

INVESTIGATING THE INTERACTIONS AMONG GENRE, TASK COMPLEXITY, AND  
PROFICIENCY IN L2 WRITING: A COMPREHENSIVE TEXT ANALYSIS AND STUDY OF  
LEARNER PERCEPTIONS

By

Hyung-Jo Yoon

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Second Language Studies—Doctor of Philosophy

2017

## ABSTRACT

### INVESTIGATING THE INTERACTIONS AMONG GENRE, TASK COMPLEXITY, AND PROFICIENCY IN L2 WRITING: A COMPREHENSIVE TEXT ANALYSIS AND STUDY OF LEARNER PERCEPTIONS

By

Hyung-Jo Yoon

In this study, I explored the interactions among genre, task complexity, and L2 proficiency in learners' writing task performance. Specifically, after identifying the lack of valid operationalizations of genre and task dimensions in L2 writing research, I examined how genre functions as a task complexity variable, and how learners' perceptions and language production interact with their proficiency. In exploring ESL students' perceptions and production of different writing tasks, I used the two genres of narrative and argumentative writing, within which I manipulated the level of task complexity operationalized as idea support (e.g., narrative task with supporting ideas is the simple narrative task). I collected essay data from 76 ESL students. Each student wrote four essays (i.e., a total of 304 essays). Immediately after each writing session, the students showed their perceptions of a task in terms of six dimensions (task complexity, difficulty, anxiety, confidence, interest, and motivation). Additionally, I collected perception data from 30 ESL instructors with regard to how their students at a proficiency level similar to that of the student participants would perform the target writing tasks. In so doing, I could compare students' perceptions with teachers' expectations of how the tasks would function.

From the task perception result, I found a gap between the student and teacher groups regarding their views of the two genres. Specifically, the teachers predicted that ESL students would have greater difficulty in completing the argumentative genre than the narrative, but

instead the students perceived both genres as involving a similar level of complexity and difficulty. Also, unlike teachers' expectations, students consistently judged the tasks with idea support as less complex and less difficult. One common result from both groups was their judgments of the narrative genre as sparking greater interest and motivation for further writing than the argumentative.

The writing result showed that the students' language varied to a greater extent across the two genres but not across the idea support conditions. I also found that most linguistic features did not differ by L2 proficiency. This result suggests that there is a very weak link between writers' task perceptions and language production, challenging the common practice of task-based writing research. Therefore, this result points to the importance of exploring these two different result types separately in written discourse because writers' language changes are largely motivated by varying communicative functions of different genres but not by a task's cognitive constraints imposed on writers.

The result of essay quality scores demonstrated that narrative essays tended to receive higher scores than argumentative essays in terms of discourse-level categories, and that there were significant interaction effects between genre and idea support. Specifically, argumentative essays composed with supporting ideas resulted in higher scores, whereas narrative essays with supporting ideas led to lower scores. Unlike the result of linguistic features, with L2 proficiency as an additional variable, the result showed that higher proficiency ESL students are likely to receive higher scores on sentence-level categories. This study offers implications for L2 writing research, pedagogy, and assessment. Particularly, L2 writing instructors and task developers will be informed about the possibility of constructing independent writing tasks with various genres and task complexity to achieve an appropriate alignment of task features with target L2 learners.

I dedicate this work to my parents.

## ACKNOWLEDGMENTS

I could not have finished this project without the support of many people around me. First, I would like to express my sincere thanks to Dr. Charlene Polio for her support and encouragement over the course of my Ph.D. studies. She has been a great mentor for my studies as well as for my life. Her constant passion for improvement as a researcher taught me how I can enjoy my life as a researcher.

I am deeply thankful to each of my dissertation committee members. Dr. Paula Winke has provided me with valuable advice on how to write a research paper with a professional tone. Dr. Shawn Loewen has constantly taught me the importance of having a keen understanding of statistics as an applied linguist. Dr. Aline Godfroid has equipped me with theoretical knowledge and practical skills necessary for conducting a decent SLA study.

I also thank instructors and teaching assistants at the English Language Center who allowed me to collect data in their classes. Thanks to their generous permission, I could finish the stage of data collection with little difficulty. Also, I am grateful to ESL instructors and students who participated in my project. I would like to extend my gratitude to my friends in the Second Language Studies program. I could have a relatively stress-free life thanks to their support, and I will never forget the time we spent together. Most importantly, I would like to say that I could focus on my studies thanks to my parents in South Korea who have always supported me. Their love and trust have been a great driving force for me.

This project was funded by The International Research Foundation (TIRF), the National Federation of Modern Language Teachers' Association, and the College of Arts and Letters at Michigan State University.

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
CHAPTER 1. INTRODUCTION .....	1
CHAPTER 2. LITERATURE REVIEW .....	12
Definitions of Genre and Other Related Terms .....	12
Cross-genre L1 Studies .....	16
Cross-genre L2 research .....	20
Task-based Writing Studies .....	25
Task-based Studies with Cross-genre Manipulations .....	29
Validation of Task Complexity Manipulations .....	33
Text Analysis in TBLT Research .....	35
Rationale for the Present Study .....	38
CHAPTER 3. METHOD .....	42
Participants .....	42
Student participants .....	42
Teacher participants .....	44
Instruments .....	45
Questionnaires .....	45
Writing prompts .....	47
Rubric .....	48
Procedures .....	49
Data collection .....	49
Essay scoring .....	52
Text Features .....	54
Syntactic complexity features .....	55
Lexical features .....	57
Discourse features .....	58
Interactional metadiscourse features .....	59
Analysis .....	62
CHAPTER 4. RESULTS .....	64
Task Perceptions .....	64
Textual Feature Changes across Task Types .....	78
Interplay of L2 Proficiency and Task Manipulations Influencing Textual Features .....	88

Essay Score Changes across Task Types .....	97
Interplay of L2 proficiency and Task Manipulations Influencing Essay Scores .....	101
CHAPTER 5. DISCUSSION.....	106
ESL Students’ and Teachers’ Perceptions of Writing Tasks.....	106
Effects of Task Type on Textual Features .....	109
Effects of Task Type on Essay Quality .....	114
CHAPTER 6. CONCLUSION.....	119
Theoretical and Research Implications .....	119
Pedagogical and Assessment Implications.....	121
Limitations and Future Research .....	123
APPENDICES .....	126
Appendix A. Writing Prompts.....	127
Appendix B. Revised Analytic Scoring Rubric .....	129
Appendix C. Cloze Test .....	130
Appendix D. Example Essays.....	134
REFERENCES .....	139

## LIST OF TABLES

Table 1. Taxonomy of Genre in This Study .....	15
Table 2. Demographic Characteristics of the ESL Student Participants .....	44
Table 3. Counterbalanced Data Collection Procedures .....	50
Table 4. Demographic Characteristics of the High and Low Proficiency Group Students .....	52
Table 5. Target Text Features .....	61
Table 6. Descriptive Statistics for ESL Students' and Teachers' Perceptions of Writing Tasks ...	66
Table 7. Interaction Effects of Genre, Idea Support, and Group on Task Perceptions .....	67
Table 8. Post-hoc Analysis Results of Genre and Idea Support Effects for Each Group's Perceptions .....	67
Table 9. Main Effects of Genre, Idea Support, and Group on Task Perceptions .....	72
Table 10. Correlations between Perception Items by Task Type .....	73
Table 11. Summary of Task Perception Results.....	77
Table 12. Descriptive Statistics for Target Text Features by Task Type .....	81
Table 13. Inferential Statistics for Genre and Idea Support Effects on Textual Features .....	82
Table 14. Interaction and Main Effects of L2 Proficiency on Textual Features .....	90
Table 15. Summary of Task Manipulation and L2 Proficiency Conditions with Significantly Higher Values of Textual Features.....	91
Table 16. Correlations of Perceived Task Complexity with Linguistic Complexity Features.....	92
Table 17. Descriptive Statistics for Essay Scores by Genre and Idea Support .....	98
Table 18. Inferential Statistics for Genre and Idea Support Effects on Essay Scores.....	98
Table 19. Descriptive Statistics for Essay Scores by L2 Proficiency, Genre, and Idea Support	102
Table 20. Interaction and Main Effects of L2 Proficiency on Textual Features .....	103



## LIST OF FIGURES

Figure 1. Students' and teachers' perceptions of task complexity and difficulty across genre conditions. ....	68
Figure 2. Interaction plots for perceived complexity and difficulty showing an interaction between genre and idea support only for teacher perceptions.....	69
Figure 3. Students' and teachers' perceptions of task confidence, interest, and motivation across genre conditions. ....	71
Figure 4. Interaction plots for complex nominals per clause, modifiers per noun phrase, and nominalization density showing an interaction between genre and idea support conditions. ....	80
Figure 5. Complex nominals per clause and modifiers per noun phrase across genre and idea support conditions. ....	84
Figure 6. Temporal connective density and self-mention density across genre and idea support conditions. ....	85
Figure 7. Nominal clause density, adverbial clause density, and adjectival clause density across genre and idea support conditions. ....	87
Figure 8. Interaction plots for content, organization, and language use scores showing an interaction between genre and idea support conditions.....	99
Figure 9. Content and organization scores across genre and idea support conditions.....	100
Figure 10. Vocabulary and language use scores across task types and L2 proficiency. ....	104

## **CHAPTER 1.**

### **INTRODUCTION**

Much first language (L1) and second language (L2) writing research has investigated the cognitive processes involved in writing and has provided important suggestions on how writers deal with different stages of writing that place varying demands on their limited cognitive resources (e.g., Beauvais, Olive, & Passerault, 2011; Chenoweth & Hayes, 2001; Hayes & Chenoweth, 2006; Manchón & Roca de Larios, 2007; Olinghouse & Graham, 2009; Quinlan, Loncke, Leijten, & Van Waes, 2012, among many others). Specifically, drawing on cognitively-oriented writing models (e.g., Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1996), L1 and L2 writing studies have attempted to explore how writers' knowledge and memory resources interact with the task environment (e.g., Barkaoui, 2016; Johansson, Wengelin, Johansson, & Holmqvist, 2010; Johnson, Mercado, & Acevedo, 2012; Kormos, 2011; Leijten & Van Waes, 2013; Wengelin et al., 2009).

While benefiting greatly from the suggestions of L1 writing studies, due to fundamental differences between L1 and L2 writers (e.g., age of acquisition, language proficiency, amount of input, and educational experience), L2 writing research began to establish its own ground by testing its empirical findings against L2-specific frameworks such as the cognition hypothesis (Robinson, 2001a, 2001b, 2005, 2007) and the limited attentional capacity model (Skehan, 1998, 2009; Skehan & Foster, 2001). As a result, we have observed an increase in the number of L2 writing studies associated with task-based language teaching (TBLT) that focuses on the effects of task-internal cognitive demands on written language production (e.g., Ellis & Yuan, 2004; Frear & Bitchener, 2015; Johnson et al., 2012; Kormos, 2011; Kuiken & Vedder, 2007, 2008; Ong, 2014; Ong & Zhang, 2010; Révész, Kourtali, & Mazgutova, in press).

While many L2 writing studies have found a significant impact of task manipulations on students' language use, their specific findings on the link between task features and linguistic features have not converged. For example, Révész et al. (in press), in which the provision of idea support was manipulated, found significant task effects on syntactic complexity and lexical diversity, but a similar task manipulation did not result in significant changes in similar linguistic units in Kormos (2011). Additionally, Ellis and Yuan (2004) suggested a positive effect of pre-task planning on L2 writers' writing fluency and linguistic complexity in narratives, but Johnson et al. (2012) failed to find such a significant impact in his study using argumentative essays. In discussing their different findings from previous studies (Ellis & Yuan, 2004; Ong & Zhang, 2010), Johnson et al. (2012) suggested the use of different genres as one of the potential reasons. Interestingly, Révész et al. and Kormos (2011) also explored different genres of argumentative and picture-based narrative writing respectively, which might be a factor leading to discrepant findings between the two studies.

Ironically, the most consistent findings of task effects have been suggested by several task-based L2 studies that used genre as a task complexity variable and examined the effect of genre on linguistic features such as syntactic complexity and accuracy (e.g., Ruiz-Funes, 2014, 2015; Way et al., 2000; Yang, 2014). Specifically, with the assumption that the cognitive demands induced by narrative tasks are lower than those by non-narrative tasks, these cross-genre studies framed narratives as a simple task and non-narratives (e.g., exposition or argumentation) as a complex task. The findings of these studies have consistently shown increased levels of syntactic complexity in non-narrative writing when compared to those in narratives, which is well aligned with task-based hypotheses such as Robinson's cognition

hypothesis (i.e., more complex tasks may elicit more complex language). Taking into account all of these findings in the L2 writing literature, I drew three conclusions:

1. Task-based writing studies have produced varying results of task manipulation effects.
2. Task manipulation effects may interact with genre.
3. Different genres elicit different linguistic features.

However, these conclusions do not confirm our understanding of how task complexity manipulations work in written discourse, but rather they emphasize what is lacking in the field and suggest the necessity to examine the validity of some presumptions supported by many researchers.

