For this reason, the review of Kormos (2014) suits better in this section. 44 L2 English learners with L1 Hungarian completed a cartoon description and a picture narration task in two modes. Both tasks were argued to be comparable as they had the same number of pictures and were about a similar topic and narrative discourses. Participants completed oral tasks first and after a month they completed the writing tasks. They received two minutes of planning time for oral tasks with no time limit. Participants were given 45 minutes for writing and asked to write a minimum of 150 words. Analyses involved D-value for lexical complexity and the National's Range program for vocabulary range. Syntactic analyses involved the ratio of subordinate clauses to the total number of clauses, the length of clauses, and the mean number of modifiers per noun phrase calculated by the Coh-metrix program. Accuracy measures involved both general and specific measures. The ratio of error-free clauses was calculated for the general measure and the ratio of error-free verbs, past tense, and relative clauses was calculated for specific measures. Uses of cohesive devices were measured by Coh-metrix program. Similar to Ellis (1987) and Ellis and Yuan (2005), higher accuracy in past-tense forms was found in writing in Kormos (2014). Lexical diversity was also found to be higher in writing than in speaking which yields the same results as Ellis and Yuan (2005), Bulté and Housen (2009, as cited in Vasylets et al., 2019) and Granfeldt (2008). Subordination as a syntactic complexity measure, on the other hand, did not reveal any significant differences between the modes. However, modifiers per noun phrase measured by Coh-metrix showed higher levels of noun-phrase complexity in writing than in speaking. Measures of cohesion revealed higher levels in writing than in speaking except for the use of connectives which showed that participants produced more connectives in speech than writing. Based on these findings, Kormos (2014) concluded that available time and the cyclical nature of writing allow learners to have closer monitoring than online characteristics of speaking as higher accuracy in writing suggests. Kormos (2014) also compared the mean and standard deviation values for the accuracy of past tense with the ratio of error-free relative clauses. Her comparison revealed that the accuracy of past-tense forms was higher in writing and showed that students seemed to master these forms while the use of this tense form showed variation in speaking. However, significant degrees of variation of accuracy in relative clauses were found in both modalities of performance which can be interpreted as the result of partially acquired syntactic structure. Based on these comparisons and findings, Kormos (2014) suggests that the cyclical nature of writing and the benefits of the availability of time and monitoring in writing may lead to high accuracy in structures that students have already mastered.

The most recent study on modality is from Vasylets et al. (2019) focusing on the mode in the context of Instructed Second Language Acquisition (ISLA) and the measures of lexical and syntactic complexity, and propositional L2 complexity. In this large-scale study, 290 L2 English learners were assigned to oral and written conditions. Participants were asked to retell an eight-minute clip of Chaplin's film Modern Times with no time limitation in their assigned conditions. They used a tool called Synlex by Lu (2010) for lexical complexity measures and D calculation for lexical diversity. Syntactic complexity measures included the mean length of AS-unit for speech and T-units for writing. Synlex tool by Lu (2011) was used for complexity by coordination, subordination, and nominal complexity. Their results revealed that participants scored significantly higher on all measures in writing than in speaking, except for one measure; the number of words. While their lexical complexity results corroborated with most of the studies (Bulté & Housen, 2009, as cited in Vasylets et al., 2019; Ellis & Yuan, 2005; Granfeldt, 2008; Kormos, 2004), syntactic complexity yielded different results from Granfeldt (2008) and Ferrari and Nuzzo (2009, as cited in Vasylets et al., 2019).

This review of the relatively small number of studies comparing L2 oral and written production yields some characteristics needing to be mentioned. Table 2 shows the summary of the results covered in this section. One common result seems to be higher lexical complexity in writing except for Yu (2009). Participant profile in these studies seems to be a second common feature since all the participants were adult L2 learners except for Ferrari and Nuzzo (2009, as cited in Vasylets et al., 2019). Another common feature among these studies can be found in tasks. Narrative tasks seem to be preferable for modality studies, especially picture description tasks (Ellis, 1987; Ellis & Yuan, 2005; Kormos, 2014; Vayslets et al., 2019).

Some differences can be detected among these studies as well. Participants were from different L1 backgrounds such as Chinese (Ellis, 1987; Ellis & Yuan, 2005; Yu, 2009), Swedish (Granfeldt, 2008), Dutch (Bulté & Housen, 2009, as cited in Vasylets et al., 2019), and Hungarian (Kormos, 2014).

Table 2. Studies Investigating Language Performance in Two Modes

	Findings			
	Syntactic Complexity	Lexical Complexity	Accuracy	Fluency
Ellis (1987)	X	X	+ writing	X
Ellis and Yuan (2005)	+ writing	+ writing	+ writing	+ speaking
Granfeldt (2008)	=	+ writing	+ speaking[d]	X
Ferrari and Nuzzo (2009, as cited in Vasylets et al., 2019)	=	X	+ speaking	X
Yu (2009)	X	=	X	X
Bulté and Housen (2009, as cited in Vasylets et al., 2019)	X	+ writing[c]	X	X
Kormos (2014)	+ writing[a]	+ writing	+ writing	X
Vasylets et al. (2019)	+ writing[b]	+ writing	X	X

Note. + shows that L2 performance was found to be higher in the modality stated; = shows a neutral effect of modality on L2 performance; X means not investigated; [a] it was found higher for noun-phrase complexity; [b] it was found higher except for the number of words; [c] it was found higher except for lexical profile measures; [d] it was found higher for lexical accuracy.

Along with their L1 backgrounds, they were also learners of different L2 such as English, French, and Italian. Determining the level of proficiency seems to be another difference changing from standardized tests (Ellis, 1987; Ellis & Yuan, 2005) to teacher ratings (Kormos, 2014). Similarly, different ranges of research focus and measures can be observed. While Ellis and Yuan (2005) stand out as the only study to employ fluency measures in comparing the two modes, Ellis (1987) stands out with a narrow research interest by focusing only on accuracy. Bulté and Housen

(2009, as cited in Vasylets et al., 2019) and Yu (2009) show the same interest by focusing only on lexical complexity. Other studies employed a broader research focus by measuring accuracy and complexity (Ellis & Yuan, 2005; Ferrari & Nuzzo, 2009, as cited in Vasylets et al., 2019; Granfeldt, 2008; Kormos, 2014). Additionally, Kormos (2014) and Ferrari and Nuzzo (2009, as cited in Vasylets et al., 2019) stand out for their measures of cohesion, and Vasylets et al. (2019) for propositional complexity. The choice of measures also shows variation making it difficult to make a comparison. Although drawing clear-cut conclusions is rather difficult considering crucial differences among these studies, some of the findings indicate consistency.

Lexical complexity shows the most consistent findings since it was found to be higher in writing among the studies except for Yu (2009). The same cannot be drawn for accuracy because it yields mixed results. While higher results were found for writing in Ellis (1987), Ellis and Yuan (2005), and Kormos (2014), Granfeldt (2008) and Ferrari and Nuzzo (2009, as cited in Vasylets et al., 2019) revealed higher results in speaking. Syntactic complexity also revealed mixed results. While some studies reported higher scores in writing (Ellis & Yuan, 2005; Kormos, 2014; Vasylets et al., 2019), others did not find any statistically significant effects of modality on syntactic complexity (Ferrari & Nuzzo, 2009, as cited in Vasylets et al., 2019; Granfeldt, 2008; Kormos, 2014). Clearly, more research is needed to investigate the effects of modality, if any, on L2 learners' performances.

2.11 Task complexity studies in oral and written mode

The primary aim is to review relevant studies which investigated task complexity effects of L2 oral and written performances within the same study.

As the earliest study investigating the task complexity effects in L2 oral and written performances, Kuiken and Vedder (2011) focused on the role of L2 proficiency in relation to modality and employed a between-subject design. The aim was to investigate how proficiency affects task complexity in the two modes. Following the Cognition Hypothesis (Robinson, 2001a, 2011a), they employed +/- few elements as the definition of task complexity and formed argumentative tasks involving giving advice to a friend in the oral and written mode. Writers received 40 minutes. No time limit was given for oral production. Participants were L2 Italian with L1 Dutch. They were assigned to oral or written conditions. The oral mode consisted of 44 participants while the written mode involved 91 with the same participant background. They employed accuracy and syntactic and lexical complexity measures (see Table 3 for details). While T-unit was calculated for the written data, AS-unit was chosen for the oral data. Their analyses revealed no interaction between task complexity and proficiency level. They also reported that participants produced fewer errors in the complex task compared to the simple task in both modes suggesting complexity as the main factor for accuracy compared to modality. They did not find any significant effect of task complexity on lexical complexity in either of the modes. However, the effects of task complexity on syntactic complexity showed unexpected results in writing and in speaking. They detected no task complexity effect on syntactic complexity in the written mode. Contrary to their hypothesis, significantly more dependent clauses were found in the simple task in the oral mode. Kuiken and Vedder (2011) did not conduct any statistical analysis to compare writing and speaking scores with each other, but they reported their observations based on the mean differences. Syntactic and lexical complexity were observed to be higher in writing while speaking had higher accuracy means than writing. Overall, they concluded that the effects of task complexity on linguistic performance are not substantially bound by task modality therefore, it is not necessary to include the modality in Robinson's TCF.

A similar issue was investigated by Kormos and Trebits (2012) and Tavakoli (2014) and different conclusions emerged from their findings compared to Kuiken and Vedder (2011). Kormos and Trebits (2012) conducted a study investigating task complexity, modality, and aptitude. They employed narrative tasks with different cognitive demands. Task complexity was determined by whether participants received a coherent plot or not. The definition and justification of the conceptualization demands were inspired by the Cognition Hypothesis (Robinson, 2001a, 2011a). While pictures in a coherent story order were argued to be low in demand since it poses low cognitive demands on conceptualization, unrelated pictures with no storyline were considered to be more demanding based on the higher demands on conceptualization. Parallel versions of these tasks were provided for counterbalancing. 44 L2 learners with L1 Hungarian completed the tasks. They finished both simple and complex writing tasks in one session and were given 30 minutes for each task. They were asked to write 150 words minimum for each task. The choice of starting with a simple or complex task was left to the participants during the writing tasks. Oral tasks, on the other hand, were randomly given to participants. Language productions were analyzed for accuracy and complexity (See Table 3 for details). Their analyses suggested a possible interaction between task demands (pictures with coherent storyline vs. unrelated pictures) and modality (oral vs. written). However, they did not conduct any statistical analysis for detecting an interaction. The comparison of the two modes revealed higher performance in the written mode for lexical variety, and the ratio of error-free clauses and error-free verbs. The comparison of tasks in two different modes revealed that participants produced more varied vocabulary in the written mode in both coherent storyline and unrelated picture tasks than in the oral mode. The written mode elicited a higher proportion of error-free clauses in coherent storyline tasks than the oral mode. Similarly, a higher ratio of error-free verbs was found in coherent storyline tasks in the written mode than in the oral mode. Modality did not have a significant effect on the ratio of error-free clauses and verbs in unrelated picture tasks. Differences within the mode were also found in simple and complex tasks. Oral production revealed more varied vocabulary in the coherent storyline task than in the unrelated picture task. Oral mode elicited higher accuracy measured by error-free verbs in more demanding unrelated pictures tasks than the coherent storyline task. Task demands seemed to affect the results of syntactic complexity in the written mode as participants produced longer clauses and used more relative clauses in the unrelated pictures task than in the coherent storyline task. These findings lead to a different interpretation than Kuiken and Vedder (2011) and suggest a potential interaction between task complexity and modality. Kormos and Trebits (2012) conclude that task demands seem to lead to different patterns of performance in writing and speaking. Therefore, the effect of modality on task performance should be considered when analyzing and drawing conclusions about task complexity.

Tavakoli (2014) similarly investigated task complexity effects in L2 writing and speaking and drew the conclusion that task complexity interacts with modality and may yield different effects on L2 performances in the two modes. She employed narrative picture description tasks by characterizing their complexity based on the foreground and background storyline complexity following Skehan’s Trade-off Hypothesis. While foreground events offer one storyline, therefore categorized as simple, background events have many storylines requiring to be connected to create a coherent story, therefore categorized as complex. 40 participants completed counterbalanced simple and complex tasks in writing. The oral data for comparison was used from the study by Tavakoli and Foster (2008). Her analyses of the effects of task complexity on syntactic complexity in the two modes revealed that more complex tasks (background events) elicited more subordination and grammatical units in both modes. However, the difference did not reach statistical significance for writing. Syntactic complexity in complex tasks was significantly higher only in speaking as reported by Tavakoli and Foster (2008). This suggests that although task complexity elicits higher mean differences in syntactic complexity in both modes, the mode is a factor for the level of this strength. Although her syntactic complexity results were different from Kormos and Trebits (2012), the conclusion was similar. “This implies that task complexity does not operate in isolation. Rather, task complexity interacts with task design, learner factors, and the linguistic modality in which the task is performed” (Tavakoli, 2014, p. 232). However, same as Kuiken and Vedder (2011), Tavakoli (2014) did not analyze the difference between writing and speaking and could not make a statistical conclusion on the interaction effect because her design of the study did not allow for conducting within-participant analyses between speaking and writing performances.

Zalbidea (2017) investigated the effects of task complexity and task modality on L2 performance and the relationship of working memory capacity with L2 performance in simple and complex tasks. 32 L2 Spanish learners completed more, and less complex argumentative tasks adapted from Kuiken and Vedder (2011) in writing and speaking. Comparison of oral and written data in terms of modality came from different participants by applying between-subjects design as they were assigned to two groups: speaking and writing for task completion. Task complexity was operationalized as +/- few elements and +/- reasoning demands. Global complexity and accuracy measures along with task-specific linguistic measures were employed (See Table 3 for details). AS-units were used for speech and T-units were chosen for writing. Her analyses revealed that more cognitively complex tasks resulted in higher syntactic complexity and accuracy in both modes, but the task complexity effects on these performances did not reach the level of significance indicating that these differences in language production were minimal upon the control for modality (Zalbidea, 2017). While a marginal role was found for task complexity, her findings showed that task modality had a substantial role in promoting more accurate and syntactically and lexically more complex language production than task complexity. Comparing performances across modalities revealed that lexical complexity and accuracy were found to be mode sensitive and to be higher in writing. Speaking performance yielded syntactically more complex language overall. Zalbidea (2017) arrived at a similar conclusion as Kormos and Trebits (2012) and Tavakoli (2014) and supported Kormos' (2014) argument that modality might best be evaluated as an element of task complexity because of the potential differences in the opportunities of allocating resources and availability of time in the two modes.

Vasylets et al. (2017) also conducted a study investigating the effects of mode and task complexity on L2 performance responding to a call for more mode-sensitive TBLT research by Gilabert et al. (2016). They employed a research design where task complexity was analyzed as a within-subjects factor and modality as a between-subjects factor. While the oral data came from Gilabert (2007), the written data was collected specifically for the study. Task complexity was operationalized by

+/- reasoning demands. The same tasks in Gilabert (2007) were used to collect the written data. The analyses of data included linguistic and propositional complexity, accuracy, and time on task (See Table 3 for details). The comparison of L2 performances in the two modes revealed higher linguistic complexity for the mean length of units in the written mode. Lexical complexity measured by D-value was also found to be significantly higher in writing. The effect of modality on accuracy was found to be statistically insignificant. The analyses of task complexity effects revealed that more demanding tasks led to significantly more linguistic complexity for the mean length of units and more sophisticated words measured by lexical frequency profile than simple tasks. Although the written mode revealed significantly higher results for certain areas of L2 performance than the oral mode, the interaction between task complexity and modality was found to be statistically insignificant for accuracy, and syntactic and lexical complexity measures. As the only study reporting interactional analysis on task complexity and modality, Vasylets et al. (2017) found a statistically significant interaction between task complexity and modality only for the ratio of extended ideas and time on task. Vasylets et al. (2017) conclude that task demands lead to a better L2 performance in certain aspects in both modes which validates task complexity as a crucial task variable. They also conclude that task complexity interacts with task modality showing different effects on L2 production on the ratio of extended ideas.

Table 3. Studies of Task Complexity Effects in Oral and Written Modes

Study	Participants				Task	Conditions	Measures	Findings
	N	Age	L1	L2				
Kuiken & Vedder (2011)	91 in writing 44 in speaking	not given	Dutch	L2 Italian (high-low proficiency	argumentative task with +/- few elements	oral and written as between group conditions both simple and complex tasks were completed in the assigned conditions	lexical variety: -alternative TTR syntactic complexity: -clauses per T-unit or AS-unit -dependent clause ratio accuracy: -total number of errors per T-unit/AS-unit -division of three degrees of errors	-accuracy higher in both modes in the complex tasks, but no effects task complexity on lexical complexity -lower syntactic complexity in the complex task only in the speaking mode conclusion: no interaction between task complexity and modality task complexity effects are not constrained by mode

Study	Participants				Task	Conditions	Measures	Findings
	N	Age	L1	L2				
Kormos & Trebits (2012)	44	16-17	Hungarian	English	picture narration: 6 pictures with a coherent storyline (simple) 6 unrelated pictures with no storyline (complex)	task type and modality as a within-participants factor language aptitude as a between-participants factor	lexical variety: D-value syntactic complexity: -clause length -ratio of subordinate clauses -ratio of relative clauses accuracy: -ratio of error-free clauses -ratio of error-free relative clauses -ratio of error-free verbs -ratio of error-free past-tense verbs	-higher lexical variety in writing in both complex and simple tasks -higher accuracy overall in writing than speaking in simple tasks -more syntactic complexity in writing in the complex task conclusion: different effects of task complexity in speech and writing

Study	Participants				Task	Conditions	Measures	Findings
	N	Age	L1	L2				
Tavakoli (2014) Oral data from Tavakoli & Foster (2008)	40 in writing 60 in speaking	18-60	varied	English	picture narration: foreground events (simple) background events (complex)	Tavakoli (2014) 40 participants in writing completing complex and simple tasks Foster (2008) 60 participants in speaking completing complex and simple tasks	syntactic complexity: - mean length of T-unit -ratio of clauses to T-unit	-higher syntactic complexity in speaking complex -no effect of task complexity in writing conclusion: different effects of task complexity in speech and writing
Zalbidea (2017)	16 in speaking 16 in writing	19.6 years mean age	English	Spanish	adapted from Kuiken & Vedder (2011) +/- few elements +/- reasoning demands	Oral and written as between group conditions Participants were divided into oral and written modes completing both simple and complex tasks in their assigned conditions	syntactic complexity: -mean number of words per unit -the ratio of the number of dependent clauses to units lexical complexity: -Guiraud's index accuracy: -error categories	-higher lexical complexity and accuracy in writing -higher syntactic complexity in speaking - no main effects of task complexity conclusion: a more significant role of task modality for accuracy and L2 complexity

Study	Participants				Task	Conditions	Measures	Findings
	N	Age	L1	L2				
Vasylets et al. (2017) Oral data from Gilabert (2007)	39 in writing 39 in speaking	18-40	Spanish & Catalan	English	argumentative tasks +/- reasoning demands	task complexity (simple vs. complex) as a within-subject task modality (oral vs. written) as a between subject factor	syntactic complexity: -mean length of AS-units (MLU) -subordination measured by S nodes per AS-unit -the mean number of modifiers per noun phrase by Coh-metrix lexical diversity: -D-value lexical sophistication: -lexical frequency profile (LFP) accuracy: - errors per 100 words	-significant interaction was found only for the ratio of extended idea units and time on task -more complex language in writing -significant task complexity effects on LFP and MLU conclusion: task complexity as a crucial task variable leading to a better L2 performance in certain aspects in both modes. -possible interaction between task complexity and modality coming from the propositional complexity

2.12 Evaluation of task complexity studies in oral and written mode

The previous studies yielded mixed findings on the effects of task complexity on language production in the two modes (See Table 3). As far as syntactic complexity is concerned, different effects of task complexity in the two modes were found in some studies. Kuiken and Vedder (2011) revealed a different effect of task complexity in the two modes on dependent clauses. While task complexity did not show any significant differences in written mode, oral mode revealed significantly higher dependent clauses in simple tasks than in complex tasks. They also reported their observation of mean differences between writing and speaking indicating performance differences in the two modes. Syntactic complexity was observed to be higher in writing. Similarly, Tavakoli (2014) revealed a different effect of task complexity in two modes on syntactic complexity. While task complexity did not have any statistically significant effect on syntactic complexity in writing, it revealed a significant effect in speaking. She reported higher syntactic complexity in the complex tasks in speaking while no significant difference was found between simple and complex tasks in writing. Finally, Kormos and Trebits (2012) also found that task complexity affected syntactic complexity differently in the two modes. However, contrary to others, they found writing to be more sensitive to task demands reporting no changes in syntactic complexity in simple and complex tasks in the oral mode while revealing significantly more syntactic complexity in the complex writing task. However, their comparison of writing and speaking did not reveal any significant differences. Vasylets et al. (2017) and Zalbidea (2017) revealed a marginal role of task complexity on L2 performances while concluding a substantial role of task modality. The effects of task complexity on syntactic complexity were found to be statistically insignificant (Vasylets et al., 2017; Zalbidea, 2017).