First, we need to test the level of cognitive demands imposed by different genres. While it is reasonable to assume that making a logical argument necessitates more in-depth thinking processes than describing a story does (Beauvais et al., 2011), as the majority of cognitive models of writing suggest, a writer's task schemas and other types of knowledge (e.g., topic, genre, and audience) indisputably influence the cognitive demands of a writing task placed on the writer. Of several cognitive models of writing, this study is grounded in the model of writing proposed by Hayes and his colleagues, which has been widely accepted in writing research fields (Hayes, 1996, 2012; Hayes & Flower, 1980). Since the original model in Hayes and Flower (1980), John Hayes has constantly modified his model by including more affective and motivational factors, but the writer's task schemas and knowledge remain as important resources that moderate writing processes and cognitive constraints. Specifically, the revised model (Hayes, 1996) includes the two major factors of the task environment and the individual. The latter is composed of four components that interact with each other: motivation/affect, cognitive processes, working memory, and long-term memory. Of these components, most relevant to the

focus of this study is the writer's long-term memory component that includes task schemas, topic knowledge, audience knowledge, linguistic knowledge, and genre knowledge. During the act of writing, this knowledge-related component together with other dimensions of individual differences such as working memory and motivational attributes interacts with the task environment (e.g., task materials, writing medium, collaborators, and audience). In other words, writers' task-relevant knowledge can be used to reduce a level of cognitive demands imposed by a certain task; writers' performance is dependent on their familiarity and understanding of a given task.

One important question inferred from this model is: do adult English as a second language (ESL) students really have greater genre knowledge and task schemas for narrative tasks than for argumentative tasks, as accepted by many researchers? In this study, I set out to answer this question by exploring both perception and language production data. While the majority of TBLT studies have initially focused on oral tasks, writing researchers recently began to examine cognitive task complexity to see how it interplays within written discourse (see Plonsky & Kim, 2016 for a review), either adjusting task features within a specific written genre (e.g., Ong & Zhang, 2010; idea support condition manipulated in argumentative writing) or operationalizing genre as a cognitive complexity dimension (e.g., Ruiz-Funes, 2015; expository genre operationalized as more complex than narrative). Here, I argue that genre as a task variable needs to be manipulated and analyzed with caution because there are two research lines that address a similar issue with different starting points and purposes.

Specifically, one research tradition originated from composition studies with L1 children addresses how learners at different grades (or proficiency levels) show distinct writing skills across genres, attempting to identify an appropriate genre for a particular age group (e.g., Beers

& Nagy, 2011; Berman & Katzenberger, 2004; Berman & Nir-Sagiv, 2004; Ravid, 2005). In this tradition, researchers have explained potential genre effects on linguistic features by linking linguistic forms to discourse functions (e.g., extensive use of past tense to express temporality in narratives and increased noun-phrase complexity to express generality in non-narrative genre; Berman & Katzenberger, 2004). The other line of research sees written genres as tasks having different cognitive demands (e.g., Ruiz-Funes, 2014, 2015; Yang, 2014). Drawing on TBLT hypotheses, researchers in this line have interpreted the increased linguistic complexity of learner language in a particular genre as evidence of the genre's higher task complexity; with the consistent findings of an increase in linguistic complexity in non-narrative genre compared to narrative genre, they concluded that genre is a valid task complexity variable affecting L2 learners' cognitive processes and language production. Therefore, despite their similar methods and results, the two research lines have been established with different assumptions about written genre, generating diverging interpretations.

To problematize the presumption of different genres' varying cognitive demands (more specifically, equating argumentative writing with a high-cognitive demand task and narrative writing with a low-cognitive task), I draw on the long-term memory component of the writing model (Hayes, 1996; Hayes & Flower, 1980) that includes genre knowledge. Major assumptions about genre-specific cognitive demands are based on the findings of L1 writing research whose participants were mostly children or adolescents (e.g., Berman, 2008; Ravid, 2005). For example, researchers presuppose that students may have greater experience with narrative tasks than with argumentative tasks because children in the U.S. educational system actually work primarily on narrative tasks as a first step of developing their full range of writing skills. Regarding genre-specific writing skills, the standards of English language arts that have been

adopted by forty-two states and the District of Columbia suggest that K-5 students need to develop skills for narrative, opinion, and simple explanatory writing and those in grades 6-12 develop skills for argumentative writing (CCSS, 2017). The alignment of written genres with specific grades is a clear reflection of children's developmental trajectories of cognitive skills. Specifically, it is widely known that children undergo notable growth in cognitive abilities with age, and that their cognitive skills for rational judgment and abstract thinking start to develop in the stage of ages 7 to 11 (Ginsburg & Oppenheimer, 1988). Thus, it can be very challenging for children or young adolescents to compose an argumentative essay, and it is reasonable to assume that children would feel more comfortable with narratives. However, the same scenario cannot be applied to adult L2 learners who have already reached a high level of cognitive maturity. Furthermore, most adult L2 learners have finished primary and secondary schools in educational contexts distinct from those in the United States, leading me to assume that genre-specific difficulties for adult L2 learners may depend on their educational experience with various genres and modes of discourse.

The key focus is the quantity and quality of writing instruction that typical adult L2 learners are likely to have experienced before coming to an ESL context, as well as their motivation for learning English writing. This is particularly so when I consider the components of Hayes' (1996) writing model that include motivation, task schemas, and genre knowledge as an important part of the individual element. First, it should be noted that adult ESL students who learned English mostly in primary and secondary schools in their own countries (i.e., English as a foreign language contexts) are likely to have acquired limited English writing skills due to the English educational systems greatly influenced by high-stakes exams that focus on receptive language skills (Butler & Iino, 2005; Byun et al., 2011; Watanabe, 1996). Particularly, L2

learners in East Asian countries would likely have received English instruction that focuses on the development of grammar, vocabulary, and reading comprehension skills because of the inclusion of such skills in the English section of high-stakes college entrance exams (e.g., Cheng, 2008; Jeon, 2009; Kikuchi, 2006; Sakamoto, 2012).<sup>1</sup> For example, Shim and Baik (2004) noted that English teachers in South Korea have difficulties in teaching productive English skills due to students' expectations of having examination-oriented instruction. English teachers in Japan and Hong Kong have also expressed similar concerns (Butler & Iino, 2005; Chow & Mok-Cheung, 2004).

Given this information, what we can expect from many ESL students is that their major English writing practice would be for the preparation of standardized English tests (e.g., TOEFL or IELTS), with the scores being required to obtain admission to schools in English-speaking countries. Further, considering the fact that argumentative writing has long been a typical genre for standardized writing assessments (Qin & Karabacak, 2010), we can infer that adult ESL students would have greater genre knowledge for argumentative essays than narratives. It may still hold true for adult ESL students who have received college education in an English-speaking country for years because argumentative writing is a typical and necessary text type for the college academic curriculum (Christie, 1997; Johns, 1995; Mei, 2006). All of these points likely challenge a current understanding that implementing an argumentative task will naturally impose increased cognitive complexity on adult ESL learners and suggest a new prediction that adult L2

---

<sup>1</sup> According to the statistics of Institute of International Education (2016), students from China constitutes 31.5% of the entire international student population studying in the United States, and those from China, South Korea, Taiwan, and Japan adds up to 40.3%. The college entrance exams administered by public institutes in these countries (e.g., National Higher Education Entrance Examination in China; College Scholastic Ability Test in South Korea; and National Center Test for University Admissions in Japan) are large-scale, multiple-choice tests that do not involve actual writing.



learners would be more familiar and thus comfortable with argumentative writing, which will be tested in the present study.

With regard to the literature of TBLT research, I noted earlier that task complexity writing studies had produced conflicting findings in terms of their support for task complexity hypotheses (extensively reviewed in the Literature Review section). One possibility is that some task manipulations are not applicable to written discourse due to several fundamental differences between the two modalities (written and oral language production) (Biber, 1988, 2006a; Chafe, 1982). Researchers have expressed concerns about the validity of the direct application of cognitive complexity hypotheses to writing (Frear & Bitchener, 2015; Jackson & Suethanapornkul, 2013; Johnson et al., 2012; Yoon & Polio, 2017). That is, while the underlying assumption of cognitive task complexity is the allocation of limited attentional resources, writing would be less constrained by such cognitive limitations due to the features of writing as a recursive process that involves a series of planning, monitoring, and revising (Hayes, 1996; Hayes & Flower, 1980; Kellogg, 1996). In this regard, Yuan and Ellis (2003) argued that writers are less pressured than speakers in terms of their allocation of attention between idea conceptualization and linguistic formulation, and also that writers have more attentional resources available for planning and monitoring than speakers. Based on their finding of the lack of pre-planning effects on written language production, Johnson et al. (2012) discussed the following:

[W]riting is fundamentally different from speaking. For this reason, written L2 production may not be described accurately by the Cognition Hypothesis (Robinson, 2001, 2005, 2011a, 2011b) nor by the Limited Attentional Capacity Model (Skehan, 1998; Skehan & Foster, 2001) because such models predict the impact of pre-task

planning on L2 oral production. Because speaking is a linear process, planning time prior to L2 speaking tasks is effective in relieving attentional demands of language production ... In contrast, writing is a recursive process, thus planning time prior to L2 writing tasks does not obviate online planning as well as monitoring. (p. 271)

Another concern is the validity of the ways that researchers operationalize cognitive complexity for writing tasks. The manipulations of task complexity in L2 writing studies include the provision of planning time (e.g., Ellis & Yuan, 2004; Ojima, 2006; Ong, 2013, 2014; Ong & Zhang, 2010), number of elements (e.g., Kuiken, Mos, & Vedder, 2005; Kuiken & Vedder, 2007, 2008, 2011), here-and-now (Ishikawa, 2007), and conceptualization support through idea provision (e.g., Kormos, 2011, 2014; Kormos & Trebits, 2012; Ong, 2013, 2014; Ong & Zhang, 2010; Révész et al., in press), most of which have been directly applied from TBLT speaking studies. Specific to written discourse is the pattern of task manipulations in relation to written genre. On the one hand, researchers have manipulated task dimensions within a specific genre (i.e., within-genre manipulation studies). For example, Kuiken and Vedder varied the number of elements to be considered to decide a travel destination (3 and 6 elements) in letter writing, and Kormos adjusted the level of conceptual demands in a picture narrative task by changing the condition of supporting content. On the other hand, a few recent studies have operationalized genre as one of the resource-directing dimensions of cognitive complexity (i.e., cross-genre manipulation studies), based on the assumption that argumentative essays would involve higher cognitive complexity than narrative essays (e.g., Ruiz-Funes, 2014, 2015; Yang, 2014). There were also some L2 writing studies that investigated multiple genres composed by learners and interpreted findings with a similar assumption of genre-specific cognitive demands, although

these studies were framed as task complexity studies (e.g., Jeong, 2016; Qin & Uccelli, 2016; Way et al., 2000).

However, as discussed in Polio and Yoon (2016), genre research and task-based research have suggested varying interpretations from similar findings (e.g., higher syntactic complexity in non-narrative writing) due to different starting points of each research line (i.e., communicative functions in genre research and cognitive demands in task research). Furthermore, some previous task-based studies have suggested different patterns of task effects, potentially due to the use of different genres (e.g., Kormos, 2011; Ong & Zhang, 2010; Révész et al, in press), suggesting the need to explore the interaction between genre and task complexity effects on L2 learners' language production and perceptions. Of several existing task variables for within-genre manipulations, it seems particularly important to explore the condition of idea support in terms of its varying roles in different genres because this variable was found to influence writers' perceptions validly in one genre (i.e., idea support judged as a valid task variable in argumentative writing by Révész et al, in press), while others have not been tested in terms of their validity.

To explore the validity of genre and task manipulations, in this study, I examine ESL learners' production and perceptions of four writing tasks, together with ESL teachers' perceptions of the same tasks. The tasks targeted in this study involve argumentative and narrative genres within which a level of task complexity is manipulated in terms of the provision of supporting ideas. Students' perceptions of the tasks are collected immediately after their writing performance via a self-rating questionnaire. Going beyond the common practice of examining traditional linguistic complexity features to validate task complexity hypotheses (see Robinson, 2011 for a review), I analyze textual features at multiple levels (i.e., syntactic, lexical,

discourse, and metadiscourse levels), attempting to explain the motivation for linguistic changes on the basis of their communicative functions. In the following chapters, I review L1 and L2 genre studies, as well as task-based writing studies in order to suggest specific gaps in the literature and to introduce how this study addresses them appropriately.

## CHAPTER 2.

### LITERATURE REVIEW

#### Definitions of Genre and Other Related Terms

There has been a large body of research into the effect of genre on learners' language use (e.g., Beers & Nagy, 2009, 2011; Lu, 2011; Qin & Uccelli, 2016; Yoon & Polio, 2017, among others). Researchers have also shown much variation in writing processes and essay scores arising from genre differences (e.g., Beauvais et al., 2011; Bouwer et al., 2015; Hamp-Lyons & Mathias, 1994; Jeong, 2016; Way et al., 2000). Findings from such extensive genre research have suggested the need to control for genre in developmental research and to employ different genres to obtain a more comprehensive understanding of learners' writing proficiency in assessment contexts. However, there is still some confusion about the notion of the term *genre* because researchers have used genre and other related terms such as register, text type, and mode of discourse in different ways. For example, early research used the term *mode of discourse* in discussing traditional types of rhetorical categories such as narrative, description, and argumentation (e.g., Crowhurst, 1979, 1980; Engelhard, Gordon, & Gabrielson, 1992; Kegley, 1986; Steen, 1999), while recent studies referred to such categories as genres (e.g., Jeong 2016; Lu, 2011; Qin & Uccelli, 2016). Some authors used these terms interchangeably with no explicit distinction among terms (Stubbs, 1996). Accordingly, to avoid potential confusion, before reviewing L1 and L2 genre studies, I clarify my use of genre-related terms in this study.

Several studies have attempted to address the elusive nature of these text-classifying terms by elucidating their different nuances (e.g., Biber, 1988; Lee, 2001; Nunan, 2008; Paltridge, 1996). An early attempt to differentiate between genre and text type is Biber (1988), in which genre is considered a classification based on external criteria such as purpose and

audience, and text type a category based on text-internal criteria such as linguistic features. That is, although some texts have very similar linguistic characteristics, they could be seen as different genres when they have different purposes. In line with Biber's distinction, Paltridge (1996) also suggested the criteria of external and internal dimensions to explicate the meanings of genre and text type. Nunan (2008) noted that a collection of texts can be grouped into the same genre when they have a common communicative function, while acknowledging great difficulty in building confirmatory taxonomies of genres.

According to Lee (2001), there is some additional difficulty distinguishing between *genre*, *register*, and *style* clearly due to some overlap in their meaning and the interchangeable use of these terms in previous research. In Biber and Conrad's (2009) book-length study, the authors endeavored to define each of these terms for clarification purposes. They noted that register and style are categories based on the frequently occurring linguistic features. That is, some linguistic features pervasive in one register would be infrequent or rare in another register (the same premise applicable to styles). The difference between register and style is that the former primarily involves varying linguistic features arising from different situations and contexts, while the latter involves linguistic variation related to an individual writer's linguistic choices, which is a widely-accepted classification now. Register and genre are in fact the two terms that requires further elucidation. Drawing on the concepts of systemic-functional linguistics, Biber and Conrad (2009) distinguished between genre and register:

Register variation focuses on the pervasive patterns of linguistic variation across such situations, in association with the functions served by linguistic features; genre variation focuses on the conventional ways in which complete texts of different types are structures. Taken together, register/genre variation is a fundamental aspect of human

language. All cultures and languages have an array of registers/genres, and all humans control a range of registers/genres. (p. 23)

Other genre researchers have also noted that genre is likely to be associated with its relevant cultural context, while register concerns the immediate context of situation (Martin, 1993, 2001; Swales, 1990). More importantly, it has been noted that the analysis of register variation begins with inductive text analysis that contributes to identifying different registers, while genre variation is analyzed in terms of the occurrence of a particular rhetorical organization that reflects the predicted structure of a genre (see Lee, 2001). Specifically, the two unique features of genre are the use of external criteria such as communicative purpose and the use of pre-identified categories. Biber and Conrad viewed register as the most important category for text analysis because it fully recognizes linguistic features as units fulfilling situational functions and *all* types of texts can be analyzed in terms of their linguistic features that contribute to register variation. Their support for inductive text categorization based on register variation is well aligned with their dedication to the multi-dimensional approach, which identifies a set of co-occurring text features through factor analysis and assign composite scores on each text to categorize them into different registers or text types. The current study, however, does not involve any inductive grouping of texts based on a set of linguistic features; instead, it focuses on potential changes in linguistic features across prearranged genres to identify the linguistic representations of genre-specific communicative functions, and this is my rationale for using the term genre as the feature under investigation.

The last, but important, typological feature of genre is its varying levels of generality (Martin, 1993; Steen, 1999). This means that a particular genre can consist of multiple sub-genres, each of which can function as a superordinate genre that include further sub-genres. In

this regard, drawing on prototype theory in cognitive science, Steen (1999) suggested that genre could be conceptualized as having multiple hierarchies that include super-genre (superordinate level), genre (basic-level), and sub-genre (subordinate level), pointing to the importance of understanding the flexible nature of the level of generality in recognizing genres. The application of this taxonomy to the focus of this study is presented in Table 1. Therefore, the two genres used in this study are argumentative and narrative writing (more specifically, position-setting argumentative and personal narrative writing). In one sense, timed writing can be considered too specific to be a super-genre, but given the prevalence of timed writing in a wide range of academic setting (e.g., standardized tests, placement tests, and in-class tasks), I believe that timed writing merits a superordinate category that can be further divided into its genres and subgenres.

Table 1.

*Taxonomy of Genre in This Study*

Classification	Examples relevant to the present study
Superordinate (Super-genre)	Timed writing
Basic-level (Genre)	Argumentative writing, narrative writing
Subordinate (Sub-genre)	Position-setting argumentative writing (agreement/disagreement), solution-suggesting argumentative writing (deciding on the best solution), personal narrative writing, imaginative narrative writing, picture narrative writing ...

Writers are expected to fulfill different functions and communicative purposes in different genres. Based on their primary rhetorical functions, written genres can be divided into narratives and non-narrative types (Bruner, 1986); narratives entail an event description with a focus on people's actions in a specific time frame, while non-narrative essays involve the



argumentation or explanation of general ideas (Berman & Slobin, 1994). In this study, I target two genres that elicit strikingly different communicative functions: timed argumentative and timed narrative tasks (i.e., making arguments to convince readers in argumentative and telling an interesting story to entertain readers in narrative). I use *task type* as a looser term when referring to different writing tasks manipulated in terms of either genre or task complexity (idea support). That is, different genres are always different task types, and the same genre can include different task types when manipulated in terms of the condition of idea support.

### **Cross-genre L1 Studies**

Over the past thirty years, there have been many L1 writing studies on genre differences. This sustained attention to genre in L1 writing research reflects the implementation of varying genres for assessing students in different grades, which is aligned with the state standards (CCSS, 2017). As described above, children in different grade levels are expected to focus on developing skills for different genres (i.e., K-5 students for narrative, opinion, and explanatory genres; 6-12 grade students additionally for argumentative genre). The majority of L1 genre studies have consistently demonstrated that children have greater difficulty in composing non-narrative essays than narrative essays (e.g., Berman, 2008; Berman & Sobin, 1994; Hickman, 2003; Peterson & McCabe, 1983; Ravid, 2005), which was often interpreted as the consequence of teachers' tendency to use narrative tasks as major writing assignments for young learners (Engelhard et al., 1992). Specifically, Ravid (2005) noted that children are capable of writing personal narratives that entail people, events, and places, while young adolescents still have difficulty with expository writing that requires abstract content knowledge, indicating an expectation that children would not be able to accomplish argumentative tasks.

Previous studies provided further support for a genre-cognition connection by showing higher essay scores in narratives than in non-narrative writing (e.g., Bouwer, Béguin, Sanders, & van den Bergh, 2015; Crowhurst, 1980; Engelhard et al., 1992; Kegley, 1986; Sachse, 1984). Kegley (1986), for example, reported varying proportions of adequate and inadequate writing performance across four genres (description, narration, exposition, and persuasion). Kegley collected data from seventh-grade students and categorized their competency as either adequate performance (scores 2 or lower) or inadequate performance (scores 3 or higher) using a holistic rubric (scores from 0 to 4). Her result showed the highest proportion of adequate performance in narrative genre and the lowest proportion in persuasion (i.e., adequate performance proportion in narration: 56%; description: 43%; exposition: 41%; and persuasion: 31%), suggesting that more than one fifth of students may be given different evaluations and categorized in different proficiency groups according to genre. Similarly, Engelhard et al. (1992) explored eighth-grade students' performance on the three genres of narrative, descriptive, and expository tasks, and their results also demonstrated the highest scores on personal narratives and the lowest scores on expository tasks. Some additional information from this study is that, unlike Kegley (1986), Engelhard et al. employed an analytic rubric that includes content/organization, style, sentence formation, usage, and mechanics and suggested that the effect of genre was stronger on discourse-level development (i.e., content/organization and style) than on sentence-level sophistication (sentence formation, usage, and mechanics).

Recently, Bouwer et al. (2015) analyzed 67 sixth-grade children's writing and statistically verified genre as an important factor explaining 11% of the variance in writing scores (i.e., higher scores in narrative tasks than argumentative tasks). The authors suggested genre knowledge as one of the possible explanations for a clear genre effect on essay scores and,

specifically, assumed that children might have build more stable schemata for narrative writing than those needed for argumentation. Based on the results of generalizability theory, they further suggested that at least two raters should evaluate three texts in each of four genres to draw generalizable writing proficiency. Given the practical impossibility of such a testing setting, their conclusion seemed intended to warn us not to judge one's writing proficiency based on one writing performance.

Unlike the majority of L1 genre research findings that indicated significant genre effects on essay scores, Beers and Nagy (2009), who collected data from 41 seventh and eighth grade students, showed a different pattern. They explored the two genres of persuasive and narrative writing for their holistic essay scores and syntactic complexity (clauses per T-unit, words per clause, and words per T-unit). While having different levels of syntactic complexity (i.e., higher syntactic complexity in persuasive essays than narratives), the students obtained similar essay scores on the two genres. More strikingly, Quellmalz, Capell, and Chou (1982) analyzed expository and narrative essays composed by high school students (those in eleventh and twelfth grades) and found significantly higher ratings for expository writing than narratives. Noting the older ages of their participants compared to other studies (i.e., high school students in contrast with elementary or middle school students in other L1 genre studies), the authors interpreted this unexpected finding as either the outcome of greater focus of the high school curriculum on expository genre (strongly established schemata for expository writing) or raters' varying leniency across genres.

L1 genre research that focused on linguistic form variations has consistently demonstrated that syntactic complexity tends to increase in non-narrative writing compared to narrative writing (e.g., Beers & Nagy, 2009; Crowhurst & Piche, 1979; Ravid, 2005). For

example, an early L1 study by Crowhurst and Piche (1979) examined how syntactic complexity measures (production unit length and subordination) differ across three genres (narrative, descriptive, and argumentative writing). The authors found the highest syntactic complexity in argumentative essays and the lowest in narratives, suggesting initial evidence of the variability of syntactic complexity across genres. With more findings that support this pattern of language variation, it has been concluded that child and adolescent writers modify their language across genres to fulfill different rhetorical functions (e.g., Beers & Nagy, 2009; Berman & Katzenberger, 2004; Ravid, 2005).

While much attention has been given to the dimension of syntactic complexity, there has also been a body of research that focused on the effect of genre on lexical features (e.g., Gardner, 2004; Grobe, 1981; Olinghouse & Leaird, 2009; Olinghouse & Wilson, 2013). For example, in their repeated-measures design study, Olinghouse and Wilson investigated how various dimensions of lexical features would vary by genre (narrative, persuasive, and informative tasks) and how such lexical features predict the writing quality of each genre. With regard to the effect of genre on lexical features, Olinghouse and Wilson found that lexical diversity was the highest in the narrative texts, while content vocabulary and elaboration were the highest in the informative texts. There were no statistical differences among the three genres in the use of academic words. In terms of the prediction of each genre's text quality, the authors identified lexical diversity as the strongest predictor of narrative writing quality and content vocabulary as the strongest predictor of the quality of persuasive and informative writing. This result might indicate a similar expectation of extensive use of topic-relevant content words for high quality persuasive and information writing, which was not the case for narrative genre. The findings of this study offered empirical evidence for varying lexical features elicited by different genres,

suggesting that the effects of genre might be on a wide range of linguistic features beyond the traditional scope of syntactic complexity.

To summarize, the findings of L1 genre studies provide sufficient evidence for the variation in writing performance across genres. Although there were a few exceptions (e.g., Beers & Nagy, 2009; Quellmalz et al., 1982), most studies have demonstrated higher scores on personal narrative essays than on argumentative essays. In terms of language variation, it has been suggested consistently that non-narrative essays tend to contain more complex language than narratives. While informative, these findings need to be complemented with more comprehensive findings because there is little evidence suggesting that higher ratings for narrative tasks actually reflect learners' better performance and lower challenges. Because the majority of previous research followed the tradition of making inferences about cognitive challenges from essay scores or linguistic features, future research needs to include independent measures of learner perceptions (Révész, 2014; Sasayama, 2016) to better understand the cognitive demands of distinct genres. Furthermore, a few genre studies that adopted an analytic scoring rubric commonly found greater genre effects on discourse-level subscales (e.g., organization) than sentence-level ones (e.g., mechanics), indicating the greater sensitivity of discourse-level scores to genre variation (e.g., Kegley, 1986; Quellmalz et al., 1982). In this study, I attempt to shed further light on these areas by adopting an analytic scoring rubric and a task perception questionnaire.

### **Cross-genre L2 research**

So far, I have reviewed the literature of L1 genre studies. The findings from L1 research have generally shown a significant impact of genre on essay quality (e.g., Bouwer et al., 2015; Engelhard et al., 1992; Kegley, 1986; Quellmalz et al., 1982) and language use (e.g. Crowhurst

& Piche, 1979; Olinghouse & Wilson, 2013; Ravid, 2005), as well as the mediating effect of genre on the relationship between essay quality and linguistic features (e.g., Beers & Nagy, 2009; Crowhurst, 1980; Olinghouse & Wilson, 2013). While it has been more than 30 years since some early attempts to explore genre effects in L1 research, genre has begun to attract researchers' attention fairly recently in L2 writing research, and the major focus of L2 genre studies has also been on the effect of genre on learners' language use (e.g., Lu, 2011; Qin & Uccelli, 2016; Yoon & Polio, 2017).