However, it is important to note that they did not conduct a paired-sample *t*-test as in the previously mentioned studies. Therefore, the results were for the main effects. While Zalbidea (2017) found higher syntactic complexity in speaking, Vasylets et al. (2017) reported higher syntactic complexity in writing. Studies indicate that the mode in which a task is performed affects the syntactic complexity. However, predicting the directionality of this effect becomes challenging due to the mixed results. Similarly, the effects of task complexity on syntactic complexity yielded mixed results making it difficult to predict any directionality.

The previous studies indicated a pattern of effects of task modality and task complexity on lexical complexity. The effects of task complexity on lexical complexity were mostly found to be statistically insignificant (Kuiken & Vedder, 2011; Vasylets et al., 2017; Zalbidea, 2017). Only Kormos and Trebits (2012) reported different effects of task complexity in the two modes by revealing more lexical complexity in simple tasks than in complex tasks in speaking. When the two modes were analyzed in simple and complex tasks, lexical complexity was found to be significantly higher in writing in all studies. Only Kuiken and Vedder's (2011) report did not include the statistical significance of the effects of task modality on lexical complexity because they analyzed the task complexity only within the same mode. However, their observation based on the mean differences in the two modes revealed a higher tendency in the written mode for lexical complexity. The fact that all the studies indicated higher lexical complexity in writing and no effect of task complexity on lexical complexity suggests an insignificant interaction between task complexity and modality in terms of lexical complexity due to the possibility of a parallel increase from speaking to writing. Modality seems to be a leading variable for higher lexical complexity regardless of task complexity.

Accuracy measures showed an overall pattern regarding task complexity. Vasylets et al. (2017) and Zalbidea (2017) reported no significant main effects for task complexity on accuracy. Similarly, Kormos and Trebits (2012) reported that task complexity did not have a statistically significant effect on accuracy within the same mode. Only Kuiken and Vedder (2011) found a significant effect of task complexity on accuracy both in the written and oral modes. They revealed that complex tasks were significantly more accurate than simple tasks. They also concluded that task complexity is the main factor for accuracy as participants produced fewer errors in the complex task than in the simple task in both modes. In terms of the effects of task modality, accuracy performances seem to be affected differently in writing and in speaking although the direction of this effect is rather mixed. All the studies indicated a different effect of writing and speaking on accuracy, except for Vasylets et al. (2017). Zalbidea (2017) reported higher accuracy in writing than in speaking. Kormos and Trebits (2012) also found higher accuracy in writing but only in simple tasks. Their analysis of complex tasks did not reveal any significant effect of modality on accuracy. A different effect of the two modes was also observed in Kuiken and Vedder (2011) but in speaking. They reported higher accuracy mean difference in speaking than in writing.

Variations in results among previous studies can be explained by the differences in the operationalization of task complexity, tasks, and procedures such as using data from different years and failing to counterbalance modality, and different measures. Keeping all these important methodological differences in mind, the previous studies suggest a possible interaction between task complexity and task modality based on the different effects of task complexity in the two modes. However, only one study reports a statistical analysis for the hypothesis of an interaction. Vasylets et al. (2017) did not find any statistically significant interaction between task complexity and modality in terms of accuracy and syntactic and lexical complexity. Although previous studies suggested a possible interaction between task complexity and modality based on the different effects of task complexity and modality on L2 performance, it is difficult to infer an interaction from mere comparisons of task complexity in the two modes because the interaction is affected by the magnitude of the mean differences between the independent variables and the nature of these differences such as whether they occur in a parallel fashion. Therefore, different effects of task complexity in modality need to be examined more closely to infer an interaction.

Considering relevant studies covered in the previous sections, the current study investigates the following research questions:

i. How does task modality affect L2 complexity and accuracy?

ii. How does task complexity affect L2 complexity and accuracy?

iii. How do the effects of task complexity and task modality on L2 complexity and accuracy interact with each other?

Corresponding hypotheses were formed based on previous studies and underlying theories.

i. Combining the previously outlined models of speech and writing (Hayes, 1996, 2012; Kellog, 1996; Kormos, 2006; Levelt,1989) with the previous studies (Bulté & Housen, 2009, as cited in Vasylets et al., 2019; Ellis & Yuan, 2005; Granfeldt, 2008; Kormos & Trebits, 2012; Kormos, 2014; Vasylets et al., 2017; Zalbidea, 2017), lexical complexity is hypothesized to be higher in writing than in speaking. Similarly, accuracy is hypothesized to be higher in writing. Given the mixed results for the effects of task modality on syntactic complexity, only different effects of task modality are hypothesized on syntactic complexity.

ii. Considering the mixed results and a few studies conducted on task complexity and modality (Kuiken & Vedder, 2011; Kormos & Trebits, 2012; Tavakoli, 2014; Vasylets et al., 2017; Zalbidea, 2017), it becomes difficult to build hypotheses for task complexity. Therefore, the null hypothesis is assumed for syntactic complexity and accuracy between simple and complex tasks. However, a pattern for lexical complexity emerges between simple and complex tasks. The null hypothesis is again assumed for lexical complexity but based on the statistically insignificant findings of the effects of task complexity on lexical complexity in most studies (Kuiken & Vedder, 2011; Kormos & Trebits, 2012; Vasylets et al., 2017; Zalbidea, 2017).

iii. The review of relevant studies showed that only one study reported statistical analysis on the interaction of task complexity and modality and found the interaction to be statistically insignificant for accuracy, and syntactic and lexical complexity (Vasylets et al., 2017). Other studies concluded a possible interaction based on their comparison analyses (Kormos & Trebits, 2012; Tavakoli, 2014; Zalbidea, 2017) except for one (Kuiken & Vedder, 2011). It is hypothesized to detect an interaction between task complexity and task modality for accuracy and syntactic complexity based on the conclusions from previously mentioned studies. However, lexical complexity findings indicate that writing as a task modality is the leading variable in simple and complex tasks and that the effects of task complexity are statistically insignificant in the two modes (Kormos & Trebits, 2012; Kuiken & Vedder, 2011; Vasylets et al., 2017; Zalbidea, 2017). This suggests a parallel decrease in lexical complexity from writing to speaking. Therefore, the null hypothesis is assumed for lexical complexity interaction.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter describes the data collection and analysis procedures. It also provides background for the participants, research design, materials, and the tools used in analyzing learner performance.

3.2 Participants

Participants consisted of 7th-grade students from a private school in İstanbul. All the 7th-grade students (*N* = 174) in the school were informed about the study. The parents of 60 students submitted the consent form. Out of 60 participants, 52 of them completed the tasks in all conditions. 8 participants failed to participate in the tasks due to absence on the days of data collection stemming from health issues. Additionally, some of them wanted to stop being part of the study because of having high anxiety during the speaking tasks. The primary school involved in this study reports that their objective is for their students to achieve a B1+ level on the Common European Framework of Reference for Languages (CEFR) at the end of middle school, and to serve this purpose, their curriculum includes communicative activities catering towards four skills as writing, speaking, listening, and reading. The school also offers clubs that the medium of language is English where students are engaged in their interests.

The participants (*N* = 52) were randomly assigned to two orders of modality in the study: writing-speaking (*N* = 23) and speaking-writing (*N* = 29). The data collection occurred while students were studying their regular curriculum.

3.3 Research design

This study has a 2x2 factorial within-subjects design as task complexity (simple vs. complex) and modality (writing vs. speaking) within-subjects factors. The order of modality (oral-written vs. written-oral) was designed as a between-subjects factor to examine the effects of task modality (refer to Table 4 for the design).

Table 4. The Design of the Study

	Order of Modes	
	Condition 1 Writing-Speaking	Condition 2 Speaking-Writing
Task Complexity & Modes (Within-Participants)	Simple	Simple
	Complex	Complex

Tasks given in the study were counterbalanced in each condition. Having two conditions speaking-writing and writing-speaking provided counterbalancing for modality. Participants completed the first task in writing and the second task in speaking for the first condition. The second condition required participants to complete the first task in speaking and the second task in writing.

3.4 Materials

Each participant completed four tasks: two counterbalanced simple tasks in speaking and writing, and two counterbalanced complex tasks in speaking and writing. All participants completed the tasks according to the condition they were assigned to which means that some of them started with writing first and others with speaking.

Tasks in this study required participants to narrate a story based on a series of pictures (See Appendix A). Picture narration tasks in the simple task condition have an already constructed storyline with 6 pictures. One of the sets of pictures for the simple task condition was taken from Heaton (1966, 1975) and the other set was used in other complexity and modality studies before (Ellis, 1987; Ellis & Yuan, 2005; Tuzcu, 2019; Tuzcu & Yalçın, 2020). The set of pictures by Heaton (1966, 1975) involves a story of a boy and a girl going on a picnic. They realize at the end that their dog was in the basket eating all the food (*Dog and Basket* task). The second set of pictures from the simple task condition presents a story of a boy having several boxes and being chased by a man. First, the boy gets spooked by the man but in the end, realizes that the man just wants to give him the box that he dropped earlier (*Lost and Found* task). In order to guarantee that simple tasks are comparable, similar code complexity, storyline complexity, and task structure were established between simple tasks.

Both sets of pictures in simple tasks had similar code complexity in a way that both required easy key vocabularies such as girl, boy, mother, dog, picnic, and basket in the picture set for the first simple task, and boy, box, bus, catch, and run in the picture for the counterbalanced simple task. These two tasks also matched in terms of storyline complexity. Tavakoli and Foster (2008) argued that storylines have foreground and background information. Foreground information is related to the central propositions that contribute to the development of the theme while background information is about elaborating on the foreground information. Both tasks in the simple task condition have foreground information along with some background events such as a dog secretly entering the basket in one of them and the boy dropping his box in the background and therefore being chased in the other task.

Tavakoli and Foster (2008) mention connecting background events to the foreground main storyline promotes using subordinate clauses, therefore it is important to match storyline complexity between tasks to avoid the promotion of different linguistic elements. Another issue to establish between the simple tasks was structure. In both sets of pictures in the simple task condition, tasks were sequenced. Whether the story has a sequence presented to the audience affects the processing burden of telling a story in an L2 (Tavakoli & Foster, 2008). The same number of pictures, similar code complexity, storyline complexity, and structure were provided for both simple tasks to make sure they are comparable.

Sets of pictures for the complex tasks included random pictures without any structured storyline. The sets of pictures for complex tasks were adopted following Kormos (2011b), Kormos and Trebits (2011), and Kormos and Trebits (2012). Tasks in this condition involved six unrelated pictures and all the pictures were asked to be included in the story. One set of pictures includes a house, lightning, a ring, mountains, a plane, and a locked door. The other set of pictures consists of an island, a book, big waves, an open door, a boat, and a street. In order to guarantee that complex tasks are comparable, the same number of unrelated pictures and the presentation of similar elements were provided.

Both sets of pictures include similar elements to balance language production between the two versions. In both versions, an object was presented (book in one task and ring in the other), a picture showing bad weather was included in both versions (lightning in one task and stormy sea in the other), a transport was present in both versions (airplane in one and boat in the other one), a geographical location was present (mountains in one task and island in the other), a house was presented in both versions (a house in the forest in one task and houses in the street in the other) and finally a door was included (locked in one version and open in the other) (Kormos & Trebits, 2012). Successful completion of these two tasks required participants to both rely on their linguistic knowledge and relate the pictures to one another to create a story. Kormos and Trebits (2012) argue that this poses greater demands on conceptualization compared to the simple task condition. Tasks sequenced with a coherent storyline in the simple condition do not require conceptualization of the plot and therefore do not pose a high cognitive load for the participants (Kormos, 2011b).

3.5 Procedure

Data collection was completed in several sessions over the course of two months. Data collection started on April 19 and ended on June 9. Each condition started with their tasks in their assigned order of mode (i.e., writing-speaking or speaking-writing). After finishing the first task, 7 to 10 days later, each participant completed the task in the alternative mode. Tasks were counterbalanced for all conditions. For the speaking task, prior to the administration of the tasks, participants were informed that they will be recorded. Table 5 provides the timeline of the study.

Participants were given three minutes to plan after seeing the pictures in both speaking and writing tasks. The pictures were available to the participants during the entire task completion time in both modes. They were given one lesson period (40 minutes) to complete the tasks for both modes although participants did not need 40 minutes to finish their speaking tasks. Most participants finished in a 5 to 6 minutes time range including the planning time. Regardless, the same amount of time for planning and completing the task was allowed for both modes to eliminate any linguistic differences in the output that might result from imposing different planning and completion time. Writing tasks were given to participants during their English lessons by their English teachers. English teachers administering the writing tasks were informed by the researcher regarding the instructions. A written instruction letter was given to teachers to be read to the participants (See Appendix B). Speaking was conducted by the researcher and a colleague who is currently an MA student in English Language Education. She was also informed about the study and given instructions to be read to the participants (See Appendix B). For speaking, participants were taken from their English lessons one by one, and the tasks were administered and recorded in an empty classroom. All instructions were presented in English because the private school cooperating with this study wanted to keep its English-only policy.

Table 5. Timeline of the Study

	Order of Modes	
	Condition 1	Condition 2
Task Complexity & Modality (Within-Participants)	Writing Simple 1st Task April 19 - Speaking Simple 2nd Task April 26/27/28	Speaking Simple 1st Task May 17/18/19 - Writing Simple 2nd Task May 24/26/27
	Writing Complex 1st Task May 10 - Speaking complex 2nd Task May 17/18/19	Speaking Complex 1st Task May 30/31, June 1 - Writing Complex 2nd Task June 7/8/9

3.6 Operationalization of task complexity

In this study, related pictures with a coherent storyline are considered cognitively simple tasks while unrelated pictures with no sequence and storyline are considered cognitively complex tasks. This study takes Levelt's (1989) speech model to explain what is cognitively complex and simple. Cognitive demands posed by tasks in the conceptualization stage can be considered as a part of task complexity (Kormos & Trebits, 2012). According to Levelt (1989), the conceptualization stage involves the planning of what to say and the content of the speech. While the process of conceptualizing a message in L1 needs attention, the other stages which are formulation and articulation work smoothly (Kormos, 2006). However, in L2 or non-balanced bilinguals, formulation and articulation stages might require attention along with the conceptualization stage. The need for conscious attention disrupts the parallel processing between the stages (Kormos, 2006). Based on the cognitive demands a task requires in the conceptualization stage, this study identifies the tasks with a set of related pictures and a storyline sequence as cognitively simple because participants do not need to give as much attention to the content of the story and are not required to be heavily engaged in selecting and relating relevant concepts and ordering them during conceptualization. This results in easing the cognitive demands in the conceptualization stage (Kormos & Trebits, 2012; Skehan, 2009). On the other hand, the task with a set of unrelated pictures and no storyline sequence is considered cognitively complex because of placing higher cognitive demands in the conceptualization stage by requiring participants to find and relate relevant concepts and put them in the order of a coherent storyline.

Having less or more cognitive demands can affect the formulation stage and how learners manage attentional resources. Limited Attentional Capacity Model suggests that humans have limited attention and memory capacity (Skehan, 2009). Therefore, increasing task complexity decreases the availability of attention and memory resources. Furthermore, cognitively more demanding tasks place attention on the content of the message while taking it away from linguistic forms (Skehan, 2009). Skehan (1998) divides cognitive complexity into two subdimension: cognitive processing and cognitive familiarity. Manipulation of tasks in this study can fall under cognitive processing since cognitive processing is related to online processing demands such as organizing information, and online computation (Skehan, 1998). The task complexity is manipulated and increased through posing more online processing, organizational, and information demands in unrelated pictures while the complexity is decreased through demanding less online processing, organizing information, and planning in related pictures with a storyline and a sequence.

When the Cognition Hypothesis is concerned, the manipulation of task complexity in the current study can be related to the resource-directing dimension, specifically in causal reasoning (Robinson, 2001a, 2001b, 2011a). Related picture sequences and a set of unrelated pictures with no storyline pose different levels of reasoning. A given storyline through pictures can be considered to require a decreased level of reasoning since it mainly involves transforming given information (Kormos, 2014; Tavakoli, 2014). A set of unrelated pictures with no storyline, on the other hand, can be considered to require more complex reasoning since it involves reasoning about forming and linking random events and situations in a way that constructs a coherent storyline.

3.7 Measures

Both oral and written transcriptions were analyzed in terms of accuracy, and syntactic and lexical complexity. The following sections provide the accounts for the transcription of the data and the measures for accuracy and syntactic and lexical complexity.

3.7.1 Oral data units of analysis

Transcription and coding of written data are less controversial as it is more organized and punctuated and offers more clear-cut boundaries than spoken data. Therefore, carefully chosen transcription and unit coding was important for the spoken data. When studies concerning the analysis of speech production were investigated, many of them did not provide a definition of a unit they used and the remaining did not give enough detail or a sample of coding in their studies (Foster, Tonkyn, & Wigglesworth, 2000). Following Foster et al. (2000), this study used Analysis of Speech units (AS-unit) to transcribe oral data and divide it into units. AS-units were chosen following the argument that coding oral data into AS-units is more reliable than using T-units because it takes intonations, pauses, unique syntactic characteristics of oral data, and dysfluency markers into consideration and offers guidelines about how to handle them (Foster et al., 2000). For example, the coding of dependent and independent clauses can show variations in oral data depending on the units chosen. AS-units offer sensitive coding of subordinate conjunctions such as because. Such a conjunction can function as an ellipted version of an independent clause by carrying the meaning of *I say this because…* in speech and needs to be coded as such. However, T-units do not offer such sensitive coding and these ellipted versions are rendered as dependent clauses. The definition of AS-unit was taken from Foster et al. (2000) as an independent clause together with any subordinate clauses associated with it. An independent clause includes a finite verb. A subordinate clause can include a finite verb or a non-finite verb and a minimum of one other clausal element. AS-units for oral data analysis have been widely used in other modality studies before (Ferrari & Nuzzo 2009, as cited in Vasylets et al., 2019; Kuiken & Vedder, 2011; Zalbidea, 2017). Following Foster et al. (2000), a sample AS-unit coding is provided for the oral data (See Appendix C).