Findings from L2 genre research generally indicated that learner language in argumentative writing tend to be more complex than that in narrative writing, which is well aligned with findings from L1 studies. Specifically, using his own automated processing tool for syntactic complexity, Lu (2011) examined the syntactic complexity of Chinese learners of English in narrative and argumentative texts. According to Lu's results, L2 learners showed higher values of production unit length (e.g., mean sentence length and mean clause length) and phrase-level syntactic complexity (e.g., complex nominals per clause and coordinate phrases per clause) in argumentative essays than in narrative essays. The findings of L2 research that are similar to those of L1 research may indicate that both L1 and L2 writers have a certain level of genre awareness, leading to language variations arising from genre-specific communicative functions. Based on such notable linguistic changes across genres, L2 studies have also confirmed the role of genre as an important task variable that should be taken into account when research explores language development (e.g., Yoon & Polio, 2017).

Unlike the consistent findings regarding the association between genre and language, previous L2 research into the effect of genre on text quality have suggested mixed findings (e.g., Hamp-Lyons & Mathias, 1994; Jeong, 2017; Qin & Uccelli, 2016; Way et al., 2000). For

example, Way et al. (2000) explored three different tasks (descriptive, narrative, and expository writing) composed by low-level L2 French learners and found the lowest essay scores on expository and the highest scores on descriptive writing. Focusing on task-internal challenges, the authors concluded that the expository task might have been most challenging and the descriptive task least challenging for low-level L2 learners. The finding of Hamp-Lyons and Mathias (1994) showed the opposite direction of genre effects on writing scores (i.e., higher holistic scores on argument/public writing tasks than expository/private tasks). Unlike Way et al., Hamp-Lyons and Mathias focused on task-external features such as raters' perceptions of the prompts and interpreted their results as the outcome of raters' adjustment of rating severity based on perceived task difficulty (e.g., assigning higher scores for argument tasks that are perceived as more difficult by raters).

In addition, there have been recent studies that showed a different picture of genre effects by taking into account additional variables such as writers' L2 proficiency (Jeong, 2017; Qin & Uccelli, 2016). For example, Jeong examined the narrative and expository essays written by 180 Korean learners of English at three different proficiency levels (60 students from each of the novice, intermediate, and advanced levels). Based on the results of a multi-faceted Rasch analysis, Jeong showed there is no significant difference in EFL writing scores between the two genres; instead, the author suggested a significant interaction between genre and L2 proficiency. Specifically, it was revealed that beginning writers tended to obtain higher scores on narrative, while advanced writers tended to have higher scores on expository writing, suggesting the complex nature of the role of genre in writing performance and the need to take into account L2 proficiency in exploring genre effects.

Qin and Uccelli (2016) investigated how secondary-school Chinese EFL students

perform differently on argumentative and narrative essays. Analyzing 200 texts produced by 100 EFL students, they showed that the students' writing performance, which was measured using a holistic scoring rubric, did not significantly differ by genre. Despite no clear effect of genre on writing scores, the authors found that the quality of each genre was best predicted by a different set of textual features; narrative writing quality was predicted by stance marker frequency, while argumentative writing quality was predicted by lexico-syntactic complexity and organization marker diversity. To summarize, there have been some interesting, but conflicting, findings from L2 genre studies in terms of the influence of genre on essay scores. While different studies interpreted their findings with different foci (e.g., rater severity, task difficulty, and L2 proficiency), one methodological commonality of these L2 studies is its reliance on holistic text scores (possibly due to practical reasons), but as shown by some early L1 studies (e.g., Kegley, 1986; Quellmalz et al., 1982), different categories of a scoring rubric have varying levels of sensitivity to genre variation, indicating the need to employ an analytic rubric in L2 genre research for the identification of more specific patterns and, more generally, for the advancement of the field.

Given the consistent findings of language changes across genres and somewhat contrasting findings of essay score changes, there is one important question that remains to be resolved in the area of L2 genre research: how can we determine the reason for these well-attested language differences across genres? To provide empirical evidence related to this question, Yoon and Polio (2017) analyzed linguistic complexity, accuracy, and fluency (CAF) features in the narrative and argumentative essays composed by 37 ESL students and 46 native English-speaking college (NS) students. The starting point of this study was the premise that ESL students would have greater cognitive pressure for timed writing than NS students due to



their limited command of the language. Then, based on the cognition hypothesis (Robinson, 2001b, 2005, 2007) suggesting that L2 learners' greater use of attentional resources for language forms in a more cognitively demanding task would lead to their use of more complex language, Yoon and Polio predicted that ESL students' language would be influenced more strongly by genre than that of NS students if different genres in fact pose greatly different cognitive demands on writers. On the other hand, if both ESL and NS writers show similar genre effects, it may provide evidence that different linguistic features elicited from different genres are indicative of their fulfillment of genre-specific functional needs.

The results of Yoon and Polio showed similar patterns of language differences from both groups, and the authors concluded that language variation across genres may be better explained as the outcome of different communicative functions expected in different genres than genre-specific cognitive demands. For example, narrative writing is likely to contain more personal pronouns, while argumentative include more nominalizations and nominal post-modifiers, leading to higher linguistic complexity in argumentative writing than in narratives (see Biber & Conrad, 2009 for a detailed description of discourse features). Yoon and Polio paved the way for questioning the validity of linking genre and cognitive demands, but there is still a need to use an independent measure of a writer's task perceptions (Révész, 2014) in order to clearly disentangle genre effects on linguistic features from those on learners' perceptions in written discourse. In the following sections, I review task-based writing studies that examined the effect of various cognitive complexity variables on L2 learners' language use and justify the adoption of supporting idea provision as a target task complexity dimension in this study.

## **Task-based Writing Studies**

Over the past decade, there has been an increase in the number of L2 researchers who have shown an interest in the effects of task complexity on learners' language. Task complexity, defined as the “attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner” (Robinson, 2001a, p. 29), has been argued to influence the amount of attentional and cognitive resources available for language constructions during task performance (Robinson, 2003; Skehan, 1998). To gather evidence of potential task complexity effects on language production, L2 researchers have explored how the manipulation of various task features (e.g., planning time availability: Yuan & Ellis, 2003; number of elements: Kuiken & Vedder, 2011; here-and-now: Gilabert, 2007), identified by Robinson's Triadic Componential Framework (Robinson, 2001b, 2005, 2007), can lead to changes in traditional CAF measures (see Housen, Kuiken, & Vedder, 2012; Norris & Ortega, 2009 for a review of CAF). While most research initially focused on the effects of cognitive task complexity on oral language production, authors of L2 writing studies began to examine cognitive task complexity to see how it interplays within written discourse (e.g., Ellis & Yuan, 2004; Frear & Bitchener, 2015; Johnson et al., 2012; Kormos, 2011; Kuiken & Vedder, 2007, 2008; Ong, 2014; Ong & Zhang, 2010; Révész et al., in press).

In cognitively-oriented TBLT studies, manipulations of task complexity are expected to create differing cognitive demands in the conceptualization stage that may lead to changes in the amount of attentional resources allocated to language constructions and, accordingly, in the complexity level of linguistic forms (see Robinson, 2001b, 2005). In terms of the causal relationships between task complexity (with regard to cognitive demands) and language production, two competing hypotheses—cognition hypothesis (Robinson, 2001a, 2001b, 2005,

2007) and trade-off hypothesis (also referred to as limited attentional capacity model; Skehan, 1998, 2009; Skehan & Foster, 2001)—suggest different explanations of how varying cognitive demands of language tasks lead to a difference in task performance. Specifically, Robinson's cognition hypothesis presumes that there are multiple dimensions of attentional resources that language learners can access simultaneously. Dividing task complexity into resource-directing and resource-dispersing dimensions, Robinson (2001a, 2005) argued that increasing task complexity along resource-directing dimensions (e.g., adding more elements or increasing reasoning demands) leads learners' attentional resources to complex language constructions, facilitating language development. On the other hand, increasing task complexity along resource-dispersing dimensions (e.g., no planning time) imposes greater demands on attentional and working memory resources, leading to learners' dispersed attention to language formulation.

In contrast, Skehan's trade-off hypothesis (Skehan, 1998, 2009; Skehan & Foster, 2001) presupposes that learners have limited attentional resources and working memory; during task performance, learners are not capable of attending to content and language at the same time. In other words, paying attention to one area leads to the reduced attentional resources available for other areas. Thus, more complex tasks requiring higher conceptual demands direct learners to focus less on linguistic aspects. The limited amount of attentional resources available for language further leads to a competing relation between linguistic complexity and accuracy. Of the two dimensions of cognitive complexity (i.e., resource-directing and resource-dispersing dimensions), the discrepancy between the two hypotheses exists mainly in terms of the effect of resource-directing cognitive demands on language production. Therefore, there has been a larger body of task-based writing research into cognitive complexity effects along the resource-directing dimension in an attempt to test the predictions of each of the two hypotheses (e.g.,

Frear & Bitchener, 2015; Kormos, 2011, 2014; Kuiken & Vedder, 2007, 2008; Ruiz-Funes, 2015; Tavakoli, 2014; Yang, 2014; Yang, Lu, & Weigle, 2015), as compared to those along the resource-dispersing dimension (e.g., Ellis & Yuan, 2004; Ishikawa, 2007; Johnson et al., 2012).

We can view task manipulations in the written modality with regard to genre, classifying them into within-genre or cross-genre manipulations (Polio & Yoon, 2016). Of several within-genre variables (e.g., here-and-now, number of elements, and reasoning demands), this study focuses on the level of conceptual demands operationalized as the provision of supporting ideas. Previous research has explored how varied conceptual demands lead to a difference in written language production (e.g., Kormos, 2011; Kormos & Trebits, 2012; Ong & Zhang, 2010; Révész et al., in press; Tavakoli, 2014), with the prediction that a task with greater complexity at the level of idea conceptualization would lead a writer to formulate more complex language (see Robinson, 2001b, 2005). First, picture-based writing tasks have been adopted for conceptualization-level manipulations (e.g., Kormos, 2011; Kormos & Trebits, 2012; Tavakoli, 2014); the picture narration task that requires participants to develop a story plot based on the pictures given in random order is considered more complex than the cartoon description task that provides a clear storyline. For example, Kormos (2011) explored how the difference in conceptualization demands affected linguistic- and discourse-level features in NS and NNS writing, and found that the writers showed increased lexical sophistication and connective use in a more complex task (i.e., random-order picture narration) but no difference in lexical diversity, accuracy, syntactic complexity, and cohesion. While adopting more fine-grained accuracy measures (ratio of error-free clauses, ratio of error-free relative clauses, error-free verbs, and error-free past-tense verbs), Kormos and Trebits (2012) still revealed no task complexity effect on linguistic accuracy.

The variable of conceptualization demands has also been operationalized as the provision of supporting ideas in argumentative writing (e.g., Ong, 2013, 2014; Ong & Zhang, 2010; Révész et al., in press). For example, Ong and Zhang (2010) manipulated multiple task variables, one of which was idea support (i.e., three tasks: no ideas given, ideas given, and both ideas and macro-structure given, in order of decreasing cognitive demands). Their findings showed significant effects of idea support on lexical diversity (more complex tasks leading to greater lexical diversity) but no effect on fluency. A recent study by Révész et al. also attempted to examine the effect of idea support on linguistic complexity (lexical sophistication, lexical diversity, and syntactic complexity), as well as on participants' writing behaviors (pausing and revision behaviors) using the keystroke logging software. For this purpose, the authors collected data from advanced-level participants (CEFR C1 level). The text production results of this study showed that task complexity had a significant influence on lexical sophistication, but no clear effect on lexical diversity and syntactic complexity. Therefore, while both studies examined the effect of idea support on linguistic features in argumentative genre, their findings exhibited somewhat contrasting patterns. From the findings of previous studies on conceptual demands, we can conclude tentatively that the manipulation of idea support does not exert prevalent effects on linguistic features and that studies have found different patterns of idea support effects on language due to task-internal (genre: Kormos, 2011; Révész et al, in press) or learner-internal factors (L2 proficiency: Ong & Zhang, 2010; Révész et al, in press). Particularly, given potentially different patterns of idea support effects across genres, I attempt to explore both genre and idea support as target task variables and illuminate their distinct effects on L2 learners' language use and perceptions, which would enable us to gain a comprehensive picture of task manipulation effects in written discourse.

## **Task-based Studies with Cross-genre Manipulations**

In exploring the effect of reasoning demands on language use, a few studies have operationalized genre as a task complexity variable (e.g., Ruiz-Funes, 2014, 2015; Yang, 2014). This line of research is based on the prediction that argumentative genre that involves logical causal reasoning would be more cognitively demanding to L2 learners than narrative genre. For example, Yang (2014) examined the four genres of narrative, expository, expo-argumentative and argumentative essays (in order of increasing cognitive complexity) composed by adult Chinese EFL learners. Using CAF values as the outcome of task complexity effects, Yang found the lowest values of lexical density and syntactic complexity (e.g., unit-length and phrasal coordination measures) in narrative writing and the highest values of the same syntactic complexity measures in the argumentative task, but no significant genre effects on fluency, accuracy, and some of the complexity measures (e.g., clausal coordination and subordination measures). Based on her findings with a pattern of increased linguistic complexity in the argumentative task, she concluded that her findings provide partial evidence for Robinson's cognition hypothesis.

Additionally, Ruiz-Funes (2015) reported on the findings of two repeated-measures studies regarding task complexity effects on CAF measures: one that explored the writings of foreign language learners of Spanish at an advanced proficiency level and the other at an intermediate level. The participants in the first study were required to complete two writing tasks (analytic and argumentative essays), in which the analytic writing task was predicted to be less cognitively demanding than the argumentative task. In her second study, the participants also completed two writing tasks (personal narrative and expository essays) that concerned the shared topic of study abroad. As regards their relative task complexity, personal narrative writing was

operationalized as the low-complexity task and expository writing as the high-complexity task, with an a priori assumption that providing a thesis and relevant evidence would be more cognitively demanding to L2 learners than narrating personal experience. While there were no statistically significant effects of genre on linguistic features due to a small sample size (study 1:  $N = 8$ ; study 2:  $N = 24$ ), the author found a pattern of increased complexity but lower accuracy and fluency in the writing tasks operationalized as more complex. Ruiz-Funes interpreted these results as evidence in support of Skehan's trade-off hypothesis.

Frear and Bitchener (2015) is another task-based writing study that examined how L2 writers of English show different syntactic and lexical complexity features in three letter-writing tasks manipulated in terms of reasoning demands and number of elements. One methodological issue in this study was that the level of reasoning demands was manipulated only to differentiate between the low- and medium-complexity tasks (but the number of elements manipulated for all three tasks), resulting in a wider gap in cognitive demands between the low- and medium-complexity tasks, compared to that between the medium- and high-complexity tasks. Moreover, while all three tasks were categorized as letter writing, the low- and medium-complexity tasks seem different in their purpose of writing (i.e., the low-complexity task for description and the medium-complexity task for persuasion); therefore, these two tasks can be regarded as distinct genres: descriptive and persuasive writing in a letter format. The findings of this study indicated L2 writers' increased lexical diversity in the high-complexity task, compared to the low-complexity task. Interestingly, while showing no significant change in general subordination (dependent clauses per T-unit), Frear and Bitchener found a significant decrease in a more specific measure of subordination (adverbial clauses per T-unit) with the increase of task

complexity, pointing to the importance of exploring different types of subordination for a clearer understanding of language use and development (Lambert & Kormos, 2014; Rimmer, 2008).

The methodological operationalization of these cross-genre task studies is similar to that of the L1 and L2 genre studies reviewed earlier (e.g., Beers & Nagy, 2009, 2011; Crowhurst, 1980; Lu, 2011; Ravid, 2005; Yoon & Polio, 2017), as the majority of these studies involve the analysis of linguistic feature changes across written genres. As a result, the studies of these two lines have produced very similar results (e.g., higher syntactic complexity, particularly mean unit length, in non-narrative writing than narrative). The two lines of research, however, have been grounded in different assumptions. Specifically, task-based research into cross-genre effects predicts that linguistic features would change notably due to varying levels of cognitive demands imposed by different genres, whereas genre research focuses on functional and pragmatic motivations for different linguistic features. As a result of these contrasting starting points, their findings, although very similar, offer different types of implications. Specifically, the major implication of task-based research would be related to how to promote language development more effectively, so more complex language possibly leading to language development is greatly valued in this research line (i.e., more complex language is better). On the other hand, genre research has its implication for raising better awareness of genre-appropriate communicative purposes and language (i.e., more complex language is not necessarily better).

In this regard, although originally framed as a task complexity study, Frear and Bitchener (2015) argued that their findings might have resulted from functional, which they called “pragmatic”, requirements of the tasks and participants’ personal language choices, which has little to do with cognitive factors. Specifically, the authors suggested that their result of task complexity effects only on a certain type of subordinate clauses (i.e., significant decrease in



adverbial clauses per T-unit in a more complex, persuasion-related task) can be better explained as the outcome of different needs for clause types “as a means of fulfilling the pragmatic requirements of the task” (p. 52).

Similarly, Polio and Yoon (2016) attempted to associate their linguistic complexity findings in two genres (argumentative and narrative writing) with a set of functionally motivated lexico-grammatical features (Biber & Conrad, 2009). Their findings indicated that longer average words (higher lexical sophistication) in argumentative genre actually come from the extensive use of nominalization (e.g., *transportation* and *conclusion*) in argumentative and that of personal pronouns (e.g., *I*, *my*, and *our*) in narratives. Polio and Yoon also found that higher production unit length and complex nominals in argumentative writing arise from the increased use of nominal post-modifiers (e.g., that-relative clauses: *the second reason that the cost of living is*; wh-relative clauses: *students who live off campus are*; and prepositional phrases: *the experience of living off campus*) that are needed more in making logical arguments than in narrating a personal story. Given this plausible explanation of the findings of cross-genre research via functional needs, we need to clarify the roles of genre by exploring the impact of genre differences on learners’ writing performance and task perceptions simultaneously. Additionally, the inclusion of another variable of task complexity, which has been shown to influence L2 writers’ conceptual constraints (i.e., provision of supporting ideas), will enable us to understand how different task manipulations exert varying effects on writers’ perceptions and production.

A way to achieve the separation between learners’ language production and perceptions is to implement an independent measure of task complexity (and other task features such as task difficulty and task motivation) aimed at examining whether task manipulations actually cause

intended cognitive effects (Révész, 2014; Révész, Michel, & Gilabert, 2016). Further, for a meaningful comparison between perception and production results, the use of a wide range of textual features (not limited to traditional CAF measures) would contribute to providing a fuller picture of what textual features interact with learners' task perceptions as a result of the manipulation of genre and task complexity.

### **Validation of Task Complexity Manipulations**

As argued by Révész (2014), the majority of cognitively-oriented TBLT studies have focused on testing how their findings correspond to the existing task-based hypotheses: the cognition hypothesis (Robinson, 2001a, 2001b, 2005, 2007) or the trade-off hypothesis (Skehan, 1998, 2009). Thus, by manipulating task complexity variables in keeping with preexisting assumptions, many researchers have predicted that their participants would experience differing levels of cognitive demands imposed by different tasks, which underlies their language changes (e.g., Kormos, 2011; Kormos & Trebits, 2012; Kuiken & Vedder, 2007, 2008, 2011; Ong, 2013, 2014; Ong & Zhang, 2010; Ruiz-Funes, 2014, 2015; Yang, 2014). However, as discussed above, it should not be presumed that participants perceive and perform different language tasks in full accordance with researchers' intention without an independent perception measure (Révész, 2014). To address the possible incongruence between intended and actual task effects, recent studies began to employ a separate measure of cognitive demands via Robinson's (2001a, 2007) self-rating questionnaire items (e.g., Malicka & Levkina, 2012; Révész, Sachs, & Hama, 2014). In implementing an independent measure of task features, researchers asked their participants to perform a language task and then to complete their self-ratings of perceived task qualities such as task difficulty and task interest.

Thus far, there have been two studies that attempted to validate the impact of task design manipulations on cognitive complexity (Révész et al., 2016; Sasayama, 2016). The authors of the two studies employed very similar methodologies in their validation process and commonly provided evidence for the validity of a self-rating questionnaire. Specifically, Révész et al. examined how three techniques (dual-task methodology, self-ratings, and expert judgments) assess the task complexity of three oral tasks (a picture narrative, a map task, and a decision-making task in order of increasing task complexity). In dual-task methodology, participants are required to complete a primary task simultaneously with a secondary task (e.g., reacting to background color changes) with the prediction that participants will show inferior performance on the secondary task (e.g., slower response or lower accuracy) if the primary task is more complex. Self-ratings and expert judgments are subjective measures of task complexity by way of questionnaires. The authors' findings suggested that participants' subjective self-ratings were consistent with the intended manipulations of task complexity (i.e., more complex tasks rated as more complex and difficult) as well as with other validation techniques, confirming the high validity of subjective self-ratings for assessing the function of task complexity manipulations. Sasayama employed four oral narrative tasks manipulated in terms of the number of elements (different numbers of characters in a story) as a primary task, together with a secondary task of reacting to a color change. Sasayama used participants' reaction time, estimated time for task completion, and self-ratings as independent measures of cognitive complexity, and revealed that large differences in task complexity (e.g., between the simplest task and the most complex task) could be detected by all of the measures. Of the three measures, participants' self-ratings were found to detect cognitive task complexity with the largest effect sizes. These findings from both

studies assured me to employ a self-rating questionnaire as a major independent measure to assess L2 students' perceptions of writing tasks in this study.

### **Text Analysis in TBLT Research**

Traditionally, task-based writing studies have examined CAF measures to test the impact of task feature manipulations (e.g., Ishikawa, 2007; Kormos, 2011; Kuiken & Vedder, 2007, 2008; Ong & Zhang, 2010; Ruiz-Funes, 2015; Tavakoli, 2014; Yang, 2014), providing evidence for either of the two competing task complexity hypotheses (i.e., Robinson's cognition hypothesis and Skehan's trade-off hypothesis). For example, Kuiken and Vedder (2008) examined the letters composed by foreign language learners of Italian and French in terms of syntactic complexity (clausal subordination measures), lexical diversity (type-token ratio measures), and accuracy (number-of-error measures). In their study, the level of task complexity was operationalized as the number of requirements to be considered for choosing the destination. The results of this study indicated that learner language tends to be more accurate (fewer errors) in a more complex task, while there was no significant change between the two tasks in syntactic complexity or lexical diversity. With these results, the authors concluded that their findings offer evidence for L2 writers' greater attention to language forms when involved in a task more complex along the resource-directing dimension, thus giving support to the cognition hypothesis rather than to the trade-off hypothesis.

However, previous task-based research has produced contrasting findings in terms of task complexity effects on CAF probably because different studies have employed different linguistic measures for a single construct (e.g., linguistic accuracy operationalized as ratio of error-free clauses by Kormos, 2011; number of errors per T-unit by Kuiken & Vedder, 2008), manipulated task complexity inconsistently (e.g., conceptual demands operationalized as the availability of

storyline by Kormos, 2014; the number of storylines by Tavakoli, 2014), and adjusted the level of task complexity in different genres (e.g., planning time in narrative by Ellis & Yuan, 2004; planning time in argumentative by Johnson et al., 2012). Furthermore, as discussed above, linguistic features in written discourse are likely to reflect functional needs demanded by a particular task (Frear & Bitchener, 2015; Yoon & Polio, 2017), and TBLT researchers need to go beyond comparing their findings to the two task complexity hypotheses (Kormos & Trebits, 2012) in order to gain a more comprehensive understanding of what aspects of task features motivate notable changes in learner language. These arguments clearly point to the importance of conducting a more comprehensive analysis of textual features in the domain of genre and task complexity research.

In fact, there have been some L2 writing studies that showed significant task type effects on various discourse-level text features (e.g., stance markers: Biber, 2006b; Qin & Uccelli, 2016; Hong & Cao, 2014; temporal cohesion: Kormos, 2011; quantity of ideas: Ong, 2013). For example, the study by Kormos (2011), which I described above, examined cohesion features (causal, temporal, and spatial cohesion features based on the use of connectives, particles, nouns, and verbs) together with widely-used syntactic complexity, lexical complexity, and accuracy measures. The finding of this study indicated that L2 writers tend to use more temporal and logical connectives in the simple condition (narrative task with a given storyline), which was interpreted as the outcome of L2 writers' greater attentional resources available for an explicit indication of cohesive relations in a cognitively less demanding task.

Another plausible area of textual analysis in TBLT research is the use of interactional metadiscourse features (K. Hyland, 2005), which corresponds to increasing attention to the role of stance, authorial voice, or writer-reader interaction (all of which are closely related to the

concept of interactional metadiscourse) in academic writing (Biber, 2006b; Hong & Cao, 2014; Jeffery, 2009; Yoon, in press; Zhao, 2012, 2017; Zhao & Llosa, 2008, among others). In examining quantifiable, text-based features of interactional metadiscourse, researchers have based their studies on Hyland's (2005) model of interactional metadiscourse (e.g., Aull & Lancaster, 2014; Hong & Cao, 2014; Yoon, 2017a; Zhao, 2012, 2017). Hyland's model consists of *stance* (one side including categories such as hedges, boosters, attitude markers, and self-mention) and *engagement* (the other side including categories such as reader pronouns, questions, and directives). In brief, stance involves writer-oriented linguistic features, helping writers to present their opinions and feelings toward a proposition. On the other hand, engagement involves reader-oriented features, intending to perform the recognition of readers' presence and invitation of them as discourse participants.