Oral data were transcribed and coded by the researcher. 15% of the oral data were coded by a native speaker of English, independently to establish reliability. Inter-coder reliability of 15% of oral data was calculated for the accuracy measures and syntactic complexity measures. The second coder received training from the researcher regarding the coding of the data. A coding guideline was given to the second coder (See Appendix D).

3.7.2 Written data units of analysis

A different unit of analysis was used in the written mode. T-unit was chosen for the analysis of written samples. While AS-unit is considered more appropriate for oral data, t-unit is considered to be more appropriate for written data which mostly consists of full clauses and sentences (Norris & Ortega, 2009). When modality studies are reviewed, it is common to choose different and appropriate units of analysis for the oral and the written data (Ferrari & Nuzzo, 2009, as cited in Vasylets et al., 2019; Kuiken & Vedder, 2011; Vasylets et al., 2019; Zalbidea, 2017). Following this approach, written data was divided into T-units based on Hunt's (1970) definition. A T-unit refers to a minimal terminable unit that includes at least one main clause but can also have subordinate clauses, phrases, and words attached to it or embedded in it (Hunt, 1970). After written data was divided into T-units, inter-coder reliability of 15% of written data was calculated for the accuracy and syntactic complexity measures. The second coder received training and guidelines for coding T-units (See Appendix E). Inter-rater reliability was checked for accuracy and syntactic complexity measures.

3.7.3 Rationale for measures chosen

General and distinct L2 measures were employed in the current study in order to capture the multidimensional nature of L2 performance and serve the purpose of the study. Norris and Ortega (2009) state there are measures of L2 performance that are complementary to each other and should be interpreted together. Norris and Ortega (2009) further argued in their analysis of studies that there are some measures and metrics targeting the same dimension of performance and therefore they are redundant when interpreted together. Therefore, global measures of accuracy, and syntactic and lexical complexity were chosen. As Skehan and Foster (1997, 1999) stated generalized measures are more sensitive to differences in language production in different experimental conditions. Another important reason to use general measures was to provide comparability of the results of this study to other modality studies (Kormos & Trebits, 2012; Kuiken & Vedder, 2011; Tavakoli, 2014; Vasylets et al., 2017; Zalbidea, 2017).

There is a need for more fined measures of L2 performance along with the general measures. Norris and Ortega (2009) argue that measures of the mean length of a multi-clausal unit will give global linguistic complexity while missing certain elements of complexity stemming from subordination. While global syntactic complexity is measured, any increases stemming from pre- or post-modification within a phrase or the use of nominalizations are disregarded (Norris & Ortega, 2009). Therefore, measures that reveal subclausal, phrasal complexity should be considered a complementary measure to global syntactic complexity measure. In response to this call for specific measures, this study adds the ratio of clauses to AS-units or T-units to tap into subordination complexity and the mean length of a clause to reveal phrasal complexity. Having mentioned the reasons for the measures chosen in this study, the next section provides the operationalization for these measures.

3.7.4 Operationalization of the measures employed

The current study assessed learners' performances by using various syntactic and lexical complexity, and accuracy measures. Measures on the oral data were conducted based on the pruned narratives. This narrative included only the final version of all repetitions, false starts, and self-corrections. Sample AS-unit coding is provided for the procedure (See Appendix C). Additionally, pruned narratives excluded fillers such as *"You know"*, *"I guess"*. More details on the exclusion of fillers and comments from the oral data are provided in the data analysis section. The following two sections, on the other hand, explain the syntactic and lexical complexity and accuracy measures employed in this study.

3.7.4.1 Accuracy measures

Unit accuracy was measured through the proportion of error-free t-units to the total number of t-units for the written data and the proportion of error-free AS-units to the total number of AS-units for the oral data.

Clausal accuracy measure was calculated by dividing the total number of error-free clauses by the total number of clauses.

The ratio of error-free units and clauses is widely used in other complexity and modality studies as global measures of accuracy (Ellis & Yuan, 2005; Kormos & Trebits, 2012; Kormos, 2014; Kuiken & Vedder, 2007, Tuzcu & Yalçın, 2020).

3.7.4.2 Syntactic complexity

This study used three different syntactic complexity measures that were manually calculated to reveal overall complexity, subordination complexity, and phrasal complexity. These measures can be found in other complexity and modality studies (Ellis & Yuan, 2005; Ferrari & Nuzzo, 2009, as cited in Vasylets et al., 2019; Gilabert, 2007; Kormos & Trebits, 2012; Kormos, 2014; Kuiken & Vedder; 2007; Kuiken & Vedder, 2011; Tavakoli, 2014; Tuzcu & Yalçın, 2020; Vasylets et al., 2017; Zalbidea, 2017).

Overall complexity was manually calculated by finding the mean number of words per AS-unit for oral data and per T-unit for written data. This measure has been used in modality studies before to reflect the global syntactic complexity (Tavakoli, 2014; Tuzcu & Yalçın, 2020; Vasylets et al., 2017; Zalbidea, 2017).

Subordination complexity was calculated by the ratio of clauses to AS-units or T-units. This measure has also been widely used in other complexity and modality studies (Kormos & Trebits, 2012; Kuiken & Vedder, 2007; Kuiken & Vedder, 2011; Tavakoli, 2014; Tuzcu & Yalçın, 2020; Zalbidea, 2017).

Phrasal complexity was calculated by the mean length of a clause followed by Norris and Ortega (2009) arguments.

3.7.4.3 Lexical complexity

This study calculated lexical complexity through *_lognostics D_Tools* program available at: http://www.lognostics.co.uk/tools/D_Tools/D_Tools.htm (Meara & Miralpeix, 2016). Table 6 presents the summary of measures and tools and Appendix F provides explanations for the manual measures used in this study.

Table 6. Measures in the Study

Tools	Measures	Definition
Manual	Unit Accuracy	the proportion of error-free T-units or AS-units to the total number of T-units or AS-units, dividing the total number of error-free units by the total number of units
Manual	Clausal Accuracy	The ratio of error-free clauses, dividing the total number of error-free clauses to total number of clauses
Manual	Overall Syntactic Complexity	Mean length of AS-unit or T-unit, dividing the number of words by the number of AS-units or T-units
Manual	Subordination Complexity	the ratio of clauses to AS-units or T-units, diving the total number of clauses to the total number AS-units of T-units
Manual	Phrasal Complexity	Mean length of clause, dividing the number of words by the number of clauses
_lognostics D_Tools Program	Lexical Complexity	lexical diversity (LD) measured by *D* calculations

Lexical complexity has become part of complexity studies as Skehan (2009) argues that complexity, accuracy, and fluency are important measures of L2 performance, but they need to be supported by lexical measures. Type-token ratio (TTR) is widely known and used as a lexical complexity measure, but Skehan (2009) argues that this measure is strongly related to text length and therefore needs to be corrected. With the aim of decreasing the effects of text length, this study employs lexical diversity (LD) measure calculated by *D* calculations in _lognostics D_Tools.

Malvern and Richards (1997) created a D-formula also known as *vocd* to offer a more reliable and less text-length-base lexical diversity measure. Adopting this computation, _lognostics D_Tools program calculates LD through type-token ratio samplings and curve fittings (Meara & Miralpeix, 2016). Lexical measures can be distinguished as text-internal measures which take text itself as enough for the calculation and as text-external measures which ask for a reference such as a word frequency (Skehan, 2009). While the type-token ratio would belong to the former one and is strongly affected by text length, the *D* calculation offers an alternative measure to TTR and is less affected by the text length. It is based on a mathematical probabilistic model where it takes 100 random samples of 35 tokens and repeats the sampling procedure for a sample of 36 tokens, 37 tokens until 50 tokens and calculates the mean TTR for each sample (Malvern & Richards, 1997; McCarthy & Jarvis, 2007; Meara & Miralpeix, 2016). Malvern and Richards (1997) accept the premise that different text lengths result in different TTR values and therefore calculate a set of TTRs from text samples of different sizes and give the best-fitting curve as the value. To do so, D calculations generate 100 samples. Each sample has words ranging from 35 to 50 and then D calculates the mean TTR for each of these 100 samples (Meara & Miralpeix, 2016). After having the plot for 35-50 random samples of tokens, _lognostics D_Tool program uses D coefficient formula to generate a theoretical curve so that it can find the best fitting random-sampling TTR curve to the theoretical curve. Therefore, the program needs a text length of a minimum of 50 words to calculate samples of up to 50 words. Some participants (*N* = 9) in the current study failed to produce 50 words in their narratives. These participants' lexical complexity measures were left as missing data. Since *D* calculation is less sensitive to the text length than the type-token ratio, the current study employs _lognostics D_Tools program for measuring lexical complexity.

3.8 Data analysis

The data was transcribed by the researcher. Oral data revealed that some participants made comments on the task itself, their thoughts, or the story. Off-task utterances were removed from the data analysis. Additionally, oral data revealed that participants used formulaic phrases such as "*I mean*" or "*You know*" as fillers. These phrases were excluded from the pruned narratives.

The data was coded by the researcher and a second coder who is a native speaker of English, independently. The second coder received a training session provided by the researcher and information regarding how to code the data (See Appendix D and Appendix E). Two coders did not include phrases that function as fillers in the analyses. An inter-coder reliability analysis included syntactic complexity and accuracy measures and was conducted for 30% of the data. Pearson Correlation reached more than 95% reliability for all measures (See Appendix G for the list of reliability scores). After the reliability was met, the data was entered into SPSS version 25.0.

Before starting any analyses, the data were checked for the assumptions by referring to skewness and kurtosis values, normality tests, histograms, Q-Q plots, and box plots. The results of checking for the assumptions revealed some outliers and non-normal distributions in some variables. There are several ways to deal with non-normal data. One way is to use non-parametric tests (Field, 2009). Non-parametric tests conduct analyses based on ranking the data to circumvent the distributional assumptions of parametric tests. This procedure results in losing some information regarding the size of differences between the data and can lead to less powerful results than their parametric counterparts (Field, 2009). A decrease in statistical power can lead to a higher possibility of a Type II error but only when non-parametric tests are conducted on normally distributed data. Then non-parametric tests have an increased chance of leading to a Type II error (Field, 2009). In this study, only some variables showed non-normal distributions. Therefore, adopting non-parametric tests for the current study would mean that some non-parametric tests would be conducted between non-normal and normal distributions which can lead to a Type II error.

Another way to deal with non-normal data is data transformation. Transforming data does not create any statistical consequences as it does not affect the relationships between variables. However, it changes the differences between variables because transformed data have different units of measurement than untransformed data. Therefore, other variables need to be transformed as well to ensure the same unit of measurement (Field, 2009). This suggests that the variables that are normally distributed in this study need to be transformed as well to be comparable to non-normal transformed data. Transforming normally distributed data does not seem to be the best option and further, transformation results in "addressing a different construct to the one originally measured, and this has obvious implications for interpreting that data" (Field, 2009, p. 156). Transformation leads to changes in the scale and properties of the data. The issue particular to this study is that transformation may not maintain the within-subject structure of the data. Franz and Loftus (2012) discuss this problem by referring to circularity in repeated measures design. Franz and Loftus (2012) define circularity as the homogeneity of variance (used for between-subjects design) of repeated measures design. They argue that when pairwise comparisons show small variability between some conditions (e.g., levels A and B and levels C and D) but large variability is observed in other conditions (e.g., levels B and C), circularity can be violated. The normalization of data to fix such a violation can result in propagating the large variability (e.g., between B and C) to other conditions (e.g., between levels A and D) and therefore, creating the impression that the violation of circularity is fixed (Franz & Loftus, 2012). Since the entire data in this study are based on the measures of the same individuals over time, the transformation of data creates concerns in terms of reaching incorrect inferences and violating the within-subject structure of the data.

The statistical test employed for data analysis can be another way to deal with non-normal data. The robustness of a statistical test becomes a critical issue. Therefore, this study follows the argument that "the question of whether to transform is linked to this issue of robustness (which in turn is linked to what test you are performing on your data)" (Field, 2009, p. 155). The current study uses repeated measures ANOVA which is considered to be robust (Field, 2009). Early findings on this issue showed that slightly skewed and moderately skewed distributions yielded very small deviations from the theoretical power, so the transformation of data might not be needed (Games & Lucas, 1966). Another study reviewing the power of F regarding normal and non-normal data showed that even if data are skewed and non-normal, F controls the Type I error (Glass, Peckham & Sanders, 1972). However, there seems to be an oversimplification of the F test being accurate on non-normal data. Therefore, details in Lunney's study (1970) become important to justify the use of repeated measures ANOVA on non-normal data. Lunney (1970) showed that ANOVA yielded accurate results when the group sizes were equal and contained at least 20% degrees of freedom. Additionally, the smallest response category needed to consist of at least 20% of all responses (Lunney, 1970). The current study has 52 participants (*N* = 52) completing all four tasks and equal group sizes for each dependent measure. However, there are 9 missing values for lexical complexity because these participants produced less than 50 words, therefore could not be analyzed for lexical complexity. This means that the smallest response category consists of a group size of 43. This equals 21% of all responses. "This evidence suggests that *when group sizes are equal* the *F*-statistic can be quite robust to violations of normality" (Field, 2009, p. 360). Following these arguments, repeated measures ANOVA was used without transforming the data. However, to be certain about the effects of non-normal data, the non-parametric equivalent of repeated measures ANOVA, the Friedman test, was conducted. Non-parametric tests matched the results of the parametric test by finding similar significant or insignificant differences in within-subject variables. In addition, some preliminary analyses were conducted with and without the outliers, and no great differences were found. Therefore, outliers were kept in the analyses because they came from real data and were not errors. As a result, the data was examined by conducting a series of 2x2 factorial repeated measures ANOVAs for research questions 1, 2, and 3. For this design, two modality levels (i.e., writing and speaking) and two task complexity levels (i.e., simple and complex) were entered into repeated measures ANOVA.

The main effects and the interaction effects were obtained in ANOVA for each dependent variable. The main effects revealed the effects of modes and task complexity on dependent variables while the interaction effects showed whether there was any interaction between independent variables (task complexity and modality). In other words, it analyzed if the effects of task complexity were statistically affected by the mode of performance in which tasks were completed. Following Cohen (1988), the effect sizes for partial eta squared (η_p^2) were interpreted as .01 = small, .06 = medium, and .14 = large. Additionally, following Field's (2009) arguments, confidence intervals were provided to visualize and supplement the main effects and interactions. 95% confidence intervals become supplementary to the comparisons of the conditions as they can indicate a likelihood of a significant difference between them. A significant difference is likely to occur when there is no overlapping between the mean scores of different conditions, and the confidence interval of one condition exceeds the upper limit of the confidence interval bar of the other condition (Field, 2009). Finally, when a significant interaction was detected, following Field (2009) and Franz and Loftus (2012) further pairwise comparisons were conducted in one-way repeated measures ANOVA to supplement and interpret the nature of the interaction between independent variables.

CHAPTER 4

RESULTS

4.1 Introduction

The data collected from the same individuals over time were examined through descriptive statistics (i.e., means and standard deviations), confidence intervals, and effect sizes.

A series of 2x2 factorial repeated measures ANOVA was conducted on syntactic complexity, lexical complexity, and accuracy measures. The alpha for achieving statistical significance was set at .05. This chapter presents the results of 2x2 factorial repeated measures ANOVA along with the descriptive statistics, confidence intervals, and effect sizes.

4.2 Syntactic complexity

Syntactic complexity was analyzed in terms of overall complexity, complexity by subordination, and phrasal complexity. Table 7 presents the means and standard deviations of dependent variables for modality and task complexity.

Table 7. Descriptive Statistics for Syntactic Complexity Measures

Dependent Variables	Writing Simple Task	Writing Complex Task	Speaking Simple Task	Speaking Complex Task
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Syntactic Complexity				
Overall	8.83 (1.87)	9.09 (2.95)	9.40 (2.89)	9.04 (2.87)
Subordination	1.76 (0.36)	1.69 (0.43)	1.75 (0.48)	1.61 (0.45)
Phrasal	5.02 (0.61)	5.33 (0.96)	5.36 (0.84)	5.62 (0.70)

4.2.1 Overall syntactic complexity

Overall syntactic complexity means of task modality indicate that speaking had higher scores (*M* = 9.40, *SD* = 2.89) than writing (*M* = 8.83, *SD* = 1.87) in simple tasks. However, in complex tasks, writing (*M* = 9.09, *SD* = 2.95) had very slightly higher scores than speaking (*M* = 9.04, *SD* = 2.87).

Overall syntactic complexity means of task complexity show that complex tasks had higher scores (*M* = 9.09, *SD* = 2.95) than simple tasks (*M* = 8.83, *SD* = 1.87) in writing. However, in speaking, simple tasks (*M* = 9.40, *SD* = 2.89) had higher scores than complex tasks (*M* = 9.04, *SD* = 2.87). The 95% confidence intervals in Figure 3 demonstrate an almost overlapping pattern for both task modality and task complexity.

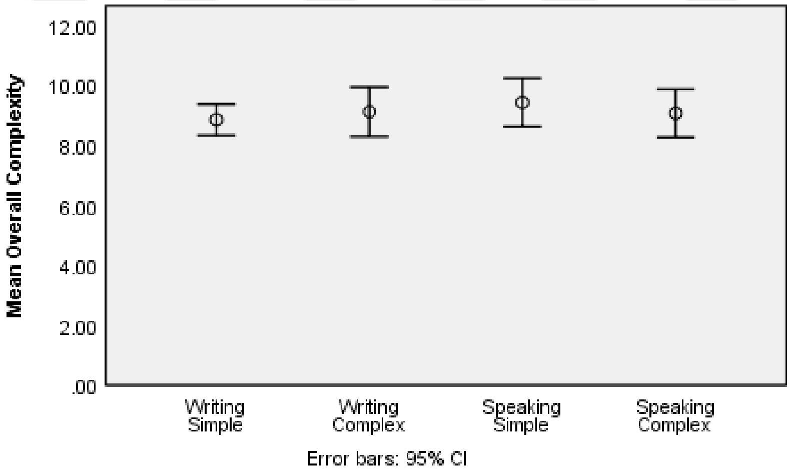


Figure 3. Overall syntactic complexity scores with 95% confidence intervals

2x2 factorial repeated measures ANOVA was conducted to investigate the statistical significance of the mean differences. The main effects of task modality on overall syntactic complexity did not indicate any statistically significant difference, $F(1,51) = .752$, $p = .390$, $\eta_p^2 = .015$, observed power = .136 (see Table 8). A small effect size also suggests that the independent variable is not a strong predictor of the dependent variable. Therefore, there was no statistically significant difference between speaking and writing in terms of overall complexity. Similarly, no significant difference was found for the main effects of task complexity with a small effect size, $F(1,51) = .020$, $p = .888$, $\eta_p^2 = .000$, observed power = .052. This suggests that simple and complex tasks did not differ significantly from each other in terms of overall complexity. Considering overall syntactic complexity is a common measure for task complexity and modality, small effect sizes for both independent variables suggest that there might be other factors that are more important predictors of the dependent variable such as proficiency levels or sample size.