One study particularly relevant to the present study is Hong and Cao (2014), in which two written genres (descriptive and argumentative essays) were examined with regard to the occurrence of interactional metadiscourse features. Specifically, the authors investigated the essays composed by Chinese, Spanish, and Polish EFL learners, and targeted the categories of hedges, boosters, attitude markers, self-mentions, and engagement markers). The finding of this study showed a significant effect of genre on the amount of hedges and self-mentions (increased use of hedges and self-mentions in argumentative writing), and the authors explained this finding as EFL writers' tendency to take a tentative stance in making arguments. Given these findings in the literature, we can see the necessity to examine textual discourse and metadiscourse features, along with traditional syntactic and lexical features, for a fuller understanding of genre and task complexity effects on written language production. I will discuss the target textual features of this study and justify my selection of them in the Method section.

## **Rationale for the Present Study**

In the previous chapters, I have discussed the importance of a valid operationalization of genre and task complexity in L2 writing research, as well as the necessity of examining a wide range of textual features. Of particular note is that, in task-based writing research, genre has been employed as a task complexity variable with the a priori prediction that the argumentative genre will be more difficult and challenging to L2 writers. However, I believe that researchers should take into account learners' experience and knowledge, as Johns (2008) noted in her review study:

When we read or write in a genre with which we are familiar, and for which we have a schema, we instantiate our schema for what typifies that genre, its conventions, as we read or write, and we use our knowledge of conventions as we produce a new text. The conventions of a genre can refer to a variety of features: the text structure, the register, the relationships between the writer and the audience. (p. 241)

What we can infer commonly from Johns' explanation and Hayes' (1996, 2012) model of writing is that genre schemas do play an important role for learners' writing performance in different genres. Then, it can be further assumed that if adult ESL students are familiar with argumentative writing, a typical genre in post-secondary education and testing, they will not find argumentative writing more difficult or cognitively demanding than other genres such as narrative writing, thus challenging a stereotype about genre-specific cognitive demands. It is even conceivable that, for ESL students whose English writing practice have mostly been in settings of standardized test preparation, narrating a story can be very challenging due to their limited schema for this unfamiliar genre.

To address these issues, I explore shared and unique effects of genre and task complexity manipulations on ESL students' language production and task perceptions. For task perception

data, I use a self-rating questionnaire that has been shown to measure participants' cognitive processes validly (Révész et al., 2016; Sasayama, 2016). Here, by collecting data from ESL teachers and ESL students, I attempt to reveal a potential gap between ESL teachers' expectations of different task types and ESL students' actual perceptions of the tasks.

As regards text analysis, I do not limit the focus of this study to traditional CAF measures or to the validity of the competing two task complexity hypotheses, but rather I examine a wide array of textual features at multiple construction levels (i.e., syntactic, lexical, discourse, and metadiscourse) to shed light into the interaction between learners' linguistic features and task perceptions, thereby explicating distinct roles of communicative functions and cognitive constraints in shaping particular linguistic features. Additionally, I examine how a group of textual features predict essay scores of each genre distinctively in an attempt to gain a detailed picture of task type effects on L2 written language production and writing performance. For the benefit of speed and reliability, I take advantage of automated processing techniques, which I will introduce in detail in the Method section.

This study is grounded in Hayes' (1996, 2012) model of writing that fully recognizes moderating effects of L2 writers' linguistic, genre, and task knowledge on writing processes and performance. With regard to the relationship between L2 proficiency and task complexity effects, unlike the majority of task-based writing studies that targeted L2 writers at one proficiency level, Ruiz-Funes (2015) examined the essays written by advanced- and intermediate-level L2 learners, and found an interaction between L2 proficiency and task complexity effects, that is, a clear difference between the two learner groups in the way they responded to the writing tasks with different levels of cognitive complexity. It is also plausible that a language task is already too complex or too simple for a learner at a certain proficiency level to observe any significant



change in language induced by the manipulation of task complexity. For example, Frear and Bitchener (2015) suggested that their limited task effect might have resulted from the incorrect alignment of task manipulations with participants' L2 proficiency. To take into account the interplay of L2 proficiency and task variables for language production, several task-based studies (e.g., Kuiken & Vedder, 2008; Sasayama, 2016; Yang, 2014) employed an additional measure of L2 proficiency, a cloze test, encouraging me to make the same methodological decision of implementing a cloze test in this study.

With this study, I intend to offer implications for L2 writing research, pedagogy, and assessment. The present study will provide evidence of valid interpretations regarding the effect of genre and task complexity manipulations in writing research. L2 writing instructors and task developers will be informed about the possibility of constructing independent writing tasks with various genres and task complexity levels to achieve an appropriate alignment of task features with target language learners. It has been suggested that learners' performance varies across task types (Bouwer et al., 2015). There is also a belief that argumentative writing is the most suitable genre for assessment, making it a dominant genre in the testing context (Qin & Karabacak, 2010). With the findings of this study, I will offer insights into the relationship between genre/task manipulations and valid writing assessment. The present study is guided by the following research questions:

1. How do ESL students and teachers perceive writing tasks manipulated in terms of genre and task complexity?
2. What are the effects of genre and task complexity on textual features in ESL writing?
  - 2.1. How does ESL students' L2 proficiency interact with task type effects on textual features?

3. What are the effects of genre and task complexity on ESL writing scores?
  - 3.1. How does ESL students' L2 proficiency interact with task type effects on writing scores?

## CHAPTER 3.

### METHOD

#### Participants

**Student participants.** For this study, I collected data from 76 ESL students enrolled in an English language program at a large U.S. university. The course levels in the program ranged from 0 to 5, and I recruited participants from the highest level of the program. Level 5 courses, which are also referred to as English for Academic Purposes (EAP) courses, are intended to develop academic writing proficiency of international students who do not use English as their native language. International students admitted to the university with an iBT TOEFL score below 79 (paper-based test 550) are required to take the placement test administered by the English language program (multi-skills test including reading, listening, and writing sections) and are assigned to a certain course level based on their performance. One issue here is that some level 5 students advanced from level 4, while others placed into level 5 based on their performance on the placement test. Students assigned directly to level 5 are often more proficient than those who moved up from a lower level, making me anticipate that some proficiency variation exist among students in the same level. Therefore, I employed a cloze test as an objective measure of L2 proficiency at the beginning of data collection.

The participants received four to six hours of L2 writing instruction (2 or 3 classes) per week during a 15-week-long semester. The primary objectives of level 5 courses were to prepare students for university-level academic courses and to help them build a clear understanding of the audience in academic writing. The largest portion of the course grades involved multi-draft essay writing and revision (50% to 70% of the course grade), indicating the emphasis of the courses on writing processes. The specific goals of the courses were to develop students'

summarizing, paragraphing, and revising skills, as well as to guide them in accomplishing several multi-draft and timed writing tasks. Quotation and citation skills were also targeted. Much of the class time was used to develop various academic writing skills (e.g., how to construct a well-developed paragraph), so students' participation in this research was their major practice for timed writing.

Based on the minimum test score needed for course enrollment as well as course objectives, the majority of the participants could be described as ESL students at the high intermediate or low advanced level, who are capable of meeting practical writing needs and composing a multi-paragraph essay within 30 minutes. Forty-six male and 30 female students participated, and they were all undergraduate students. Their ages ranged from 18 to 27, with a mean of 19. They came from various countries (e.g., Angola, China, France, Japan, Malaysia, South Arabia, South Korea, Taiwan, Thailand, and Turkey). Fifty participants spoke Chinese as their L1, and seven were Arabic native speakers. The remaining participants were either native speakers of Korean ( $n = 6$ ), Japanese ( $n = 3$ ), Portuguese ( $n = 3$ ), Malay ( $n = 2$ ), Thai ( $n = 2$ ), Turkish ( $n = 2$ ), or French ( $n = 1$ ). Their mean length of English study was 104.7 months, and their mean length of staying in the United States was 15.1 months. From these responses, it can be inferred that, despite their stay in the United States at the time of data collection, the participants' English learning had been mostly in their own countries, that is, English as a foreign language (EFL) settings. Table 2 presents the demographic information of the student participants.

Table 2.

*Demographic Characteristics of the ESL Student Participants*

Characteristics		<i>N</i> = 76
Age: Mean ( <i>SD</i> )		19.11 (1.45)
Gender	Male	46
	Female	30
First language	Chinese	50
	Arabic	7
	Korean	6
	Japanese	3
	Portuguese	3
	Malay	2
	Thai	2
	Turkish	2
	French	1
Length of English study (months): Mean ( <i>SD</i> )		104.66 (48.04)
Length of stay in the United States (months): Mean ( <i>SD</i> )		15.07 (15.36)

**Teacher participants.** For the comparison between students' perceptions and teachers' expectations of writing tasks, I collected survey data from 30 ESL teachers using an online survey platform, Qualtrics. Most teachers were English native speakers ( $n = 26$ ), and there were four teachers whose native language was not English (Iranian:  $n = 1$ ; Korean:  $n = 1$ ; Russian:  $n = 1$ ; and Turkish:  $n = 1$ ). Their mean length of teaching English was 11.7 years ( $SD = 7.9$  years), and that of teaching English writing skills was 9.4 years ( $SD = 7.0$  years). Their ages ranged from 26 to 65 with a mean age of 39 ( $SD = 9.15$ ). Twenty-three teachers were female, and seven were male. While all teachers were teaching English in the United States at the time of data collection, three teachers reported that their major field of English teaching had been college-level EFL contexts (i.e., the major setting of 27 teachers had been college-level ESL classrooms).

## Instruments

**Questionnaires.** I devised questionnaires for participants' background information and task perceptions. The background questionnaire includes the items related to participants' demographic information (e.g., age, gender, educational background, language background, and length of stay in the United States). The questionnaire for ESL students' task perceptions contains six statements relating to each writing task. The six items aim to measure how participants perceive (1) the mental effort induced by the task (task complexity), (2) task difficulty, (3) task anxiety, (4) task confidence, (5) task interest, and (6) task motivation (adapted from Robinson, 2001a; Révész et al., 2016). Traditionally, Robinson (2001a, 2007) assessed the five dimensions with no inclusion of the item related to the mental effort, but following the suggestion by Révész et al. (2016), I differentiated between task complexity and task difficulty in the questionnaire so that their distinct constructs can be fully captured (Brünken, Seufert, & Paas, 2010). Immediately after completing each writing task, participants were asked to judge each statement on a 9-point Likert scale. The following were the exact items:

This task required no mental effort at all.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	This task required extreme mental effort.
This task was not difficult at all.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	This task was extremely difficult.
I felt really relaxed doing this task.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	I felt frustrated doing this task.
I didn't do well on this task.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	I did well on this task.
This task was not interesting at all.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	This task was very interesting.
I don't want to do more tasks like this.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	I want to do more tasks like this.

The teacher participants were given the online survey that contains Likert scale items for task perceptions and open-ended response items for additional comments. For a direct comparison with the perception results of the students, it was clearly stated at the beginning of the teacher survey that the writing tasks were designed as 30-minute timed writing tasks for L2

learners at high-intermediate or low-advanced proficiency levels (i.e., level 5 students in the English language program). For teacher perceptions of the tasks, I asked the teacher participants to read each of the writing task prompts carefully and complete six questionnaire items measuring the same constructs as those targeted by the student survey (i.e., task complexity, task difficulty, task anxiety, task confidence, task interest, and task motivation).

To attain this goal of tapping the same constructs, instead of constructing new questionnaire items, I simply manipulated the wording of the task perception statements. For example, I modified a task difficulty statement *This task was not difficult at all* to a hypothetical statement *This task will not be difficult for ESL students at all*, and a task motivation statement *I don't want to do more tasks like this* to *ESL students will not want to do more tasks like this*. Four sets of Likert scale items were prepared for the four writing tasks, and these sets were given to the teacher participants in a randomized order, which was an attempt to control for potential sequence effects on task judgments. The exact items were as follows:

This task will require no mental effort of ESL students at all.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	This task will require extreme mental effort of ESL students.
This task will not be difficult for ESL students at all	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	This task will be extremely difficult for ESL students
ESL students will be really relaxing doing this task.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	ESL students will be frustrated doing this task.
ESL students will not do well on this task.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	ESL students will do well on this task.
This task will not be interesting to ESL students at all.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	This task will be very interesting to ESL students.
ESL students will not want to do more tasks like this.	1 - 2 - 3 - 4 - 5 - 6 - 7 - 8 - 9	ESL students will want to do more tasks like this.

The open-ended questions included in the teacher survey were about (1) their impression of the writing tasks, (2) reasons for different task perceptions, and (3) possibility of using the tasks in their class. It took approximately 30 minutes for the teachers to finish the survey. In return for their participation, they received a \$10 gift certificate via email.

**Writing prompts.** I developed two argumentative and two narrative writing prompts. The argumentative prompts required participants to make logical arguments on foreign language learning and use. The narrative prompts involved narrating a personal story related to a similar topic. When developing the task prompts, I consulted three SLA experts and one test developer, and improved the quality of the prompts based on their comments and recommendations. Within each genre, I manipulated the level of conceptual demands operationalized as the provision of supporting ideas. The writing tasks with lower conceptual demands were given to the participants with some information (i.e., example storylines for narrative and main ideas for argumentative) that they could utilize while writing. I intended for the ESL students to find the tasks with idea support cognitively less demanding and less difficult than the tasks with no such support. Throughout the manuscript, the argumentative and narrative prompts with idea support are labeled as *Arg/+Support* and *Nar/+Support* respectively (then, the prompts with no idea support as *Arg/-Support* and *Nar/-Support*).

To avoid potential topic effects on learner language (e.g., Hinkel, 2002; Tedick, 1990), I devised the prompts that shared the topic of foreign language learning or use, but, at the same time, to minimize task repetition effects, I attempted to develop somewhat distinct prompts. Specifically, a narrative prompt with no idea support (*Nar/-Support*) elicited a positive experience related to foreign language use (*Tell a story about ONE of your positive experiences related to foreign language use*), while the other narrative prompt (*Nar/+Support*) elicited a



difficult experience related to interactions using a foreign language (*Tell a story about ONE of your difficult experiences related to interactions using a foreign language*). For argumentative writing, one prompt (Arg/-Support) involved the necessity of using a foreign language fluently in the globalized era (*Write an essay about whether you agree or disagree with the statement about the necessity of foreign language abilities*), while the other (Arg/+Support) entailed the relationship between the ability to speak a foreign language and the possibility of having a successful life (*Write an essay about whether you agree or disagree with the statement about the relationship between foreign language abilities and success*) (Appendix A for the full prompts).

**Rubric.** The essays were evaluated using a revised analytic scale (Polio, 2013), adapted from the ESL composition profile (Jacobs et al., 1981) that comprises the five subscales of content, organization, vocabulary, language use, and mechanics (see Appendix B). According to Connor-Linton and Polio (2014), the revised rubric includes the same five categories as the original scale, but their descriptors and weighting were modified to better reflect what trained raters had noted regarding actual changes in L2 writing skills over time. The full score of each subscale is 20 points, except for one subscale, mechanics, whose full score is 10 points (i.e., total score of the rubric = 90 points); this unequal weighting is based on much fewer points for mechanics in the original Jacobs et al. rubric.

The content subscale evaluates a full development of ideas, inclusion of detailed and interesting content, and topic relevance. The organization subscale assesses overall organization, clear thesis statement, coherence (unity within and across paragraphs), and cohesion (use of connectors and transition words). The vocabulary subscale includes descriptors related to lexical sophistication, lexical accuracy, idiomatic vocabulary use, and academic register. The language use subscale evaluates syntactic complexity, syntactic variety, and syntactic and morphological

accuracy. The last subscale, mechanics, evaluates paragraph indentation, spelling accuracy, and punctuation accuracy. This revised analytic rubric was chosen because this study aims to reveal the effect of genre and task complexity manipulations on various categories of writing scores, and because this particular rubric was found to produce more reliable and valid scores than other rubrics (Connor-Linton & Polio, 2014; Polio, 2013).

## **Procedures**

**Data collection.** After obtaining approval from the IRB, I contacted ESL teachers who were teaching levels 5 writing courses at the English language program. Five instructors teaching a level 5 writing course (course names: ESL 220 and ESL 221) gave me permission to use their class time for data collection. After discussing my study details with the instructors, I visited their classes to obtain permission from students. All writing tasks were administered to all students as a part of the classroom curriculum, which helped students prepare for their final timed writing exam. All students were informed that research participation was completely voluntary and that I would not have access to any of the essays without their consent. I also informed students that their participation would be compensated with a \$15 gift certificate and my feedback on their essays (specifically, receiving error-code feedback within a week from a writing session). To make students take this study seriously, I also told them that in each class two students who write the best essays would be selected and given a \$25 gift certificate. For students who were not willing to participate, their instructors were going to give a similar type of feedback so that no one would be at a disadvantage, but all students enrolled in the five writing courses agreed to participate.

To minimize potential testing effects from a repeated-measures design, I collected data at one-week intervals, with the order of the writing prompts fully counterbalanced (see Table 3 for

the summary of data collection procedures). I only used data from the participants who completed all four writing sessions, and the essays written by the students who missed one or more sessions were excluded from analysis. This process led four students to be excluded (from 79 students originally to a final sample of 76 students).

Table 3.

*Counterbalanced Data Collection Procedures*

Group	Week 1	Week 2	Week 3	Week 4	Week 5
A ( <i>n</i> = 20)	Cloze	Nar/-Support Task survey	Arg/-Support Task survey	Nar/+Support Task survey	Arg/+Support Task survey Background
B ( <i>n</i> = 19)	Cloze	Arg/+Support Task survey	Nar/-Support Task survey	Arg/-Support Task survey	Nar/+Support Task survey Background
C ( <i>n</i> = 18)	Cloze	Nar/+Support Task survey	Arg/+Support Task survey	Nar/-Support Task survey	Arg/-Support Task survey Background
D ( <i>n</i> = 19)	Cloze	Arg/-Support Task survey	Nar/+Support Task survey	Arg/+Support Task survey	Nar/-Support Task survey Background

*Note.* Arg/-Support = argument task with no idea support; Arg/+Support = argument task with idea support; Nar/-Support = narrative task with no idea support; Nar/+Support = narrative task with idea support.

In the first week of data collection, to assess students' general English proficiency, I implemented a cloze (fill-in-the-blank) test, which was developed and validated by Brown (1978). I decided to use a cloze test as a measure of L2 proficiency because previous research has suggested adequate validity of cloze tests for assessing general language proficiency (e.g., Brown, 2002; Fotos, 1991; Hinofotis, 1980; Tremblay, 2011). The cloze test adopted in this study is *Man and His Progress* that includes 399 words and 50 blanks (deletion pattern of every 7th word; see Appendix C). Participants were given clear instructions and an example of how to fill

in the blanks, and then they had 30 minutes to complete the cloze test. The answers for the cloze test are scored using the acceptable-answer method that marks all contextually acceptable items as correct answers. I chose this scoring method because it was found to surpass other methods, such as the exact-answer or multiple-choice techniques, in validity, reliability, and item discrimination (Brown, 1980).

I evaluated students' performance on the cloze test using an answer sheet adapted from Yang (2014). The result of the cloze test was found to be reliable (Cronbach's  $\alpha = .84$ ), indicating its consistency in distinguishing among the participants ( $M = 28.20$ ,  $SD = 7.55$ ). For follow-up analyses that include L2 proficiency as a predictor variable, I categorized the student participants into different proficiency groups. Twenty-nine students who received cloze test scores equal or higher than 31 were assigned into the high proficiency group, while 28 students who received scores equal or lower than 25 were assigned into the low proficiency group (see Table 4). Nineteen students whose scores were in between those of the two groups were excluded for these analyses in order to assure a greater gap between the proficiency groups.

While dividing participants into separate groups based on performance scores (i.e., categorization of a continuous variable) involves the risk of losing much statistical power (see Plonsky & Oswald, in press), the sample size of each group still appears to be adequate for inferential statistics (29 for the high proficiency group and 28 for the low proficiency group), and the structure of the current data set (i.e., repeated measures with four different writing tasks) fits these statistical procedures better. For the same reasons, several recent task-based studies employed similar statistical analyses in exploring the interaction between task and L2 proficiency effects (e.g., Kuiken & Vedder, 2008; Sasayama, 2016; Yang, 2014).

Table 4.

*Demographic Characteristics of the High and Low Proficiency Group Students*

Characteristics		High proficiency ( $n = 29$ )	Low proficiency ( $n = 28$ )
Age: Mean ( $SD$ )		18.97 (1.09)	19.54 (1.88)
Gender	Male	16	19
	Female	13	9
First language	Chinese	22	14
	Arabic	2	3
	Korean	1	4
	Japanese	0	3
	Portuguese	1	2
	Malay	2	0
	Thai	0	1
	Turkish	0	1
	French	1	0
Length of English study (months): Mean ( $SD$ )		114.62 (40.75)	87.64 (50.49)
Length of stay in the United States (months): Mean ( $SD$ )		14.14 (15.85)	18.14 (17.02)

From the second to the fifth week, the participants composed timed essays (each under the time constraint of 30 minutes). They were not allowed to use dictionaries or other resource tools while writing. Immediately after writing, they were asked to complete a task perception questionnaire that contained six task statements. In the last week of data collection, following the steps of a writing task and task questionnaire, the participants completed a background questionnaire designed to obtain their demographic information. After collecting all essays, I transcribed them verbatim.

**Essay scoring.** Using the analytic scoring rubric introduced above, two expert raters evaluated the transcribed essays. Both raters were Ph.D. students in a language-related major and had previous experience in rating timed essays administered by the English language program.

The two raters first participated in a two-hour training session to ensure grading consistency. The raters examined the descriptors of the rubric and the prompts used for this study. I asked the raters to focus on assigning scores that fully reflect the quality of an essay; in other words, I asked them not to adjust their level of leniency (or stringency) according to the task type. It was my attempt to elicit essay scores that accurately reflect L2 writers' task-specific performance. With these points in mind, the raters completed an iterative process of rating an essay and discussing its scores. They were instructed to use any of the integer numbers within a score range. When subscale scores differed by 3 or more, the raters examined the rubric descriptors again and adjusted their scores after a short discussion. The raters continued this norming process until they reached full agreement or subscale scores that differed only by one or two. Eight essays that were not a part of this study data were used for training purposes.

After the norming session, each rater evaluated the entire data set (304 essays) independently. The raters were given the essays in random order, but they were informed of the task type of each essay (i.e., Arg/-Support, Arg/+Support, Nar/-Support, or Nar/+Support) so that they could assign essay scores most relevant to the topic of each task type. After their work of ratings all the essays, the raters were compensated with \$450. The inter-rater reliability of total essay scores was  $r = .84$  (each subscale: content  $r = .81$ ; organization  $r = .78$ ; vocabulary  $r = .66$ ; language use  $r = .72$ ; mechanics  $r = .60$ ), generally indicating an acceptable level of inter-rater reliability (Brown, Glasswell, & Harland, 2004). This study used the average scores of the two raters. For the essays that were assigned seriously discrepant scores (subscale scores differing by 3 or more), a third rater assigned new scores, and the two close scores were used.

## Text Features

For a detailed text analysis that take into account the multi-faceted nature of writing proficiency and the traits included in the rubric, I employed four natural language processing (NLP) tools that generate a wide array of linguistic, discourse, and metadiscourse features: L2 Syntactic Complexity Analyzer (henceforth SCA; Lu, 2010), Coh-Metrix (McNamara, Graesser, McCarthy & Cai, 2014), the Multidimensional Analysis Tagger (henceforth MAT; Nini, 2015), and the Authorial Voice Analyzer (henceforth AVA; Yoon, 2017a). The use of these automated tools was motivated to respond to a call to address multidimensional features of linguistic complexity (Lu, 2010; Norris & Ortega, 2009) and also to explore discourse and metadiscourse features beyond traditional CAF measures in task-based research. Measures of syntactic complexity were obtained using SCA, MAT, and Coh-Metrix. Coh-Metrix and MAT were further used for lexical and discourse-level features. Last, AVA was employed for interactional metadiscourse features. In this study, I decided not to explore linguistic accuracy or fluency because my previous research (Yoon & Polio, 2017) that examined genre effects longitudinally confirmed that error-count accuracy and fluency measures did not differ significantly by genre (also the lack of development over time).

Given an extremely large number of textual features that these tools compute, I selected target measures based on the criteria of redundancy, validity, and construct distinctiveness. To give examples related to SCA measures, clauses per sentence (C/S) is a measure of clausal embeddings that tap both subordination and coordination, but these two constructs should be measured using two distinct measures (clauses per T-unit (C/T) for subordination and T-unit per sentence (T/S) for coordination) to reflect language development more clearly (Norris & Ortega, 2009). This led to the exclusion of C/S. Also, verb phrases per T-unit (VP/T) and complex T-

units per T-unit (CT/T) that have been found to be less valid as language development indicators (Lu, 2011) were excluded in the present study. Of the three unit-length measures that SCA generates (mean length of sentence, mean length of T-unit, mean length of clause), I included mean length of sentence and mean length of clause because these two measures, unlike mean length of T-unit, were shown to tap two distinct constructs (Yoon, 2017b). For other measures tapping a very similar construct (e.g., complex nominals per T-unit and complex nominals per clause), following Yang et al. (2015), this study included only one measure that had the clause as its base unit.

**Syntactic complexity features.** This study involves the construct of syntactic complexity at the clause- and phrase-levels. SCA was used for the calculation of clause-level syntactic measures. They include mean length of production units (mean length of sentence and mean length of clause) and subordination (clauses per T-unit). Because clausal coordination (T-units per sentence) captures beginning-level language development (Bardovi-Harlig, 1992; Norris & Ortega, 2009), I decided not to include clausal coordination measures in this study that targets ESL students at a high intermediate or low advanced level. Unit-length and subordination measures have been widely adopted as language development indicators (see Bulté & Housen, 2012; Ortega, 2003; Wolfe-Quintero, Inagaki, & Kim, 1998); in particular, it was found that subordination functions as a valid developmental measure for intermediate proficiency levels. However, several recent studies have shown that subordination as a unitary construct (e.g., overall subordination ratio) failed to detect language development over a short period of time (e.g., Bulté & Housen, 2014; Mazgutova & Kormos, 2015), and also it was not sensitive enough to reflect genre variation (e.g., Lu, 2011; Yoon & Polio, 2017).