Table 8. 2x2 Repeated Measures ANOVA Results for Overall Complexity

Overall Complexity	df	MS	F	P	η_p^2	Observed power
Modes	1	3.557	.752	.390	.015	.136
Error(modes)	51	4.730				
Complexity	1	.120	.020	.888	.000	.052
Error(complexity)	51	6.014				
Modes*Complexity	1	5.091	2.025	.161	.038	.287
Error (modes*complexity)	51	2.513				

The interaction effects between modality and complexity were also not statistically significant, $F(1,51) = 2.025$, $p = .161$, $\eta_p^2 = .038$, observed power = .287. This result indicates that the effect of modality (i.e., speaking and writing) on overall syntactic complexity is statistically not different in simple and complex tasks and the effect of task complexity on overall syntactic complexity is statistically not different in writing and speaking. However, it is important to point out the higher effect size for the interaction than the main effects indicating the power of the test.

4.2.2 Complexity by subordination

The visual inspection of complexity by subordination means for task modality shows that writing had very slightly higher scores (*M* = 1.76, *SD* = 0.36) than speaking (*M* = 1.75, *SD* = 0.48) in simple tasks. Similarly in complex tasks, writing had higher scores (*M* = 1.69, *SD* = 0.43) than speaking (*M* = 1.61, *SD* = 0.45).

Complexity by subordination means of task complexity show that simple tasks (*M* = 1.76, *SD* = 0.36) had higher scores than complex tasks (*M* = 1.69, *SD* = 0.43) in writing. Similarly, in speaking, simple tasks (*M* = 1.75, *SD* = 0.48) had higher scores than complex tasks (*M* = 1.61, *SD* = 0.45). The 95% confidence intervals in Figure 4 show an almost overlapping pattern for task modality, especially in simple tasks. Error bars for task complexity show that simple tasks exceed the complex tasks, although slightly, in both writing and speaking.

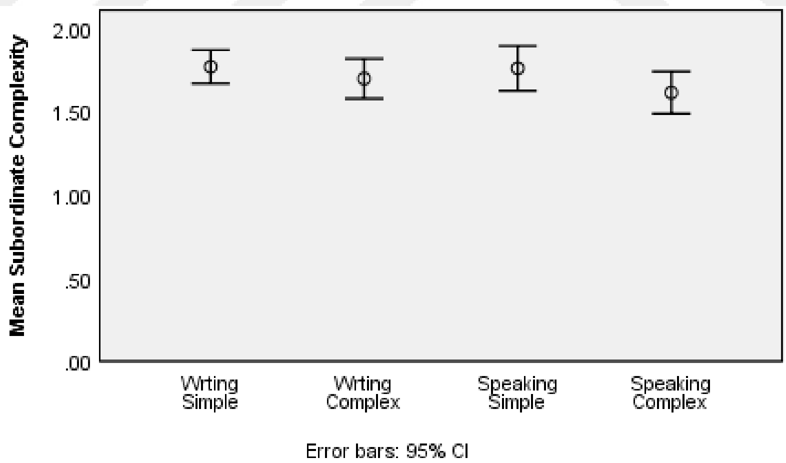


Figure 4. Subordination complexity scores with 95% confidence intervals

2x2 factorial repeated measures ANOVA was conducted on subordination scores (see Table 9). The main effects for task modality on complexity by subordination did not indicate any statistically significant difference, *F*(1,51) =

1.274, p = .264, η_p^2 = .024, observed power = .198. The small effect size can be interpreted as support for the insignificant finding. It can be concluded that no statistically significant difference was found between speaking and writing in terms of complexity by subordination. However, a statistically significant difference, although very close to the benchmark, was detected for the main effect of task complexity, $F(1,51)$ = 4.052, p = .049, η_p^2 = .074, observed power = .506. Pairwise comparisons of main effects for task complexity (see Table 10) showed that simple tasks had significantly higher complexity by subordination than complex tasks, p = .049, 95% CI [.000, .217].

The interaction between task modality and task complexity was not statistically significant, $F(1,51)$ = 1.021, p = .317, η_p^2 = .020, observed power = .168. This finding indicates that although there is a significant difference between simple and complex tasks, the difference does not significantly vary between writing and speaking.

Table 9. 2x2 Repeated Measures ANOVA Results for Subordination Complexity

Complexity by Subordination	df	MS	F	P	η_p^2	Observed power
Modes	1	.113	1.274	.264	.024	.198
Error(modes)	51	.089				
Complexity	1	.613	4.052	.049	.074	.506
Error(complexity)	51	.151				
Modes*Complexity	1	.072	1.021	.317	.020	.168
Error (modes*complexity)	51	.071				

Task complexity can be concluded as a variable leading to a better L2 performance in complexity by subordination. As a result, no statistically significant interaction was found between task modality and task complexity.

Table 10. Subordination Complexity Pairwise Comparisons of Task Complexity

(I) complexity	(J) complexity	Mean Difference (I-J)	SD	P[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Simple	Complex	.109*	.054	.049	.000	.217

*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Least Significant Difference.

It is important to refer to the interaction graph in Figure 5 to fully understand the insignificant interaction. Although there is a tendency towards lower estimated marginal means of complexity by subordination in speaking complex compared to writing complex tasks, the graph presents a parallel interaction between modality and task modality.

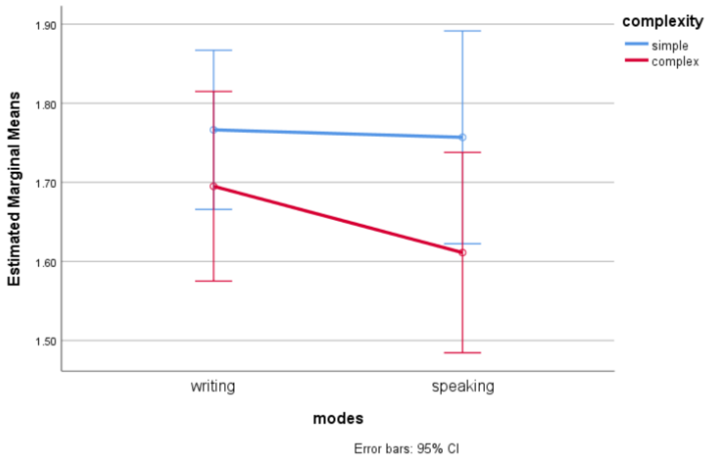


Figure 5. Modality and task complexity interaction for subordination complexity

4.2.3 Phrasal complexity

Phrasal complexity means of task modality show that speaking (*M* = 5.36, *SD* = 0.84) had higher scores than writing (*M* = 5.02, *SD* = 0.61) in simple tasks. Similar tendencies were found in complex tasks. Speaking (*M* = 5.62, *SD* = 0.70) had higher scores than writing (*M* = 5.33, *SD* = 0.96) in complex tasks.

Phrasal complexity means of task complexity showed that complex tasks (*M* = 5.33, *SD* = 0.96) had higher scores than simple tasks (*M* = 5.02, *SD* = 0.61) in writing. Similarly in speaking, complex tasks (*M* = 5.62, *SD* = 0.70) had higher scores than simple tasks (*M* = 5.36, *SD* = 0.84). The 95% confidence intervals for task modality in Figure 6 also show speaking exceeds writing in both simple and complex tasks. Error bars for task complexity show that the complex tasks exceed simple tasks in both writing and speaking. The confidence interval bars suggest a parallel increase from simple to complex and from writing to speaking.

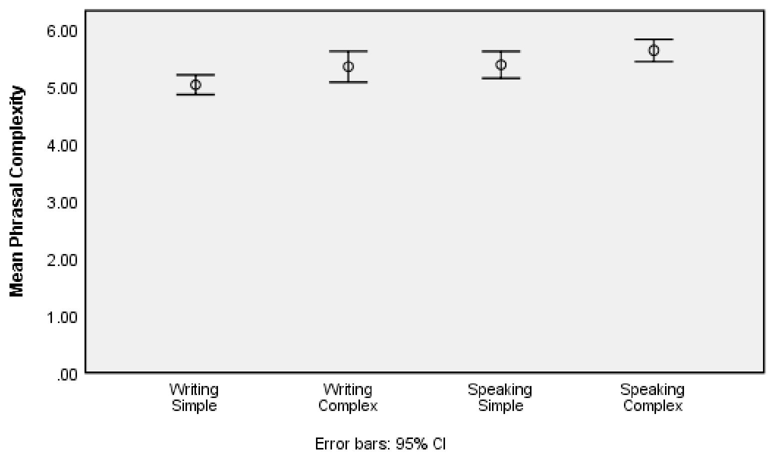


Figure 6. Phrasal complexity scores with 95% confidence intervals

2x2 factorial repeated measures ANOVA was conducted on phrasal complexity scores to further investigate the main effects and interactions (See Table 11). The main effects of task modality on phrasal complexity indicated a statistically significant difference with an almost strong effect size, $F(1,51) = 7.140$, $p = .010$, $\eta_p^2 = .123$, observed power = .746. Pairwise comparisons of modality main effects (see Table 12) revealed that participants had higher phrasal complexity in speaking than in writing, $p = .010$, 95% CI [.079, .555].

Table 11. 2x2 Repeated Measures ANOVA Results for Phrasal Complexity

Phrasal Complexity	df	MS	F	P	η_p^2	Observed power
Modes	1	5.220	7.140	.010	.123	.746
Error(modes)	51	.731				
Complexity	1	4.187	6.856	.012	.119	.729
Error(complexity)	51	.611				
Modes*Complexity	1	.055	.130	.719	.003	.065
Error (modes*complexity)	51	.419				

Table 12. Phrasal Complexity Pairwise Comparisons of Modality

(I) modes	(J) modes	Mean Difference (I-J)	SD	P[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Speaking	Writing	.317*	.119	.010	.079	.555

*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Least Significant Difference.

A statistically significant difference was also detected for the main effects of the task complexity again with an almost strong effect size, $F(1,51) = 6.856$, $p = .012$, $\eta_p^2 = .119$, observed power = .729. Pairwise comparisons of task complexity main effects (see Table 13) showed that complex tasks resulted in significantly higher phrasal complexity than simple tasks, $p = .012$, 95% CI [.066, .501]. This finding indicates that there is a statistically significant difference between simple and complex tasks in terms of phrasal complexity. However, the interaction between modality and task complexity was found statistically insignificant, $F(1,51) = .130$, $p = .719$, $\eta_p^2 = .003$, observed power = .065. The reason for the statistically insignificant interaction can be attributed to a parallel increase between the two modes (see Figure 7).

Table 13. Phrasal Complexity Pairwise Comparisons of Task Complexity

(I) complexity	(J) complexity	Mean Difference (I-J)	SD	P[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Complex	Simple	.284*	.108	.012	.066	.501

*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Least Significant Difference.

The interaction graph suggests that the significant difference between the two modes (i.e., writing and speaking) is the same or very similar to the significant difference between the levels of task complexity (i.e., simple and complex) and the small effect size for the interaction supports this finding.

Figure 7 shows that the distance between the lines in writing is statistically very similar to the distance between the lines in speaking; the lines show a parallel trend. Phrasal complexity results show that task complexity produces the same effects in writing and speaking. They both contribute to phrasal complexity in the same manner.

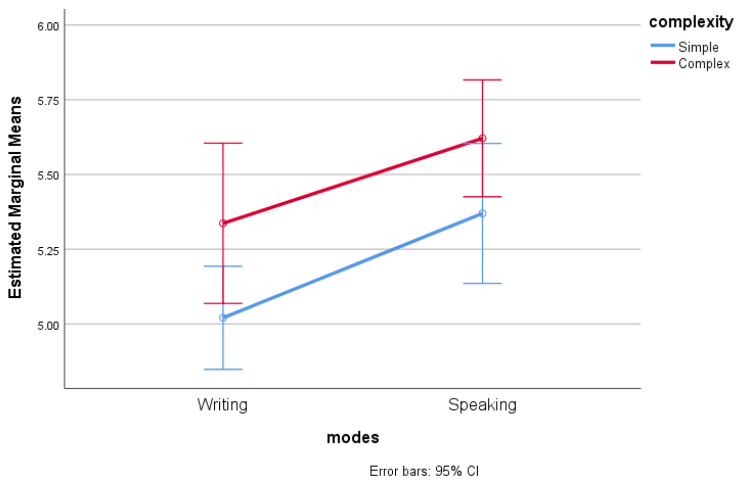


Figure 7. Modality and task complexity interaction for phrasal complexity

4.3 Accuracy

Accuracy was measured as the ratio of error-free units and the ratio of error-free clauses. The descriptive statistics of accuracy measures for all four tasks are presented in Table 14.

The following two sub-sections report the results for the ratio of error-free units and the ratio of error-free clauses.

Table 14. Descriptive Statistics for Accuracy Measures

Dependent Variables	Writing Simple Task	Writing Complex Task	Speaking Simple Task	Speaking Complex Task
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Accuracy				
Error-free units	0.63 (0.27)	0.59 (0.29)	0.53 (0.29)	0.58 (0.27)
Error-free clauses	0.71 (0.25)	0.68 (0.27)	0.64 (0.27)	0.67 (0.26)

4.3.1 The ratio of error-free units

The mean differences in the ratio of error-free units for task modality showed that participants were more accurate in writing (M = 0.63, SD = 0.27) than in speaking (M

= 0.53, *SD* = 0.29) in simple tasks. Similarly in complex tasks, participants had higher unit accuracy, although slightly, in writing (*M* = 0.59, *SD* = 0.29) than in speaking (*M* = 0.58, *SD* = 0.27).

The mean differences of task complexity in writing showed that simple tasks led to higher accuracy (*M* = 0.63, *SD* = 0.27) than complex tasks (*M* = 0.59, *SD* = 0.29). However, in speaking, complex tasks (*M* = 0.58, *SD* = 0.27) had higher accuracy than simple tasks (*M* = 0.53, *SD* = 0.29). The 95% confidence intervals in Figure 8 also indicate a large difference between modalities showing that writing exceeds speaking in simple tasks. Error bars for task complexity show an overlapping pattern both in writing and speaking.

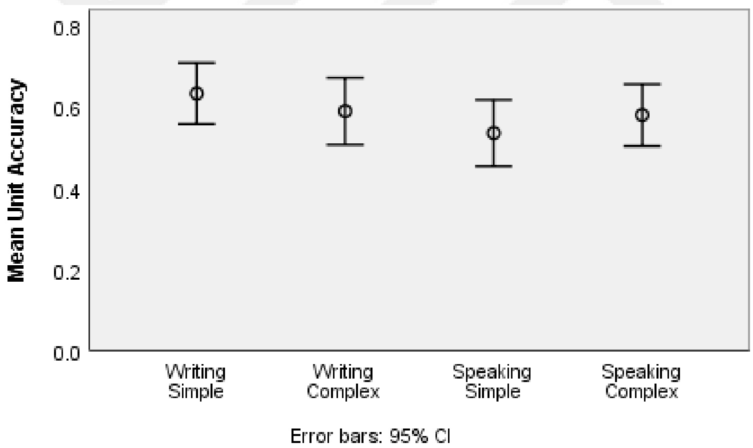


Figure 8. Error-free units scores with 95% confidence intervals

2x2 factorial repeated measures ANOVA (see Table 15) was carried out to analyze the statistical significance of the descriptive statistics. The main effect for task modality indicated a significant difference between speaking and writing with a medium effect size, $F(1,51) = 4.389$, $p = .041$, $\eta_p^2 = .079$, observed power = .538. Pairwise comparisons of modality main effects (see Table 16) showed that writing led to significantly more unit accuracy than speaking, $p = .041$, 95% CI [.002, .105].

This shows that participants' accuracy level was significantly affected by task modality. The main effect for task complexity, on the other hand, was found to be statistically insignificant, $F(1,51) = .000$, $p = .984$, $\eta_p^2 = .000$, observed power = .050. The interaction between modality and task complexity reached the level of significance with a medium effect size, $F(1,51) = 4.671$, $p = .035$, $\eta_p^2 = .084$, observed power = .564. It becomes important to check the interaction graph (see Figure 9) and conduct further pairwise comparisons using one-way repeated measures ANOVA to understand the nature of this interaction.

Table 15. 2x2 Repeated Measures ANOVA Results for Error-free Units

Error-free Units	df	MS	F	P	η_p^2	Observed power
Modes	1	.150	4.389	.041	.079	.538
Error(modes)	51	.034				
Complexity	1	7.692E-6	.000	.984	.000	.050
Error(complexity)	51	.018				
Modes*Complexity	1	.102	4.671	.035	.084	.564
Error (modes*complexity)	51	.022				

*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Least Significant Difference.

Consequently, four levels were defined for unit accuracy in one-way repeated measures ANOVA as writing simple, writing complex, speaking simple, and speaking complex to examine the pairwise comparisons.

The interaction graph in Figure 9 indicates that complex tasks do not differ across task modalities (i.e., speaking and writing). The interaction between modality and task complexity in error-free units seems to occur when the task is simple.

Table 16. Unit Accuracy Pairwise Comparisons of Modality

(I) modes	(J) modes	Mean Difference (I-J)	SD	P[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
Writing	Speaking	.054*	.026	.041	.002	.105

Pairwise comparisons were conducted to confirm this interpretation. Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $X^2(5) = 8.838$, $p = .116$ (see Table 17).

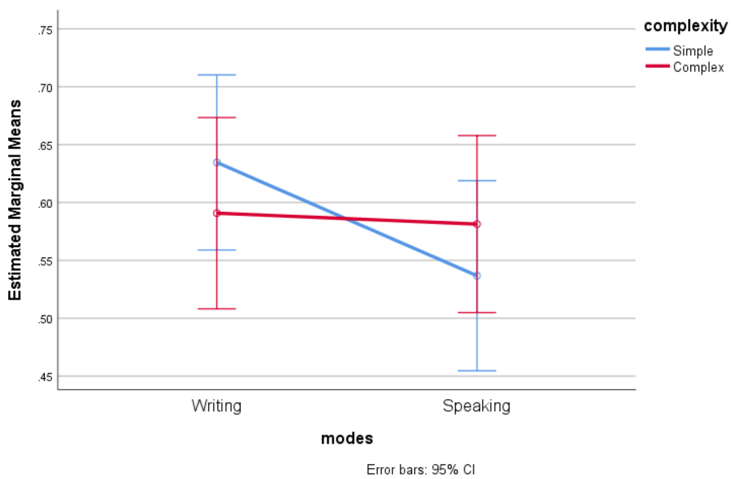


Figure 9. Modality and task complexity interaction for error-free unit accuracy

Tests of within-subjects effects (sphericity assumed) showed that there was a significant difference in unit accuracy scores between the independent variables, $F(3,153) = 3.408$, $p = .019$, $\eta_p^2 = .063$, observed power = .760 (see Table 18). Pairwise comparisons revealed that the differences between all the pairs were found to be statistically insignificant except for the writing simple and speaking simple pair, $p = .016$, 95% CI [.013, .183]. See all the pair comparisons in Table 19. These results are consistent with the significant interaction between modality and task complexity and give further information about the nature of the interaction.

Table 17. Mauchly's Test of Sphericity for Unit Accuracy

Within-subjects Effect	Mauchly's W	Approx. Chi-Square	df	P	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower bound
Unit Accuracy	.837	8.838	5	.116	.901	.956	.333

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects Table.

Table 18. Tests of Within-Subjects Effects for Unit Accuracy

		df	MS	F	P	η_p^2	Observed power
Unit Accuracy	Sphericity Assumed	3	.084	3.408	.019	.063	.760
	Greenhouse-Geisser	2.702	.093	3.408	.023	.063	.726
	Huynh-Feldt	2.868	.088	3.408	.021	.063	.745
	Lower-bound	1.000	.251	3.408	.071	.063	.441
Error(Unit Accuracy)	Sphericity Assumed	153	.025				
	Greenhouse-Geisser	137.823	.027				
	Huynh-Feldt	146.276	.026				
	Lower-bound	51.000	.074				

To sum up, the interaction between modality and task complexity in unit accuracy occurs when the task is simple. When the task is complex, the effect of modality on unit accuracy is not significantly different. In other words, the difference between the modes of writing and speaking in terms of their effect on unit accuracy is only significant when the task complexity is simple. It can be concluded that when examining the effects of simple tasks on unit accuracy, it is important to consider the mode of performance (writing or speaking) as it has a significantly different effect on unit accuracy.

Table 19. Repeated Measures Pairwise Comparisons for Unit Accuracy

(I) Unit Accuracy	(J) Unit Accuracy	Mean Difference (I-J)	SD	P[b]	95% Confidence Interval for Difference[b] LB	UB
Writing Simple	Speaking Simple	.098*	.031	.016	.013	.183
Writing Simple	Writing Complex	.044	.025	.480	-.024	.111
Speaking Simple	Speaking Complex	-.045	.030	.889	-.128	.039
Speaking Complex	Writing Complex	-.009	.034	1.000	-.104	.085

*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

4.3.2 The ratio of error-free clauses

The mean differences in the ratio of error-free clauses for task modality showed that participants produced more accurate clauses in writing (*M* = 0.71, *SD* = 0.25) than in speaking (*M* = 0.64, *SD* = 0.27) in simple tasks. Similarly in complex tasks, participants had higher clausal accuracy, although slightly, in writing (*M* = 0.68, *SD* = 0.27) than in speaking (*M* = 0.67, *SD* = 0.26).

The mean differences of task complexity in writing showed simple tasks (*M* = 0.71, *SD* = 0.25) led to higher clausal accuracy than complex tasks (*M* = 0.68, *SD* = 0.27). In speaking, on the other hand, complex tasks (*M* = 0.67, *SD* = 0.26) had higher accuracy than simple tasks (*M* = 0.64, *SD* = 0.27). Refer to Table 14 for the descriptive statistics.

The 95% confidence intervals in Figure 10 also suggest a difference between writing and speaking showing writing exceeds speaking in simple tasks. Error bars for task complexity show an overlapping pattern both in writing and speaking.

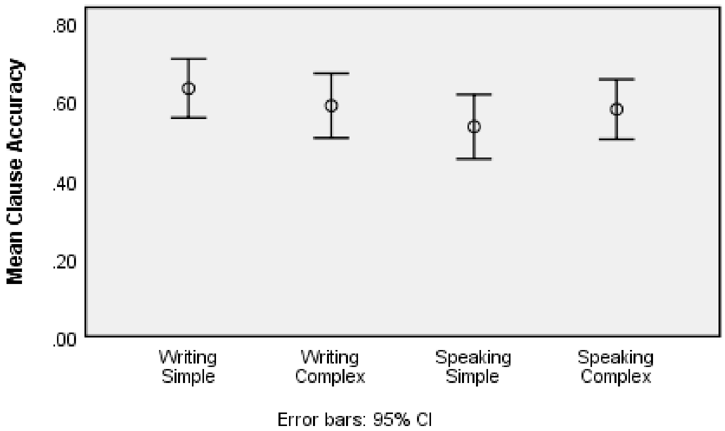


Figure 10. Error-free clauses scores with 95% confidence intervals

2x2 factorial repeated measures ANOVA was conducted to analyze the statistical significance of the descriptive statistics (see Table 20). The main effect of task modality on clausal accuracy did not indicate any statistically significant difference between speaking and writing, $F(1,51) = 3.849$, $p = .055$, $\eta_p^2 = .070$, observed power = .486. This shows that participants' clausal accuracy was not affected by the mode in which they were performing the task. However, it is important to note that the effect size was medium which suggests some effect of clausal accuracy was found. Additionally, the level of insignificance was very close to the benchmark which shows a trend toward significance and power of the statistical test.

Similarly, the main effect of task complexity was found to be statistically insignificant, $F(1,51) = .196$, $p = .660$, $\eta_p^2 = .004$, observed power = .072. This result suggests that participants' clausal accuracy scores were not affected by task complexity. Finally, the interaction between modality and task complexity was found to be statistically insignificant with a medium effect size, $F(1,51) = 3.637$, $p = .062$, $\eta_p^2 = .067$, observed power = .465. This result indicates that the effect of modality (i.e., speaking and writing) on clausal accuracy is statistically not affected by task complexity (i.e., simple and complex) or vice versa. While the interaction did not reach the level of statistical significance, the high effect size and the *F* ratio yield the power of the statistical test.

Table 20. 2x2 Repeated Measures ANOVA Results for Error-free Clauses

Error-free Clauses	df	MS	F	P	η_p^2	Observed power
Modes	1	.078	3.849	.055	.070	.486
Error(modes)	51	.020				
Complexity	1	.002	.196	.660	.004	.072
Error(complexity)	51	.009				
Modes*Complexity	1	.049	3.637	.062	.067	.465
Error (modes*complexity)	51	.013				

The descriptive statistics in Table 14 for accuracy measures indicate a similar trend between the mean scores of unit accuracy and clausal accuracy. However, no significant difference or interaction was found between independent variables for clausal accuracy while significant interaction and difference between modes were detected for unit accuracy. Considering the very similar tendencies between accuracy measures, the effect sizes, and high *F* ratios for clausal accuracy, it might be useful to replicate this study to validate the findings for clausal accuracy.

4.4 Lexical complexity

43 out of 52 participants' language outputs (*N* = 43) were analyzed for lexical complexity due to the 50-words minimum requirement. Descriptive statistics of lexical complexity (see Table 21) for task modality indicated that participants had higher lexical complexity in writing (*M* = 47.42, *SD* = 14.90) than in speaking (*M* = 33.00, *SD* = 11.58) in simple tasks. Similarly in complex tasks, participants produced more lexically complex language outputs in writing (*M* = 49.31, *SD* = 18.30) than in speaking (*M* = 36.82, *SD* = 13.86).

The means for task complexity showed that complex tasks (*M* = 49.31, *SD* = 18.30) had a higher mean for lexical complexity than simple tasks (*M* = 47.42, *SD* = 14.90) in writing. The same trend was observed in speaking as well. Complex tasks (*M* = 36.82, *SD* = 13.86) had higher lexical complexity than simple tasks (*M* = 33.00, *SD* = 11.58).

Table 21. Descriptive Statistics for Lexical Complexity

Dependent Variable	Writing Simple Task	Writing Complex Task	Speaking Simple Task	Speaking Complex Task
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Lexical Complexity	47.42 (14.90)	49.31 (18.30)	33.00 (11.58)	36.82 (13.86)

The 95% confidence intervals in Figure 11 visualized the clear mean difference between the two modes while showing an almost overlapping pattern for task complexity. The graph suggests a significant effect of task modality and an insignificant difference between task complexity.

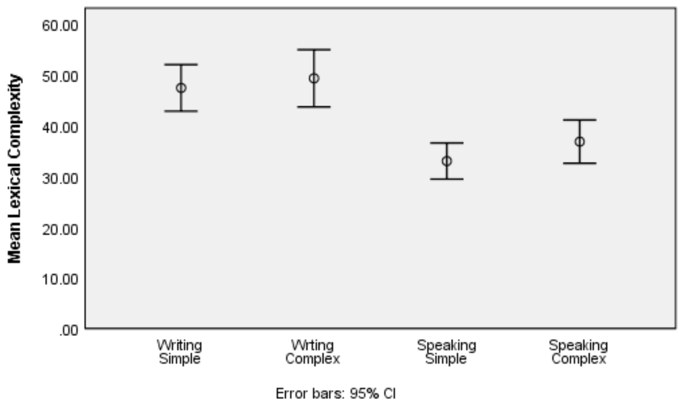


Figure 11. Lexical complexity scores with 95% confidence intervals

2x2 factorial repeated measures ANOVA was conducted on lexical complexity scores to analyze the observations from descriptive statistics and error bars (see Table 22). The main effects of task modality on lexical complexity indicated a statistically significant difference as observed from the error bars with a very strong effect size, $F(1,42) = 71.878$, $p = .000$, $\eta_p^2 = .631$, observed power = 1.000. The pairwise comparisons of the main effects (see Table 23) showed that lexical complexity in writing was significantly higher than in speaking, $p = .000$, 95% CI [10.250, 16.654]. However, the main effects of task complexity were found to be statistically insignificant, $F(1,42) = 2.366$, $p = .132$, $\eta_p^2 = .053$, observed power = .324. This finding indicates that task complexity has no significant effect on lexical complexity but the effect size on the insignificant finding indicates the power of the statistical test. Finally, the interaction between modality and task complexity was found statistically insignificant, $F(1,42) = .372$, $p = .545$, $\eta_p^2 = .009$, observed power = .092. A very small effect size supports the insignificant finding.

Table 22. 2x2 Repeated Measures ANOVA Results for Lexical Complexity

Lexical Complexity	df	MS	F	P	η_p^2	Observed power
Modes	1	7781.363	71.878	.000	.631	1.000
Error(modes)	42	108.257				
Complexity	1	351.180	2.366	.132	.053	.324
Error(complexity)	42	148.454				
Modes*Complexity	1	39.946	.372	.545	.009	.092
Error (modes*complexity)	42	107.365				

No significant interaction indicates that although there is a significant difference between writing and speaking, the difference does not significantly vary in simple and complex tasks. As a result, no statistically significant interaction was found between modality and task complexity. Figure 12 illustrates that while simple and complex tasks have very close estimated marginal means in the same mode, writing as modality results in significantly higher lexical complexity than speaking regardless of whether tasks are simple or complex.

Table 23. Pairwise Comparisons of Modality for Lexical Complexity

(I) modes	(J) modes	Mean Difference (I-J)	SD	P[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
Writing	Speaking	13.452*	1.587	.000	10.250	16.654

*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Least Significant Difference.

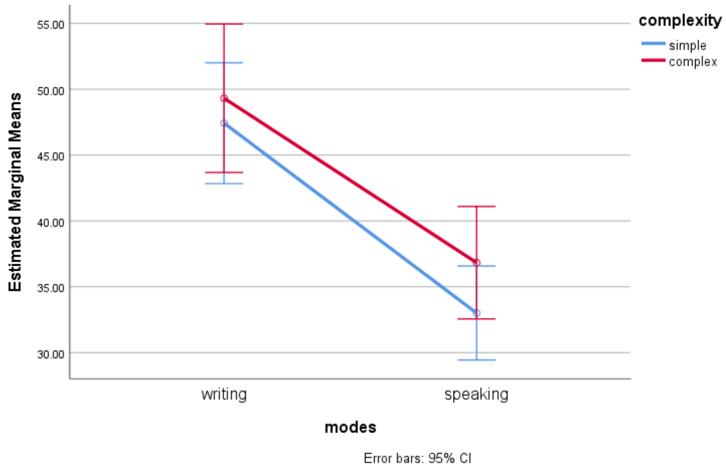


Figure 12. Modality and task complexity interaction for lexical complexity

4.5 Summary of the chapter

This section provides a discussion of the results in relation to research questions and hypotheses.

As an answer to research question 1, significant differences in speaking and writing were found for phrasal complexity, which was higher in speaking; for unit accuracy, which was higher in writing, and for lexical complexity, which was higher in writing. More lexically complex language production in writing confirms the hypothesis regarding higher lexical complexity in writing. Accuracy was also hypothesized to be higher in writing and it was partially confirmed. Only unit accuracy was found significantly higher in writing. The effect of modality on clausal accuracy was found to be statistically insignificant. Different effects of writing and speaking were hypothesized for syntactic complexity. This hypothesis was confirmed only for phrasal complexity. The hypothesis for the different effects of modality on syntactic complexity was rejected for overall syntactic complexity and complexity by subordination. The results for these dependent variables were found to be statistically insignificant.

As an answer to research question 2, significant differences in simple and complex tasks were found for complexity by subordination, which was higher in simple tasks, and phrasal complexity, which was higher in complex tasks. The null hypotheses were assumed for the effects of task complexity on all the dependent variables. While the null hypotheses were rejected for complexity by subordination and phrasal complexity, they were confirmed for overall syntactic complexity, unit accuracy, clausal accuracy, and lexical complexity.

As an answer to research question 3, only one significant interaction was found. Unit accuracy yielded a significant interaction between task complexity and modality revealing that modality has a varying effect on unit accuracy in simple tasks. In simple tasks, writing leads to significantly higher unit accuracy than speaking while in complex tasks, the effect of task modality was found statistically insignificant. This finding partially confirmed the hypothesis regarding task complexity and modality interaction for accuracy because the interaction for clausal accuracy was found to be insignificant. The interaction hypothesis for syntactic complexity was rejected. No significant interaction between task complexity and modality was found for overall, phrasal, and subordination complexity. However, it is important to note that trends toward significant interaction were observed for overall syntactic complexity and clausal accuracy. The null hypothesis was assumed for lexical complexity and was confirmed. Due to the parallel relationship between task complexity and modality, the interaction for lexical complexity was found to be insignificant. Table 24 presents the summary of all the results.

Table 24. Summary of the Results

	Task Complexity	Modality	Task Complexity* Modality
	Simple & Complex	Speaking & Writing	Interaction
Syntactic Complexity			
Overall	-	-	-
Subordination	+	-	-
Phrasal	+	+	-
Accuracy			
Unit	-	+	+
Clausal	-	-	-
Lexical Complexity			
D calculation	+	-	-

+ Marks the significant results
- Marks the insignificant results

CHAPTER 5

DISCUSSION

5.1 Introduction

This chapter discusses the main effects of task complexity and task modality and the interaction effects between the independent variables in light of the Limited Attentional Capacity Hypothesis (Skehan, 2009, 2014), the Cognition Hypothesis (Robinson, 2001b, 2011a), and the Speech and Writing production models (Hayes, 1996, 2012; Kellogg, 1996; Kormos, 2006; Levelt,1989). The results for each research question will be addressed separately.

5.2 Research question 1: The effects of task modality

The participants' written and oral performances were assessed through syntactic and lexical complexity, and accuracy. The hypotheses were partially confirmed for task modality. Different effects of task modality were hypothesized on syntactic complexity without predicting directionality. This hypothesis was partially confirmed. Only the phrasal complexity measure revealed a significant difference in the two modes while the effects of task modality on syntactic complexity and subordinate complexity were found to be statistically insignificant. Accuracy was hypothesized to be higher in writing. While the unit accuracy findings confirmed this hypothesis, the effect of task modality on clausal accuracy was statistically insignificant. Lexical complexity was hypothesized to be higher in writing and the main effects of task modality confirmed this hypothesis.

In what follows, findings for the effects of task modality on each dependent variable will be discussed separately.

5.2.1 Syntactic complexity

The results regarding the lack of difference in overall syntactic and subordination complexity are similar to those of Granfeldt (2008), Ferrari and Nuzzo (2009, as cited in Vasylets et al., 2019), Kormos and Trebits (2012), and Kormos (2014) but contrary to Ellis and Yuan (2005), Kuiken and Vedder (2011), Tavakoli (2014), Vasylets et al. (2017), Zalbidea (2017) and Vasylets et al. (2019). It is important to note that there are major differences between the studies such as research design, the statistical tests, participant background, measures, and tasks employed.

The lack of difference in overall syntactic and subordination complexity might be due to the measures used in this study. It is possible that more sensitive measures might be needed to reveal subtle syntactic differences between writing and speaking. Byrnes, Maxim, and Norris (2010) argue that a major change from the oral to literate continuum can be observed through intraclausal elaborations such as pre- and post-modifications. Norris and Ortega (2009) similarly argued for the importance of an intraclausal measure revealing the phrase level. Consequently, the phrasal complexity measure was employed in this study. While the effects of task modality on overall syntactic and subordination complexity were found to be statistically insignificant, phrasal complexity was found to be significantly higher in speaking. It might be possible that the phrasal complexity measure is more sensitive to reveal changes in writing and speaking (Byrnes et al., 2010). However, why it was found higher in speaking certainly calls for an answer.

Directional hypotheses were not made considering the mixed findings in the literature for syntactic complexity. However, the comparison of speech (Kormos, 2006; Levelt, 1989) and writing models (Kellogg, 1996) suggests that writing is more self-paced and offers more opportunities for careful planning and linguistic encoding compared to speaking (Gilabert et al., 2016; Kormos, 2014; Vasylets et al., 2017). Considering the clear differences in terms of online and offline planning favoring written mode, higher phrasal complexity in speaking bears an explanation.

One of the reasons for higher phrasal complexity in speaking can be found in Halliday's (2002) distinction between "speakable wordings" and "writeable wordings" (p. 345). Different ways of meaning-making can be observed in the two modes. While speech can be characterized as "processlike, intricate, with meanings related serially", writing can be considered "productlike, tight, with meanings related as components" (Halliday, 2002, p. 350). As a result, these characteristics might result in higher phrasal complexity in speaking as it can lead to a higher number of words within a clause. Another explanation offered by Halliday (2002) is related to lexical density. Higher lexical density in writing is not necessarily related to the increase in lexical items. Higher lexical complexity in writing can be attributed to the decrease in non-lexical items or grammatical words. This decrease results in even lower number of clauses and words within a clause in writing (Halliday, 2002). Although phrasal complexity is important to reveal differences at a clausal level, a closer inspection of varying clause lengths in the context of task modality might be needed to fully picture the nature of syntactic complexity in writing and speaking. For example, Halliday's (2002) argument suggests that writing might have shorter clauses with reduction techniques such as nominalization. While this might result in more concise sentences with fewer words and clauses than speaking, therefore rendering them as not complex under the current phrasal complexity measure, it does not necessarily make writing less syntactically complex than speaking. Granfeldt (2008) supports this argument stating that there can be qualitative differences between writing and speaking in terms of syntactic complexity. Based on these arguments, Byrnes and Manchon (2014) address the potential problem of employing identical measures for examining writing and speaking. Therefore, ratio measures might not fully account for qualitative syntactic differences and varying linguistic characteristics between writing and speaking.

Another explanation could be related to individual differences. Weissberg (2000) investigated the effects of writing and speaking on morpho-syntactic innovations. The data collection procedure involved meticulous tracking of the emergence of morpho-syntactic structures in the spoken and written samples which included an intake interview, writing samples, essays, interviews, and weekly at-home journal entries. The overall results suggested that the emergence of new morpho-syntactic forms and accuracy development tend to occur in the written mode. However, a closer analysis of each participant revealed that learners might "range along the continuum according to their preference for writing or speech as the primary vehicle for introducing new syntactic elements" (Weissberg, 2000, p. 49). Investigating what modality the participants felt more comfortable with or preferred might explain the higher phrasal complexity in speaking.

Task modality research offers mixed results regarding syntactic complexity. Although measures and units can be different across studies, they usually involve general syntactic measures calculated by a ratio of clauses to units and the length of a clause and/or a unit (Kormos & Trebits, 2012; Kuiken & Vedder, 2011; Tavakoli, 2014; Vasylets et al., 2017; Zalbidea, 2017). By acknowledging various reasons for mixed results such as different tasks, proficiency levels, and research design, the syntactic results of this study note that a closer inspection of phrasal complexity in particular and ratio-based syntactic measures in general for task modality might be needed to reveal types of subordinate clauses and modifications in a unit and a phrase to better account for the differences in the two modes.

5.2.2 Lexical complexity

The main effect of task modality on lexical complexity revealed, with a very strong effect size, $\eta_p^2 = .631$, that writing leads to significantly more lexical complexity than speaking. This result confirms the research hypothesis and is aligned with other studies (Bulté & Housen, 2009, as cited in Vasylets et al., 2019; Ellis & Yuan, 2005; Granfeldt, 2008; Kormos & Trebits, 2012; Kormos, 2014; Vasylets et al., 2017; Vasylets et al., 2019; Zalbidea, 2017). Only one study reported no effect of task modality on lexical complexity (Yu, 2009). Kuiken and Vedder (2011) could only report their observations indicating no effect of task modality by looking at the mean scores since they did not employ any statistical analysis between writing and speaking. The different results in these two studies could stem from statistical tests or lack thereof, tasks employed, and participant background.

Higher lexical complexity in writing can be explained by the cyclical and interactive nature of cognitive processes as well as opportunities for online and offline planning (Kellogg, 1996; Vasylets et al., 2017). The offline nature of writing allows learners to have a self-paced production and to integrate content planning into actual writing (Gilabert et al., 2016; Kormos, 2014). These conditions create more opportunities for careful planning, extensive lexical searches, and monitoring compared to speaking. Furthermore, learners can benefit from these opportunities both during and after language production (Williams, 2012). Consequently, one reason for higher complexity in writing could be attributed to the offline and self-paced nature of written mode facilitating retrieval and elaboration of ideational content, more careful lexical searches, and monitoring of word choices. Another reason could be related to the permanence of written output which could help learners avoid repetition of the same words. The visibility of the written output allows learners to revisit the information in their memory (Kuiken & Vedder, 2012). Speakers, on the other hand, must keep the information that they have planned and already produced in their memory. Speaking offers fewer opportunities to conduct exhaustive lexical searches during and before language production than writing as learners have to make their lexical choices very quickly at the stage of formulation. In fact, more inherent time pressure in speaking could tax learners' cognitive resources at the conceptualization stage and therefore make the preparation of a complex preverbal plan very challenging from the beginning.

5.2.3 Accuracy

Accuracy was hypothesized to be higher in writing. While the unit accuracy measure confirms the hypothesis, clausal accuracy missed the statistical benchmark for a significant difference. However, the examination of the effect sizes, descriptive statistics, and high F ratio indicated that clausal accuracy has almost the same pattern as unit accuracy and exhibits a trend toward a significant difference. Therefore, the statistically insignificant effect of task modality on clausal accuracy could be attributed to the sample size and should be retested with a larger group to confirm its tendency toward a significant difference and interaction. Higher accuracy in writing was also found in most studies (Ellis, 1987; Ellis & Yuan, 2005; Kormos & Trebits, 2012; Kormos, 2014; Zalbidea, 2017) while two studies found it higher in speaking (Ferrari and Nuzzo, 2009, as cited in Vasylets et al., 2019; Granfeldt, 2008). Vasylets et al. (2017) reported different findings from the previous studies as they revealed no significant effect of task modality on accuracy. Differences in results could be due to the different research designs, participant backgrounds, L1 backgrounds, and sample sizes. The result of this study, however, seems to be aligned with many of the studies in the literature.

Higher accuracy in writing could be attributed to the availability of different opportunities for conceptualization, formulation, and monitoring in writing compared to speaking. Considering the recursive nature of writing, learners have the opportunity to go back to the ideas that they formed at the formulation stage to evaluate, reshape and reformulate them if necessary. Consequently, learners can benefit from these opportunities to revise the language that they produced at the Execution stage (Kellogg, 1996). This interactive and cyclical nature of writing allows learners to have additional time at the stages they need and to attend to the form without overtaxing their memory. Compared to speaking, the permanency of written output also allows learners to have closer monitoring opportunities of the forms. Time pressure in speaking, on the other hand, leads to overtaxing of cognitive resources and little opportunity for monitoring. Writing with inherent in-built planning opportunities and availability of time can free up attentional resources and induce focus-on-form processes through controlled processing (Ellis & Yuan, 2005).

5.3 Research question 2: The effects of task complexity

The null hypotheses were assumed for the task complexity effects on syntactic and lexical complexity, and accuracy measures based on the mixed results and scarcity of task complexity and modality research. The null hypotheses were retained for overall syntactic complexity, lexical complexity, and accuracy measures. Subordination and phrasal complexity, on the other hand, were affected by task complexity. Phrasal complexity was found to be higher in complex tasks. Subordination complexity, however, was significantly higher in simple tasks. Each dependent variable will be discussed separately in the following section.

5.3.1 Syntactic complexity

The lack of differences between simple and complex tasks for overall syntactic complexity is similar to Zalbidea's (2017) findings and partially similar to Tavakoli (2014) and Kuiken and Vedder (2011) as they found no task complexity effects in writing. However, it contradicts Vasylets et al.'s (2017) findings as they reported a higher mean length of a unit in complex tasks. It is important to keep in mind that different tasks and participant profiles were employed in the studies making it difficult to conduct direct comparisons. One explanation for the result of the current study can stem from the overall syntactic index. The mean length of AS or T-unit might not be sensitive enough to reveal the subtle changes between simple and complex tasks. Another explanation might come from the task design. Narrating the same number of pictures in both simple and complex might have resulted in similar overall syntactic complexity. Therefore, potential changes in linguistic complexity might be more apparent at a clausal level in such task designs. Supporting this argument, the differences between simple and complex tasks in fact were found in the measures of subordination and phrasal complexity. These findings also support the argument that measures reflecting clausal complexity are needed as specific and distinct measures to reveal linguistic differences (Norris & Ortega, 2009).

Subordination complexity results revealed that simple tasks had significantly higher complexity than complex tasks while phrasal complexity was found to be significantly higher in complex tasks. Robinson (2001a, 2011a) predicts that syntactic complexity along with accuracy and lexical complexity increases simultaneously as the complexity of a task increases. Skehan (2009) also makes the prediction that learners tend to produce more complex language when the tasks require the organization and manipulation of information and content. While phrasal complexity results confirm these predictions, subordination complexity reveals the opposite. Although it seems contradictory to find higher subordination complexity in simple tasks, its explanation can be found in task design and the need for finer-grained measures.

Foster and Tavakoli (2009) argue that connecting background and foreground events in a narrative is expected to be accomplished through subordinate clauses. In this study, simple tasks had both background and foreground events (going to a picnic while a dog sneaks into the basket in the background or going home while dropping a package and being followed by a man in the background). Participants needed to connect these events in simple tasks. However, complex tasks in this study did not inherently pose background and foreground events. Participants needed to plot such events and a story, therefore were free to construct stories including foreground and/or background. In this study, simple tasks requiring a narrative of pictures ordered in a coherent storyline might promote connecting events through subordinate clauses while complex tasks might encourage learners to be engaged in intraclausal modifications through adjectives, adverbs, prepositional phrases, or nonfinite clauses to deliver their complex conceptualization. Consequently, it can be argued that simple tasks requiring the connection of events might promote certain use of structural complexity. Plotting a story and sequencing events in the complex tasks, on the other hand, were completely left to the participants. Therefore, it is possible that participants planned and plotted stories that reflect the conceptualization demands at the phrasal level.

Elaborating on Halliday and Matthiessen's (1999) systemic functional theory, Norris and Ortega (2009) argue that clause length (referred to as phrasal complexity in this study) taps a complexification that is more narrowly defined. They provide a further distinction on subordination and phrasal level complexity that greater use of phrasal complexification can reveal advanced language development while the use of coordination can be expected to reveal beginning levels of development (Norris & Ortega, 2009). It is possible that participants stretched their interlanguage and linguistic resources to match the narratives they plotted with necessary linguistic materials and reflected this linguistic complexity significantly at the phrasal level.

The overall consideration of syntactic complexity results suggests that participants produced more complexity by subordination in simple tasks, arguably stemming from connecting pictures to successfully deliver a given storyline. Since simple tasks in this study were argued to pose fewer conceptualization demands, participants might simply connect the pictures together through possibly mostly coordinators and therefore produce more clauses per unit in simple tasks than complex tasks. Higher phrasal complexity in complex tasks suggests that participants met complex conceptualization demands through modifications within a phrase and accessed the upper limits of their interlanguage revealing more advanced language production (Norris & Ortega, 2009). This could suggest that participants were engaged in contemplation of a plot supported by descriptions and elaborations in complex tasks during the conceptualization stage. Further support for the arguments and findings can be found below in example utterances of the same participant in simple and complex tasks.

An example ending from a simple task:

They were sad and went home also left the picnic for another day

An example ending from a complex task:

The man gave the beautiful cloud-shaped pearled ring to his wife

The main effects of both task modality, η_p^2 = .12, and task complexity, η_p^2 = .11, revealed very close to strong effect size for phrasal complexity. This also supports the previous arguments that the phrasal complexity measure was more sensitive to revealing subtle changes in language production. However, it is important to note that a closer and more rigorous inspection of language outputs in this study regarding subordinations and phrasal complexity is needed to confirm previous arguments.

5.3.2 Lexical complexity

Main effects of task complexity on lexical complexity were found to be statistically insignificant. This finding corroborates other studies (Kormos & Trebits, 2012; Kuiken & Vedder, 2011; Vasylets et al., 2017; Zalbidea, 2017). However, it is important to note that complex tasks resulted in higher mean scores than simple tasks (see Table 21). The effect size on the insignificant finding was very close to medium size, η_p^2 = .053 indicating the power of the test and the possibility of finding a significant effect. Most studies, on the other hand, revealed an insignificant effect of task complexity on lexical complexity so other possible explanations need to be considered.

The explanation of this finding might lie in the operationalization of task complexity or that cross-task comparison to investigate changes in lexical complexity might constitute a problem (Tavakoli & Foster, 2008). To explore the first explanation, the aforementioned studies' operationalization of task complexity was checked. It showed that different operationalizations of task complexity were employed in these studies. Kuiken and Vedder (2011) employed a manipulation of +/- few elements in argumentative tasks of holiday choices for task complexity. Zalbidea (2017) adapted Kuiken and Vedder's (2011) tasks and used +/- few elements in the argumentative tasks. Vasylets et al. (2017) used different tasks posing a fire situation and justification for a rescue plan. They operationalized task complexity based on +/- reasoning demands as the simple version depicted a less critical situation and was argued to demand lower involvement of reasoning. Kormos and Trebits (2012) used task complexity operationalization that is the same as the current study. They used narrative tasks with pictures ordered in a coherent storyline for simple tasks and with random pictures with no story order for complex tasks. Yet, all these studies found an insignificant effect of task complexity on lexical complexity. Either all the aforementioned studies failed to operationalize task complexity or task design might constitute a sensitive issue in measuring lexical complexity when comparing the effects of task complexity within the same mode.

Tavakoli and Foster (2008) elaborated on their findings of lexical complexity and stated that different stimuli in narratives might be prompting vocabulary choices and cross-task comparisons for lexical diversity can be a problem. Tavakoli (2008) did not include lexical complexity in her complexity measures based on the argument that entirely different stimuli affect vocabulary choices. While this could offer a potential answer to the insignificant findings in narrative tasks, findings for argumentative tasks still require an answer. Vasylets (2017) offers an explanation in her PhD thesis. While she found an insignificant effect of task complexity on lexical complexity measured by D calculation, she found a significant effect on lexical sophistication. The reason for lexical complexity results might be related to having argumentative tasks in both simple and complex task conditions. Although the cognitive complexity of the tasks differs, participants still require producing arguments and justifications for actions in simple and complex versions. Argumentative tasks with different cognitive demands might not really promote a variety of words but might encourage the use of different frequencies of lexical items (Vasylets, 2017). Skehan (2009) makes a similar argument, but on the narrative tasks, that more “input-driven and unforgiving” tasks such as narratives seem to result in less frequent lexis, possibly because of responding to the events within a narrative (p. 517). Considering all the studies mentioned here used either the D calculation or alternative TTR measures, a different measure of lexical complexity such as lexical sophistication might be needed to reveal, if any, the effect of task complexity on lexical complexity. Additionally, a careful task design for the comparison of different tasks might be needed to tap into differences in the choice of lexical items. Based on Skehan’s (2009) and Tavakoli and Foster’s (2008) arguments, providing more freedom for decision-making and interactions in tasks when designing task complexity, for example, might reveal possible changes in lexis.

Overall, task modality can be concluded as a stronger indicator of changes in lexical complexity than task complexity. It is also important to note that this study found a considerable effect size, $\eta_p^2 = .053$ for the insignificant effect of task complexity on lexical complexity. This suggests that there might be an effect of task complexity that was not detected in the current sample. More research is required to confirm or reject the arguments about the lexical complexity measures for task complexity.

5.3.3 Accuracy

Higher phrasal complexity results in complex tasks were argued to support the Cognition Hypothesis (Robinson, 2001a, 2011a) and the Limited Attentional Hypothesis (Skehan, 2009). Since the crucial difference between these hypotheses stem from the way they view the attentional resources, the accuracy results become a determining factor to reveal any support for one of these hypotheses. The accuracy results revealed an insignificant effect of task complexity on both unit and clausal accuracy. These results are aligned with Kormos and Trebits (2012), Vasylets et al. (2017), and Zalbidea (2017) while contrary to Kuiken and Vedder (2011).

Robison's (2001a, 2011a) multiple-pool model for the use of cognitive resources predicts that attention to one aspect of linguistic performance does not diminish the attention to other aspects. Therefore, the Cognition Hypothesis argues for a simultaneous increase in complexity and accuracy dimensions in complex tasks. While the results for accuracy measures show no support for this argument, they seem to be more compatible with Skehan's (1998) limited-resource model. Skehan (2009) argues that the need to manipulate and reorganize information induces the production of more complex language. Due to limited attentional capacity, learners tend to prioritize accuracy or complexity (Skehan, 2009). No effect of task complexity on accuracy together with an increase in phrasal complexity in complex tasks seems to fit in a trade-off account.

Interaction graphs in this study for accuracy measures, however, reveal interesting results that accuracy is affected differently by task modality and complexity. This leads to the last section of the discussion chapter.

5.4 Research question 3: Task complexity and modality interaction

Interaction analyses in this study provided statistical grounds for the claims of possible interactions between task complexity and modality in the literature. The statistical analyses and the graphs for phrasal complexity (see Figure 7) revealed no interaction with a clear parallel increase between task complexity and task modality. The results of overall syntactic complexity, on the other hand, indicated a trend for an interaction. This trend did not reach a level of significance. The potential problems of using general and identical measures of L2 performance for task modality were discussed in the previous sections. Consequently, the potential interaction effects for the mean length of a unit need to be explored in future studies with the concern of choosing more specific measures that reveal qualitative syntactic differences. It is possible that a closer investigation of syntactic complexity with finer-grained measures reveals an interaction between task complexity and task modality.

Lexical complexity revealed a very clear parallel increase from simple to complex tasks and from speaking to writing (see Figure 12). Although the interaction is not detected due to the parallel increase between the independent variables, this finding has important implications. Stretching interlanguage knowledge and trying to reach the upper limits of lexical knowledge could be strengthened through the modality in which the complex task is performed. Complex tasks when performed in the written mode could further push learners to extend and elaborate the underlying interlanguage system, specifically in lexis.

Interesting results were found for accuracy measures. Unit accuracy revealed a statistically significant interaction between task complexity and modality. Clausal accuracy results indicated a high likelihood of interaction although the current sample did not detect one. This conclusion was made based on the similar trend between the mean scores of unit accuracy and clausal accuracy interactional, and the high *F* ratio and the effect size of the clausal accuracy result, $F(1,51) = 3.637$, $p = .062$, $\eta_p^2 = .067$.

A very interesting key commonality emerged from interaction analyses of unit accuracy and clausal accuracy. Further analyses of the significant interaction and the graph interpretations revealed that it is the simple tasks affecting modes differently, not the complex tasks. The effect of complex tasks on accuracy in writing and speaking was not statistically different. This intriguing result calls for an explanation.

One explanation might be related to the choice of using attentional resources. It might be possible that complex tasks requiring increased attention to conceptualization did not result in significant accuracy differences in writing and speaking because attentional resources were allocated to content planning instead of encoding and monitoring linguistic form in the two modes (Kormos & Trebits, 2012). The possible theoretical advantages of writing such as closer monitoring opportunities and being able to read and edit the visible language output might only be utilized in a significant way in tasks that do not require high conceptualization.

The availability of planning time and different types of it has been theorized to pose different benefits (Ellis, 2005). Pretask planning, a type of planning, for example, was theorized to potentially benefit the conceptualization of a message (Ellis, 2005). However, what learners choose to spend their planning time on is crucial to unravel the potential benefits of planning. Ortega (2005) shed some light on this issue. Overall remarks from her interviews regarding what participants planned when they had extra time showed that they were engaged in “to collect one’s thoughts" or "digest everything first" (Ortega, 2005, p. 87). Further, being engaged in lexical searches was revealed to be the most frequently mentioned use of having extra time. Consequently, one explanation of the current results might be found in the self-determined and cyclical nature of writing. The written activity includes the time spent on planning the content in the writing process which allows learners to allocate extensive time for planning before starting to execute. Writing is a self-determined activity due to offline planning opportunities. This allows learners to stop the writing process and focus on retrieving information or planning (Grabowski, 2007). Therefore, depending on what learners choose to spend the availability of time in writing is important to the interpretation of the current results. Considering Ortega's (2005) and Ellis' (2005) arguments, it might be possible that participants used their possibly advantageous resources in writing for conceptualization and therefore the accuracy of complex tasks in writing and speaking did not significantly differ from each other. While this might be the case for complex tasks, simple tasks offering organized content might relieve the need to attend to the conceptualization stage and result in more time and resources for the linguistic encoding and monitoring of language output in writing. It could be suggested that the written modality can provide greater and more extensive opportunities for the control of interlanguage, attention to form, and monitoring thanks to the self-determined and recursive implementation of other cognitive processes especially when the task design involves lower levels of conceptualization and organizational demands.

CHAPTER 6

CONCLUSION

6.1 Summary and conclusions

The main aim of this study was to examine a possible interaction between task complexity and modality along with the main effects of each independent variable on the accuracy, and syntactic and lexical complexity. It aimed to provide contributions to recent debates about whether to include task modality in task complexity theories and a mode-sensitive rethinking in cognitive TBLT research. Studies leading to such debates have examined the effects of writing and speaking on L2 performances (Bulté & Housen, 2009, as cited in Vasylets et al., 2019; Ellis, 1987; Ellis & Yuan, 2005; Ferrari & Nuzzo, 2009, as cited in Vasylets et al., 2019; Kormos, 2014; Yu, 2009; Vasylets et al., 2019) and the potentially different effects of task complexity on L2 performances in the two modes (Kormos & Trebits, 2012; Kuiken & Vedder, 2011; Tavakoli, 2014; Vasylets et al., 2017; Zalbidea, 2017). However, the combined effects of task complexity and modality have been neglected in these studies, except for Vasylets et al. (2017). This study attempted to investigate the separate as well as the combined effects of task complexity and modality on L2 performances to be able to provide a more detailed and statistical account for the conclusions drawn in the previous studies regarding the potentially different effects of task complexity in the two modes.

Moreover, differently from the previous studies some of which used different participants and/or data from different years for the two modes and failed to employ counterbalancing for task complexity, this study was able to employ counterbalancing of modality and task complexity successfully and collected data from the same participants for all conditions (writing simple, writing complex, speaking simple and speaking complex). In fact, some of the studies refrained from conducting comparison analysis between writing and speaking and stated that having different participants in the two modes was an important limitation (Kuiken & Vedder, 2011; Tavakoli, 2014). Examining the language outputs of the same individuals can constitute an important part of the research design considering that the aim of the study is to investigate potentially different effects of task complexity in the two modes possibly stemming from different cognitive processes. Additionally, this study differs from the previous ones in terms of the participant profile. The data were collected from the 7th-grade students (ages between 11-13) in their school environment. Almost all the task complexity and modality studies involved adult learners or high school students as the youngest group. Therefore, representing an understudied age group constitutes another contribution of this study.

The effects of task modality on syntactic complexity showed that speaking had significantly higher phrasal complexity than writing while the results were statistically insignificant for other syntactic measures. These results led to the argument that phrasal complexity might be more sensitive to detecting differences between the two modes as pointed out by Norris and Ortega (2009) and Byrnes et al. (2010). The fact that phrasal complexity was found higher in speaking led to the conclusion that quantitative measures such as the ratio of words to units or clauses can be problematic because of different characteristics of language outputs in speaking and writing (Halliday, 2002). A qualitative investigation of clauses and units in writing and speaking can be a better account for syntactic complexity (Granfeldt, 2008). The effects of task modality on lexical complexity showed that written modality leads to significantly more lexically complex language production than speaking. This was attributed to the offline and self-paced nature of the written mode which can promote more careful lexical searches and elaboration of ideational content. Another reason was argued to stem from the visibility of the written output. This could provide more opportunities to avoid lexical repetitions and monitor word choices compared to speaking. The effects of task modality on accuracy revealed that unit accuracy was statistically higher in writing than in speaking. This was argued to stem from the different opportunities for conceptualization, formulation, and monitoring stages in writing compared to speaking (Gilabert et al., 2016; Manchon, 2014; William, 2012). The recursive nature of writing provides opportunities to revise the language that was produced at the Execution stage (Kellogg, 1996). The written modality was argued to allow learners to conduct more careful planning and linguistic encoding as well as monitoring than speaking.

The effects of task complexity on syntactic measures showed that phrasal complexity was higher in complex tasks while subordination complexity was found to be higher in simple tasks. The results for overall syntactic complexity, on the other hand, were found to be statistically insignificant. These results supported the aforementioned conclusion that overall complexity might not be a sensitive measure to reveal syntactic differences. Further, the measure of the mean length of a unit might not be sensitive enough to reveal differences in narrating the same number of pictures in both simple and complex. Consequently, potential changes in linguistic complexity might be more apparent at a clausal level in such a task design. Phrasal complexity results were concluded to provide support for the Cognition Hypothesis (Robinson, 2001a, 2011a) and the Limited Attention Capacity (Skehan, 1998, 2009) although the interpretation of phrasal complexity together with the accuracy results indicated support for the latter account. The subordination complexity was found to be higher in simple tasks which contradicts the cognitive TBLT theories (Robinson, 2001a, 2011a; Skehan, 1998, 2009). The reason was argued to stem from the task design. Simple tasks in this study included both foreground and background events. Such tasks are argued to promote the use of subordinate clauses while complex tasks consisted of random pictures requiring participants to form a story. Consequently, it might be possible that participants used more clauses per unit to deliver the given order of events possibly through coordinators. It was concluded that a deeper investigation of subordinate clauses that reveals the quality of syntactic complexity might be needed. The effects of task complexity on lexical complexity were found to be statistically insignificant. Consistency of this result with the previous studies led to the conclusion that care should be taken in the task designs when comparing them to reveal lexical differences (Tavakoli & Foster, 2008). Finally, the effects of task complexity on accuracy were found to be statistically insignificant. The interpretation of this result together with the phrasal complexity suggested a trade-off effect as learners tend to prioritize accuracy or complexity because of limited attention capacity (Skehan, 2009).

The interaction between task complexity and modality revealed some interesting results. A clear parallel increase between task complexity and task modality was detected for phrasal complexity and lexical complexity. This led to the conclusion that task complexity and modality affect phrasal and lexical complexity in the same manner. Unit accuracy, on the other hand, showed a statistically significant interaction stemming from simple tasks. Clausal accuracy results revealed a high likelihood of an interaction with a similar trend although the current sample did not detect one. These results brought the question of how learners choose to use their attentional resources freed by the availability of more time in writing. It was concluded that tasks with lower conceptualization and organizational demands might free up more time and resources for the linguistic encoding and monitoring of language output in writing while learners might choose to be engaged in conceptualization and content during tasks with high cognitive demands and therefore might have limited opportunities left for monitoring. This finding lays out the effects of learners' preferences and differences on task complexity and modality as well as the complex nature of task design.

Overall, this study shows clear evidence that task modality affects the dimensions of L2 performance. Task complexity constitutes a crucial variable as increases in task demands resulted in the enhancement of certain aspects of L2 performance. However, the critical finding of this study was that the effects of task complexity differ in writing and speaking and therefore task complexity should be evaluated and implemented considering modality. From a pedagogical perspective, the findings suggest that performing tasks with different cognitive demands in certain modes can offer learners excellent opportunities for exploiting different aspects of L2 performance. The results of the current study suggest that complex tasks performed in the written mode might promote exhaustive lexical searches in the upper limits of interlanguage while simple tasks performed in the written mode might encourage the control of interlanguage and attention to form. Further, speaking can also provide access and control of interlanguage, and sequencing different task designs in different modes could provide a mode-balanced TBLT curriculum and foster balance in interlanguage development and L2 performances. Consequently, the combined effects of task complexity and modality on language performance need more attention, and systematic and rigorous research to account for the differences in language performance and to further guide teachers, and curriculum developers in designing and sequencing tasks that can provide more balanced language learning opportunities.

6.2 Limitations and suggestions for future research

There are several limitations needed to be addressed in this study. Firstly, discussions of the results led to the conclusion that general measures might not be sensitive enough to capture changes in L2 performance, especially in writing and speaking. General measures were used following the argument that they are sensitive to reveal differences between groups and the variation on a larger scale (Foster & Skehan, 1996; Norris & Ortega, 2009). They were also widely used in previous studies. However, the results showed that examining the quality of L2 performance along with the quantitative measures might be needed, especially in the case of modality (Granfeldt, 2008). Byrnes and Manchon (2014) further problematize the use of identical measures for examining writing and speaking. In light of these arguments and the results of the current study, the use of specific measures revealing the quality of syntactic complexity and tapping into different linguistic characteristics of writing and speaking is suggested for future empirical work. Additionally, phrasal complexity findings suggested that there might be a modality preference. The mode in which participants felt more comfortable was not investigated in this study. Another potential effect of individual differences was found in interaction findings. It was argued that how learners choose to use their attentional resources might need to be considered when analyzing the combined effects of task complexity and modality on L2 performance. Therefore, future studies could employ surveys or stimulated recall protocols to reveal and examine the effects of individual differences and modality on L2 performance.

Secondly, the existence or absence of a storyline in a narrative was argued to pose different levels of task complexity in the Methodology section. Although the operationalization of task complexity fits well with the necessary theoretical grounds (Kormos, 2006; Levelt, 1989, 1993; Robinson, 2001a, 2011a; Skehan, 1998, 2009), the results for subordinate complexity revealed a sensitivity toward having foreground and background events in simple tasks. The simple tasks in this study were discussed to promote the use of subordinate clauses due to the combination of foreground and background events in pictures (Tavakoli & Foster, 2008). Consequently, future studies should address the foreground and background events when operationalizing task complexity in the continuum of the existence of a story plot and organization of information. When comparing narrative complex tasks consisting of random pictures and no storyline with narrative simple tasks presented in an order of a story, care should be given to provide narratives with only foreground events for simple tasks.

Lastly, lexical complexity results suggested a possible problem with cross-task comparisons. Different visual stimuli in simple and complex tasks might result in different vocabulary choices and therefore making cross-task comparisons for narrative tasks challenging (Tavakoli & Foster, 2008). Therefore, different task designs such as decision-making and interactional might be a better choice for task comparisons for lexical complexity, especially when assessed through some form of TTR. On this subject, future research can also include different lexical measures such as lexical sophistication that are possibly more sensitive to capture changes in lexis between simple and complex tasks (Skehan, 2009).

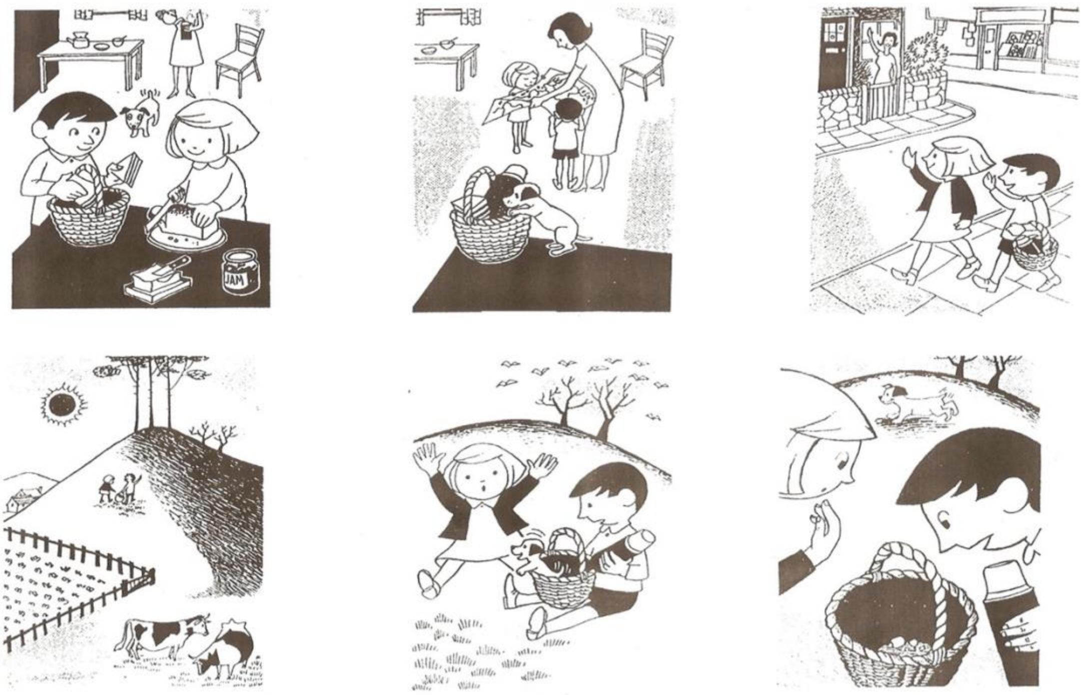
This study examined the immediate effects of task complexity and modality on L2 performance. However, there is more to discover about the potential learning benefits of the two modes when combined with task complexity. These benefits can be more apparent in a longitudinal and interventional study, especially at a curriculum level. More research is needed to better understand how speaking and writing interact with task complexity and feed one another to foster interlanguage development.

APPENDIX A

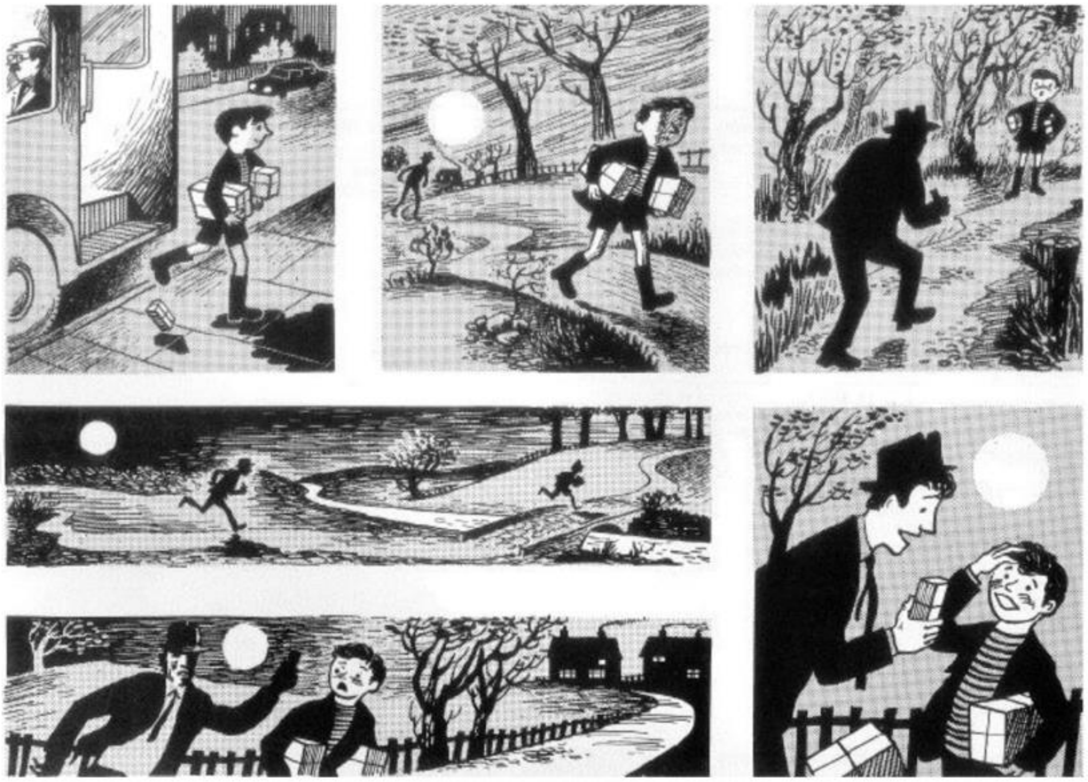
PICTURES USED IN THE STUDY

Simple Tasks

Picture 1: Dog & the Basket



Picture 2: Lost and Found



Complex Tasks

Picture 1: Unrelated Pictures

Picture 2: Unrelated Pictures

APPENDIX B

TASK INSTRUCTIONS (IN ENGLISH)

1. Task Instructions for Oral Simple Tasks

In this task, you have 6 pictures. Pictures present a storyline and they are ordered from 1 to 6 in the order of the story. You are asked to tell the story presented in these pictures. You should consider the order of the pictures while telling your story and include ALL the pictures.

- You have 3 minutes to look at the pictures and plan what you are going to say. You will be asked to tell your story at the end of 3 minutes.
- You are allowed to use the paper given for the task for planning your story.
- Please look at the pictures and the details very carefully.
- You have 40 minutes to tell your story.

If you have any questions, you should ask them before your task is handed to you. After the task starts, you should not ask any questions or use dictionaries.

2. Task Instructions for Written Simple Tasks

In this task, you have 6 pictures. Pictures present a storyline and they are ordered from 1 to 6 in the order of this story. You are asked to write the story presented in these pictures. You should consider the order of the pictures while writing your story and include ALL the pictures.

- You have 3 minutes to look at the pictures and plan what you are going to write. You will be asked to start writing your story at the end of 3 minutes.
- You are allowed to use the paper given for the task for planning your story.
- Please look at the pictures and the details very carefully.
- You have 40 minutes to write your story.

If you have any questions, you should ask them before your task is handed to you. After the task starts, you should not ask any questions or use dictionaries.

3. Task Instructions for Oral Complex Tasks

In this task, you have 6 pictures. These pictures are unrelated and randomly ordered. They do not present or suggest any story. You are asked to form and tell a story by including ALL the pictures.

- You have 3 minutes to look at the pictures and plan what you are going to say. You will be asked to tell your story at the end of 3 minutes.
- You are allowed to use the paper given for the task for planning your story.

- Please look at the pictures and the details very carefully.
- You have 40 minutes to tell your story.

If you have any questions, you should ask them before your task is handed to you. After the task starts, you should not ask any questions or use dictionaries.

4. Task Instructions for Written Complex Tasks

In this task, you have 6 pictures. These pictures are unrelated and randomly ordered. They do not present or suggest any story. You are asked to form and write a story by including ALL the pictures.

- You have 3 minutes to look at the pictures and plan what you are going to write. You will be asked to write your story at the end of 3 minutes.
- You are allowed to use the paper given for the task for planning your story.
- Please look at the pictures and the details very carefully.
- You have 40 minutes to tell your story.

If you have any questions, you should ask them before your task is handed to you. After the task starts, you should not ask any questions or use dictionaries.

APPENDIX C

EXAMPLE AS-UNIT CODING

1. Sample Transcription Before AS-unit Coding

Anna and Jack uh was get reading uh to go picnic because it was very sunny day. And while they were get ready uh dog named Chase uh wanted to also go to picnic, to park. And then he go like he go, open the basket. And after Jack and Anna uh while Jack and Anna go wait to uh [long pause] they get ready to go park. Then they were at the park. When they get hungry, they opened the basket and they they saw the Chase. And then Chase go uh go out and run away. Then they want to eat some snacks. Wh – when they saw there is no snack, they that uh they understand that Chase eat the snacks.

2. Sample Transcription After AS-unit Coding

|Anna and Jack {uh} was get reading:: {uh} to go picnic|

|Because it was very sunny day|

| {And} while they were get ready:: {uh} dog named Chase:: {uh} wanted:: to also go to picnic, to park|

| {And} then {he go like} he go:: open the basket|

| {And after Jack and Anna uh while Jack and Anna go wait to uh [long pause]} they get ready:: to go park|

| Then they were at the park|

|When they get hungry:: they opened the basket|

|and {they} they saw the Chase|

| {And} then Chase {go uh} go out:: and run away|

|Then they want:: to eat some snacks|

|{Wh} when they saw:: there is no snack:: {they that uh} they understand:: that Chase eat the snacks|

Notes.

i. |…| is used to show AS-unit boundaries

ii. :: show the clauses

iii. {…} is used to show omitted parts which are repetitions, false starts, reformulation and replacement.

3. Pruned Narrative

|Anna and Jack was get reading:: to go picnic|

|Because it was very sunny day|

| while they were get ready:: dog named Chase:: wanted:: to also go to picnic, to park|

| then he go:: open the basket|

| they get ready:: to go park|

| Then they were at the park|

|When they get hungry:: they opened the basket|

|and they saw the Chase|

| then Chase go out:: and run away|

|Then they want:: to eat some snacks|

|when they saw:: there is no snack:: they understand:: that Chase eat the snacks|

APPENDIX D

AS-UNIT CODING GUIDELINES

Please follow the instructions below to code the data.

Note: Brackets {…} in the examples below are used to mark disfluency markers which are repetitions, false starts, reformulation, and replacement.

1. What is an AS-unit?

AS-unit includes an independent clause which contains a finite verb. It can also include subordinate clauses associated with the independent clause. These subordinate clauses can include a finite or a non-finite verb.

Please mark each AS-unit with an upright slash at the beginning and at the end as |…|

Example:

Sample text before AS-unit parsing:

- ☐ When they get hungry, they opened the basket and {they} they saw the Chase and then Chase {go uh} go out and run away

Sample text after AS-unit parsing:

- ☐ |When they get hungry:: they opened the basket|

  |and {they} they saw the Chase|

  | {And} then Chase {go uh} go out:: and run away|

2. Clauses

Minimum elements consisting of a clause are a finite verb or a non-finite verb together with at least one other clause element (i.e., subject, object, complement, or adverbial).

Mark clause boundaries with a double colon (::)

2.1 Finite Verb

2.1.1 Coordination: Separate AS-unit boundaries should be applied to coordinated main clauses.

Examples:

- ☐ *It was basically a night time so he couldn't see the man's face properly*

  |It was basically a night time|

  |so he couldn't see the man's face| (2 AS-units)

- ☐ *he was making a art project then he was carrying boxes*

  |he was making a art project|

  |then he was carrying boxes| (2 AS-units)

Coordination of verb phrases should be considered to be part of the same AS-unit unless there is a clear noticeable pause of at least 0.5 seconds preceded by a falling or rising intonation. Then, the subsequent verb phrase is a new start and an AS-unit.

Example:

- ☐ *Later the man chased after and touched the kid and told that he dropped his luggage*

  |Later the man chased after:: and touched the kid:: and told:: that he dropped his luggage| (1 AS-unit, 4 clauses)

2.1.2 Subordination

Subordinated clauses should be coded in the same AS-unit.

Examples:

- *When he was going to {her} his house {umm} he dropped one of his boxes*

  |When he was going to his house:: he dropped one of his boxes| (1 AS-unit, 2 clauses)

- *Jack found out that he had actually lost a box*

  |Jack found out:: that he had actually lost a box| (1 AS-unit, 2 clauses)

Because constitutes an exception in AS-units. Normally, it is treated as part of the same AS-unit.

Example:

- *he got the goosebumps because he heard some footsteps*

  |he got the goosebumps:: because he heard some footsteps| (1 AS-unit, 2 clauses)

WITH THE EXCEPTION

In oral data, because can function as an ellipted version of an independent clause. It can mean “I say this because…”. When it functions as an independent clause, it should be coded as a separate AS-unit.

Example:

- *he got scared because everybody gets scared in that situation*

  |he got scared |

  |because everybody gets scared in that situation| (2 AS-units)

2.1.3 Embedding

Embedded clauses should be considered to be part of the same AS-unit but considered as separate clauses.

Example:

- *Once upon a time there was a boy and a girl planning to go to a picnic together*

  |Once upon a time there was a boy and a girl:: <u>planning:: to go to a picnic together</u>| (1 AS-unit, 3 clauses)

- *There was a man who wanted to get engaged to his fiancé*

  |There was a man:: <u>who wanted to get engaged to his fiancé</u>| (1 AS-unit, 2 clauses)

2.2 Non-Finite Verb together with at least one other clause element (Subject, Object, Complement, or Adverbial)

Non-finite clausal analysis requires at least one additional clause element such as subject, object, complement, or adverbial to be counted as a clause.

Examples:

- *He started to ran*

  |He started to ran| (1 AS-unit)

  (no separate clause here because "ran" doesn't contain any other element to be counted as a clause)

- *They just continued walking*

  |they just continued walking| (1 AS-unit)

- *Then they want to eat some snacks*

  |Then they want:: to <u>eat some snacks</u>| (1 AS-unit, 2 clauses)

(“to eat some snacks” is counted as a separate clause in this example because “to eat” is followed by an object.)

- ☐ *He loved watching boats*

|He loved:: watching boats| (1 AS-unit, 2 clauses)

(“watching boats” is counted as a separate clause in this example because “watching” is followed by an object.)

3. False Starts, Repetitions and Self-Corrections

The final version of the data should be counted and the previous versions should be excluded.

4. Topicalization

Topicalization is commonly found among second language learners especially when their first language is a topic-comment language (Foster et al., 2000). For this reason, topicalized noun phrases should be included in the AS-unit except when they are clearly separated by a pause or falling intonation.

Example:

- ☐ |One day a child he saw an unoccupied boat| (1 AS-unit)

APPENDIX E

T-UNIT CODING GUIDELINES

1. What is a T-unit?

A T-unit refers to a minimal terminable unit that includes at least one main clause but can also have subordinate clauses, phrases, and words attached to it or embedded in it (Hunt, 1970).

Please mark each T-unit with an upright slash at the beginning and at the end as |…|

2. Clauses

Minimum elements consisting of a clause are a finite verb or a non-finite verb together with at least one other clause element (i.e., subject, object, complement, or adverbial).

Mark clause boundaries with a double colon (::)

2.1 A finite verb

2.1.1 Coordination: Separate T-unit boundaries should be applied to coordinated main clauses except for the coordination of verb phrases.

- *There was a big storm and lightnings were falling*

  |There was a big storm|

  |and lightnings were falling| (2 T-units)

- *She let out a relieved sight when she reached the end of the street and passed the island*

|She let out a relieved sight:: when she reached the end of the street:: and passed the island| (1 T-unit, 3 clauses)

2.1.2 Subordination: Subordinated clauses should be coded in the same T-unit.

- *She thought that this was the moment*

  |She thought:: that this was the moment| (1 T-unit, 2 clauses)

- *before I can see what's at the end of the hall, I see two locations*

  |before I can see:: what's at the end of the hall:: I see two locations| (1 T-unit, 3 clauses)

2.1.3 Embedding: Embedded clauses should be considered to be part of the same T-unit.

- *Once upon a time there lived a guy whose name was Dumbo*

  |Once upon a time there lived a guy:: whose name was Dumbo| (1 T-unit, 2 clauses)

- *I found an old book that had a map in it*

  |I found an old book:: that had a map in it| (1 T-unit, 2 clauses)

3. Non-Finite Verb together with at least one other clause element (Subject, Object, Complement, or Adverbial): They should be coded in the same T-unit. Non-finite clausal analysis requires at least one additional clause element such as subject, object, complement or adverbial to be counted as a clause.

- *The man started following*

  |The man started following| (1 T-unit)

- *I decided to search for the treasure*

|I decided:: to search for the treasure| (1 T-unit, 2 clauses)

("to search for the treasure" is counted as a separate clause in this example because "search for" is followed by an object.")

- *I started swimming to the island*

|I started:: swimming to the island| (1 T-unit, 2 clauses)

("swimming to the island" is counted as a separate clause in this example because "swimming" is followed by an adverbial phrase.)

APPENDIX F

CALCULATIONS OF MANUAL MEASURES

Manual Measures in the Study

Measure	Explanation	Example
Overall Syntactic Complexity	Mean length of AS-unit or T-unit: Divide the number of words by the number of AS-units or T-units	|It was dark outside| |He was carrying the new stationary:: he bought for his new school year| |Just when he was walking towards his way home:: he hears footsteps:: heading to his direction| 33 words/ 3 T-units= 11
Subordination Complexity	The number of clauses per AS-unit/T-unit: Divide the total number of clauses by the total number of AS-units or T-units	|It was dark outside| |He was carrying the new stationary:: he bought for his new school year| |Just when he was walking towards his way home:: he hears footsteps:: heading to his direction| 6 clauses/ 3 T-units = 2
Phrasal Complexity	Mean length of clause: Divide the number of words by the number of clauses	|It was dark outside| |He was carrying the new stationary:: he bought for his new school year| |Just when he was walking towards his way home:: he hears footsteps:: heading to his direction| 33 words/ 6 clauses= 5.5
Unit Accuracy	The ratio of error-free AS-unit/T-unit: Divide total number of error-free AS-units or T-units by the total number of AS-units or T-units	|It was dark outside| |He was carrying the new stationary:: he bought for his new school year| |Just when he was walking towards his way home:: he hears footsteps:: heading to his direction| 2 units without errors/ 3 T-units total= 0.66
Clausal Accuracy	The ratio of error-free clauses to total clauses: Divide total number of error-free clauses by the total number clauses	|It was dark outside| |He was carrying the new stationary:: he bought for his new school year| |Just when he was walking towards his way home:: he hears footsteps:: heading to his direction| 5 clauses without errors/ 6 clauses total= 0.83

APPENDIX G

INTERCODER RELIABILITY SCORES

Reliability Scores

Measures	N	Pearson Correlation Coefficient *r*
Overall Syntactic Complexity	32	.97
Subordination Complexity	32	.96
Phrasal Complexity	32	.97
Unit Accuracy	32	.95
Clausal Accuracy	32	.98

Note. Oral data calculations are based on pruned narratives.

APPENDIX H

ETHICS COMMITTEE APPROVAL

Evrak Tarih ve Sayısı: 10.03.2022-57341

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ
SOSYAL VE BEŞERİ BİLİMLER YÜKSEK LİSANS VE DOKTORA TEZLERİ ETİK İNCELEME KOMİSYONU
TOPLANTI KARAR TUTANAĞI

Toplantı Sayısı	:	28
Toplantı Tarihi	:	03.03.2022
Toplantı Saati	:	11:30
Toplantı Yeri	:	Zoom Sanal Toplantı
Bulunanlar	:	Prof. Dr. Ebru Kaya, Prof. Dr. Nevra Seggie, Dr. Öğr. Üyesi Yasemin Sohtorik İlkmen
Bulunmayanlar	:	

Tuğçe Betül Aktaş Smith
İngilizce Öğretmenliği

Sayın Araştırmacı,
"Basit ve Karmaşık Bilişsel Görevlerde Yazma ve Konuşma Görevi Tekrarı ile Öğrencilerin Dil Öğreniminin Ölçülmesi" başlıklı projeniz ile ilgili olarak yaptığınız SBB-EAK 2022/14 sayılı başvuru komisyonumuz tarafından 3 Mart 2022 tarihli toplantıda incelenmiş ve uygun bulunmuştur.

Bu karar tüm üyelerin toplantıya çevrimiçi olarak katılımı ve oybirliği ile alınmıştır. COVID-19 önlemleri kapsamında kurul üyelerinden ıslak imza alınamadığı için bu onay mektubu üye ve raportör olarak Yasemin Sohtorik İlkmen tarafından bütün üyeler adına e-imzalanmıştır.

Saygılarımızla, bilgilerinizi rica ederiz.

Dr. Öğr. Üyesi Yasemin
SOHTORİK İLKMEN
Öğretim Üyesi

e-imzalıdır
Dr. Öğr. ÜyesiYasemin Sohtorik
İlkmen
Öğretim Üyesi
Raportör

SOBETİK 28 03.03.2022

Bu belge, güvenli elektronik imza ile imzalanmıştır.

REFERENCES

Baddeley, A.D. (1986). *Working memory.* Oxford, U.K.: Clarendon.

Byrnes, H., & Manchon, R. M. (2014). Task-based language learning: Insights from and for L2 writing - an introduction. In H. Byrnes & R. M. Manchon (Eds.), *Task-based language learning-insights from and for L2 writing* (Vol. 7; pp. 1-27). Amsterdam: John Benjamins.

Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *The Modern Language Journal, 94,* 1-235. http://www.jstor.org/stable/40985261

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics, 13*(1), 1-24. https://doi.org/10.1093/applin/13.1.1

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of past tense. *Studies in Second Language Acquisition, 9*(1), 1-19. https://doi.org/10.1017/S0272263100006483

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16*(4), 409-441. https://doi.org/10.1093/applin/16.4.409

Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3-37). Amsterdam: John Benjamins.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford University Press.

Ellis, R., Skehan, P., Li, A., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Ellis, R. & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167-92). Amsterdam: John Benjamins.

Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299-323. https://doi.org/10.1017/S0272263100015047

Foster, P., & Skehan, P. (1999). The influence of task structure and processing conditions on narrative tellings. *Language Learning, 49*(1), 93-120. https://doi.org/10.1111/1467-9922.00071

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning, 59*(4), 866-896. https://doi.org/10.1111/j.1467-9922.2009.00528.x

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354-375. https://doi.org/10.1093/applin/21.3.354

Franz, V. H., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review, 19*(3), 395-404. https://doi.org/10.3758/s13423-012-0230-1

Games, P. A., & Lucas, P. A. (1966). Power of the analysis of variance of independent groups on non-normal and normally transformed data. *Educational and Psychological Measurement*, *26*(2), 311-327. https://doi.org/10.1177/001316446602600205

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, *45*(3), pp. 215-240. https://doi.org/10.1515/iral.2007.010

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. https://doi.org/10.1017/S0267190515000112

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237-288. https://doi.org/10.3102/00346543042003237

Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In M. Torrance, L. van Waes & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 165-179). Amsterdam: Elsevier.

Granfeldt, J. (2008). Speaking and writing in L2 French: Exploring effects on fluency, accuracy and complexity. In A. Housen, S. van Daele, & M. Pierrard (Eds.), *Proceedings of the conference on complexity, accuracy and fluency in second language use, learning and teaching* (pp. 87-98). Brussels, Belgium: Koninklijke Vlaasmse Academi.

Halliday, M. A. K. (2002). Spoken and written modes of meaning. In J. J. Webster (Ed.), *On grammar* (pp. 323-351). London: Continuum.

Halliday, M. A. K., & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition.* Cassell.

Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Randall (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Erlbaum.

Hayes, J. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369-388. https://doi.org/10.1177/0741088312451260

Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3-30). Hillsdale, NJ: Erlbaum.

Heaton, J. B. (1966). *Composition through pictures*. Essex: Longman.

Heaton, J. B. (1975). *Beginning composition through pictures*. London: Longman.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461-473. https://doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1-20). Amsterdam: John Benjamins.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57-71). Mahwah, NJ: Erlbaum.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.

Kormos, J. (2011a). Speech production and the cognition hypothesis. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 39-60). Amsterdam: John Benjamins.

Kormos, J. (2011b). Task complexity and linguistic and discourse features of narrative writing performance. *Language Learning*, *20*(2), 148-161. http://dx.doi.org/10.1016/j.jslw.2011.02.001

Kormos, J. (2014). Differences across modalities of performance: An investigation of linguistic and discourse complexity in narrative tasks. In H. Byrnes & R. M. Manchon (Eds.), *Task-based language learning-insights from and for L2 writing* (pp. 193-217). John Benjamins Publishing Company.

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 267-289). Amsterdam: John Benjamins.

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance: Task complexity and aptitude. *Language Learning*, *62*(2), 439-472. https://doi.org/10.1111/j.1467-9922.2012.00695.x

Kuiken, F. & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics, 45*(3), 261-284. https://doi.org/10.1515/iral.2007.012

Kuiken, F. & Vedder, I. (2011). Task performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 91-104). John Benjamins Publishing Company.

Kuiken, F., & Vedder, I. (2012). Speaking and writing tasks and their effects on second language performance. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 364-379). New York: Routledge.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590-619. https://doi.org/10.1093/applin/aml029

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall, & C-W. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1-15). Berlin: Walter de Gruyter.

Long, M. H. (1985). A role for instruction in second language acquisition. In K. Hyltenstam and M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77-99). Clevedon, Avon: Multilingual Matters.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Oxford: Wiley-Blackwell.

Long, M. H., & Crookes, G. (1991). Three approaches to task-based syllabus design. *University of Hawai'i Working Papers in ESL*, *10*(1), 1-36.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36-62. http://www.jstor.org/stable/41307615

Lunney, G. H. (1970). Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement, 7*(4), 263-269. https://doi.org/10.1111/j.1745-3984.1970.tb00727.x

Malvern, D. D., & Richards, B. J. (1997). A measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon, UK: Multilingual Matters.

Manchon, R. M. (2014). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R. M. Manchon (Eds.), *Task-based language learning-insights from and for L2 writing* (Vol. 7; pp. 27-53). Amsterdam: John Benjamins.

Manchon, R. M., & Williams, J. (2016). L2 writing and SLA studies. In R. M. Manchon & P. K. Matsuda (Eds.), *The handbook of second and foreign language writing* (pp. 567-586). Boston: de Gruyter.

McCarthy, M. P., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488. https://doi.org/10.1177/0265532207080767

Meara, P., & Miralpeix, I. (2016). D_Tools v2.0. *Tools for researching vocabulary* (pp. 21-44). Multilingual Matters.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578. https://doi.org/10.1093/applin/amp044

Nunan, D. (1993). *Introducing discourse analysis*. London: Penguin.

Nunan, D. (2001). Second language acquisition. In R. Carter and D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 87-92). Cambridge: Cambridge University Press.

Nunan, D. (2004). *Task-based language teaching.* Cambridge University Press.

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77-111). Amsterdam: John Benjamins.

Ortega, L. (2012). Epilogue: Exploring L2 writing – SLA interfaces. *Journal of Second Language Writing*, *21*(4), 404-415. http://dx.doi.org/10.1016/j.jslw.2012.09.002

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*(4), 590-601. https://doi.org/10.1093/applin/amp045

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language, 29*(02), 417-447. doi:10.1017/S0305000902005111

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge: Cambridge University Press.

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

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*(2), 45-105.

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

Robinson, P. (2007a). Criteria for grading and sequencing pedagogic tasks. In M.P. Garcia Mayo (Ed.), *Investigating tasks in formal language learnin*g (pp. 7-27). Clevedon: Multilingual Matters.

Robinson, P. (2007b). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching, 45*(3), 193-213. https://doi.org/10.1515/iral.2007.009

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243-268). Amsterdam: John Benjamins Publishing.

Robinson, P. (Ed.). (2011a). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. John Benjamins Publishing Company.

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

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching, 45*(3), 161-176. https://doi.org/10.1515/iral.2007.007

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge: Cambridge University Press.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition*, (pp. 237-369). Rowley, MA: Newbury House.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*(1), 38-62. https://doi.org/10.1093/applin/17.1.38

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

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

Skehan, P. (2014). Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspective on task performance* (pp. 211-260). John Benjamins Publishing Company.

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

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge: Cambridge University Press.

Stauble, A. (1978). The process of decreolization: A model for second language development. *Language Learning*, *28*(1), 29-54. https://doi.org/10.1111/j.1467-1770.1978.tb00303.x

Swain, M. **(**1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development: In S. Gassand, & C. Madden (Eds.), *Input and second language acquisition* (pp. 235-253). Rowley, MA Newbury House.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, *50*(1), 158-164. https://doi.org/10.3138/cmlr.50.1.158

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371-391. https://doi.org/10.1093/applin/16.3.371

Swain, M. & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research, 37*, 285-304. https://doi.org/10.1016/S0883-0355(03)00006-5

Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks In H. Byrnes & R. M. Manchon (Eds.), *Task-based language learning-insights from and for L2 writing* (pp. 193-217). John Benjamins Publishing Company.

Tavakoli, P. & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439-473. https://doi.org/10.1111/j.1467-9922.2008.00446.x

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-273). Amsterdam: John Benjamins.

Tuzcu, A. (2019). *Time perspective and online planning as mediating factors in L2 oral performance: A quasi-experimental study* [Unpublished master's thesis]. Boğaziçi University.

Tuzcu, A., & Yalçın, Ş. (2020). The effects of task condition on L2 oral performance: Exploring complexity measures. *Boğaziçi University Journal of Education, 37*(1), 25-42. https://dergipark.org.tr/tr/pub/buje/issue/55539/760473

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*(3), 287-301. https://doi.org/10.1017/S0272263100009177

Vasylets, O. (2017). *Task-modality effects: A study of task complexity effects in speech and writing* [Doctoral dissertation, Universitat de Barcelona]. TDX Repository. http://hdl.handle.net/10803/404299

Vasylets, O., Gilabert, R., & Manchón, R. M. (2017). The effects of mode and task complexity on second language production: Mode and task complexity. *Language Learning, 67*(2), 394-430. https://doi.org/10.1111/lang.12228

Vasylets, O., Gilabert, R., & Manchón, R. M. (2019). Differential contribution of oral and written modes to lexical, syntactic and propositional complexity in L2 performance in instructed contexts. *Instructed Second Language Acquisition, 3*(2), 206-227. https://doi.org/10.1558/isla.38289

Weissberg, B. (2000). Developmental relationships in the acquisition of English syntax: Writing vs. speech. *Learning and Instruction, 10*(1), 37-53. https://doi.org/10.1016/S0959-4752(99)00017-1

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159-177. https://doi.org/10.1080/14639220210123806

Wickens, C. (2007). Attention to the second language. *International Review of Applied Linguistics, 45*(3), 177-191. https://doi.org/10.1515/iral.2007.008

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, *21*(4), 321-331. http://dx.doi.org/10.1016/j.jslw.2012.09.007

Willis, D. & Willis, J. (2001). Task-based language learning. In R. Carter and D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 173-179). Cambridge: Cambridge University Press.

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236-259. https://doi.org/10.1093/applin/amp024

Zalbidea, J. (2017). 'One task fits all'? The roles of task complexity, modality, and working memory capacity in L2 performance. *The Modern Language Journal*, *101*(2), 335-352. https://doi.org/10.1111/modl.12389