In this regard, challenging the tradition of L2 research that measures subordination as a single construct, Lambert and Kormos (2014) specifically argued for the need to explore these different clause types separately to show “developmental variation during task performance” (p. 608). Also, recent L2 research has begun to examine such more specific clause types as target measures and suggested distinct patterns of linguistic variation across tasks (e.g., Frear & Bitchener, 2015; Staples & Reppen, 2016) as well as L2 developmental trajectories (e.g., Vercellotti & Packer, 2016). Therefore, together with a general measure of subordination ratio, I computed more specific measures of subordination that involve three distinct syntactic relations (nominal clauses, adverbial clauses, and adjectival clauses).

Simply put, nominal clauses that may have a complementizer optionally serve as objects of superordinate verbs (e.g., *I discovered that each culture has its own communication method.*). Adverbial clauses are dependent clauses that modify superordinate verbs and are associated with main clauses using a subordinate conjunction (e.g., *While I was talking, other people started to interrupt.*). Adjectival clauses (also called relative clauses) modify nouns to specify their meaning with the optional use of a relative pronoun (e.g., *Individuals who can speak foreign language can spread their own culture to foreigners.*) (Collins & Hollo, 2010; Nippold, Hesketh, Duthie, & Mansfield, 2005).

To obtain the density values (occurrences per 1,000 words) of nominal, adverbial, and adjectival clauses, I availed myself of MAT, an automated processing tool originally developed for the multidimensional analysis. MAT annotates specific tags to raw texts using the Stanford POS Tagger (Toutanova, Klein, Manning, & Singer, 2003) and then calculates normalized frequencies of various linguistic features that include verb tenses, syntactic patterns, discourse markers, and so forth. The density of nominal clauses was calculated based on the summed

frequencies of that verb complements, subordinator that deletion, and Wh-clauses. The density of adverbial clauses was based on the occurrences of past participial clauses, present participial clauses, causative adverbial subordinators, concessive adverbial subordinators, conditional adverbial subordinators, and other adverbial subordinators. Last, the density of adjectival clauses was calculated using the frequencies of that relative clauses (both subject and object positions), pied-piping relative clauses, Wh-relative clauses (both subject and object positions), past participial relatives, and present participial relatives (see Nini, 2015).

Based on the findings that showed advanced writers' increased use of grammatical metaphor and nominalization (Halliday & Mathiessen, 1999), writing researchers are giving increasing attention to phrasal-level complexity (e.g., Biber & Gray, 2010; Biber, Gray, & Poonpon, 2011; Ortega, 2003; Parkinson & Musgrave, 2014). Phrasal-level complexity measures, as distinct features of academic writing, have been found to be valid predictors for language development and overall writing quality (e.g., Biber et al., 2011; Bulté & Housen, 2014; Lu, 2011). In this study, I explore phrasal-level syntactic complexity by investigating indices such as the number of complex nominals per clause, number of words before the main verb (degree of left embeddedness), and number of modifiers per noun phrase, all of which tap the multidimensional construct of noun phrase sophistication. Additionally, I measure the number of coordinate phrases per clause, which was found to differ across genres (Lu, 2011).

**Lexical features.** I assessed two lexical measures: lexical sophistication (word frequency) and lexical diversity (vocc-D) obtained from Coh-Metrix. Lexical sophistication addresses the various constructs of average word length, word frequency, and nominalization, whose measures have been regarded as effective predictors of lexical proficiency because L2 writers tend to use longer, infrequent words with an increasing proportion of nominalizations as

their proficiency improves (e.g., Biber, 1988; Crossley, Cobb, & McNamara, 2013; Jarvis, Grant, Bikowski, & Ferris, 2003; Laufer & Nation, 1995). In addition, for lexical diversity, I examined vocd-D that is known to sufficiently address text length effects (McCarthy & Jarvis, 2010) and to validly reflect language development (Crossley, Salsbury, McNamara, & Jarvis, 2010; Treffers-Daller, 2013). These lexical measures were also found to reflect genre differences effectively: higher lexical sophistication and lower lexical diversity in argumentative writing (Yoon & Polio, 2017). In task-based research, there have been some contrasting findings about the effect of conceptual demands on lexical diversity (i.e., increased lexical diversity in the complex task by Ong & Zhang, 2010; little difference in lexical diversity across tasks by Révész et al., in press). Given these findings, the exploration of word frequency and vocd-D in this study will advance our understanding of how these lexical features interact with task type and L2 proficiency.

**Discourse features.** I measured five discourse features obtained from Coh-Metrix and MAT. They include coreference cohesion, conceptual cohesion, causal connective density, temporal connective density and nominalization density. Cohesion generally indicates the link between ideas in the text that can be achieved with the help of text cohesion cues at three different levels: local, global, and text levels (Crossley, Kyle, & McNamara, 2016). Of these different levels of cohesion, this study targets local cohesive devices that involve lexical and semantic overlap between adjacent sentences. Coreference cohesion is measured through argument (nouns and pronouns) overlap for adjacent sentences. Conceptual cohesion is measured in terms of how two adjacent sentences are related conceptually and thematically. For conceptual cohesion, Coh-Metrix exploits Latent Semantic Analysis (LSA), a statistical method to explore the underlying semantic associations between textual segments (Landauer, Foltz, & Laham, 1998). Additionally, I assessed the normalized frequencies of three discourse markers that

apparently contribute to genre-specific communicative functions: causal connective density (e.g., *because*, *consequently*, and *accordingly*), temporal connective density (e.g., *first*, *until*, and *finally*), and nominalization density (number of normalized words with a derivational suffix; e.g., *carelessness*, *difficulty*, and *investigation*).

It has been widely acknowledged that the extensive use of these cohesion measures helps the reader better understand the text by facilitating the association between the ideas in the text (Crossley, Yang, & McNamara, 2014; Gernsbacher, 1990), but previous studies have produced inconsistent findings regarding the contribution of cohesive devices to writing quality (e.g., Crossley & McNamara, 2012; McNamara, Crossley, & McCarthy, 2010; Yang & Sun, 2012). L1 writers are likely to experience a transition from a stage of extensive local cohesion use to a next stage focusing on constructing complex sentences (Haswell, 2000; McCutchen & Perfetti, 1982), and I postulate that L2 writers will show different patterns of trade-offs, for example, between linguistic complexity and cohesion depending on their proficiency, and that genre and task complexity manipulations exert some influence on L2 writers' use of discourse-level features.

**Interactional metadiscourse features.** Using AVA, I examine the density of various interactional metadiscourse features. Built using a regular expression function in Python and Stanford Parser (Klein & Manning, 2003), AVA calculates normalized frequencies of interactional metadiscourse features (i.e., hedges, boosters, attitude markers, self-mention, reader mention, directives, and questions), motivated by the model of interactional metadiscourse (K. Hyland, 2005). Using 261 EFL argumentative essays, Yoon (2017) examined how AVA measures predict the holistic ratings of voice strength and essay quality. The finding of this study showed that three features (i.e., self-mentions, boosters, and attitude markers) explained 26% of the variance in voice strength scores, while none had a notable contribution to essay quality.

Relevant to this study is a recent corpus study that found clear genre effects on the use of hedge and self-mention markers (Hong & Cao, 2014). In this study, I focus on the density of hedges, boosters, self-mentions, and reader-mentions that have been specifically targeted by many EAP and corpus studies (e.g., Hu & Cao, 2011; K. Hyland & Milton, 1997; Lee & Deakin, 2016, among others). Table 5 presents a summary of the text features explored in this study.

Table 5.

*Target Text Features*

Construct	Measure	Description	Tool
Length of production unit	Mean length of sentence (MLS)	# of words / # of sentences	SCA
	Mean length of clause (MLC)	# of words / # of clauses	SCA
Subordination	Clauses per T-unit (C/T)	# of clauses / # of T-units	SCA
	Nominal clause density (NOMC)	# of nominal clauses * 1000 / # of words	MAT
	Adverbial clause density (ADVC)	# of adverbial clauses * 1000 / # of words	MAT
	Adjectival clause density (ADJC)	# of adjectival clauses * 1000 / # of words	MAT
Phrasal complexity	Coordinate phrases per clause (CP/C)	# of coordinate phrases / # of clause	SCA
	Complex nominals per clause (CN/C)	# of complex nominals / # of clauses	SCA
	Left embeddedness (LEFT)	# of words before the main verb	Coh-Metrix
	Modifiers per noun phrase (MOD/N)	# of modifiers / # of noun phrases	Coh-Metrix
Lexical features	vocd-D (D)	Based on vocd-D formula	Coh-Metrix
	Word frequency (WF)	Based on the CELEX corpus	Coh-Metrix
Discourse	Coreference cohesion	Argument overlap between adjacent sentences	Coh-Metrix
	Conceptual cohesion	Semantic overlap between adjacent sentences	Coh-Metrix
	Causal connective density	# of causal connectives * 1000 / # of words	Coh-Metrix
	Temporal connective density	# of temporal connectives * 1000 / # of words	Coh-Metrix
	Nominalization density	# of nominalizations * 1000 / # of words	MAT
Metadiscourse	Hedge density	# of hedges * 1000 / # of words	AVA
	Booster density	# of boosters * 1000 / # of words	AVA
	Self-mention density	# of self-mentions * 1000 / # of words	AVA
	Reader pronoun density	# of reader pronouns * 1000 / # of words	AVA

## Analysis

For the first research question regarding task perceptions, I performed three-way mixed ANOVAs with group (student and teacher) as a between-subjects variable and genre (argumentative and narrative) and idea support (no support and support) as within-subjects variables. Prior to the main statistical analysis, I checked assumptions for mixed ANOVAs. As a first step, I checked for the normality of distribution by using Shapiro-Wilk tests ( $\alpha = .05$ ). For the dependent variables that had a significant result of this normality test, I calculated z-scores of skewness and kurtosis to determine whether their distribution was within acceptable limits (i.e., absolute z-score values under 3.29; Kim, 2013). This analysis revealed that the distribution of all variables was within acceptable limits (z-scores of skewness ranging from -2.10 to 2.27; z-scores of kurtosis ranging from -2.04 to 1.13). Additionally, I performed Levene's test for homogeneity of between-group variances and found that all dependent variables failed to reject the null hypothesis ( $\alpha = .05$ ), confirming the appropriacy of the data set for mixed ANOVAs. The alpha level of all inferential statistic results was set with the Bonferroni adjustment.

To answer the second research question regarding task manipulation effects on textual features, I computed a series of two-way ANOVAs with genre and task complexity as within-subjects variables. The dependent variables that included 21 textual features at different construct levels (syntactic complexity, lexical complexity, discourse, and metadiscourse) were found to have limited correlations. Given the lack of linear relationship between dependent variables, I decided not to run multivariate analyses. Before conducting the main analysis, I checked assumptions for repeated-measures ANOVAs. I tested the assumption of normality by examining Shapiro-Wilk test results and z-scores of skewness and kurtosis. For the variables that rejected

the null hypothesis of Shapiro-Wilk tests, I calculated z-scores of their skewness and kurtosis values. This analysis informed me that four variables were not within acceptable limits: mean length of sentence, coordinate phrases per clause, left-embeddedness, and self-mention density. Considering their moderately positively skewed distribution, I transformed their values using a square root transformation (Tabachnick & Fidell, 2001). As a result, the distribution of these variables became suitable (i.e., skewness and kurtosis within acceptable limits) for two-way ANOVAs. While using transformed values for inferential statistics, I report untransformed values for means and standard deviations for ease of interpretation in Table 12.

For the third research question about genre and task complexity effects on writing scores, as I did for the analysis related to the second research question, I checked the assumption of normality by testing the significance of Shapiro-Wilk tests and, subsequently, by examining z-scores of skewness and kurtosis for variables with significant results. This analysis showed that the distribution of all variables was within acceptable limits (z-scores of skewness ranging from -2.76 to 0.14; z-scores of kurtosis ranging from -1.11 to 2.28).



## CHAPTER 4.

### RESULTS

#### Task Perceptions

The descriptive results of the perception data are presented in Table 6. The first column *Item* indicates each of the statements included in the questionnaires. *Complexity*, for example, refers to a statement tapping into the construct of task complexity (*this task required extreme mental effort*) rather than the actual manipulation of task complexity (provision of supporting ideas). To avoid confusion, throughout this chapter, I use *idea support* in indicating the manipulation of task complexity and *complexity* in indicating perceived task complexity. Scores of each item range from 1 to 9. Generally, the descriptive results showed complex patterns of perceived complexity and difficulty across different conditions, while task anxiety seemed to have little variation across the conditions. Additionally, the levels of interest and motivation for the writing tasks were apparently distinct between the student and teacher groups. To examine the effect of genre and idea support on perceptions statistically, I computed mixed ANOVAs with the Bonferroni adjustment ( $\alpha = .05/6$  or  $.0083$ ). Throughout the section, I first report the results of interaction effects and their post-hoc results, followed by those of main effects.

Table 7 shows the interaction effects of the three independent variables (genre, idea support, and group) on task perceptions. Complexity was the only item that showed a significant three-way interaction ( $F(1, 104) = 9.25, p = .003, \eta_p^2 = .082$ ). That is, the students and teachers had different perceptions about how genre and idea support manipulations influence task complexity. Specifically, the teachers predicted that providing supporting ideas would make the argumentative task less complex, but the same task manipulation would make the narrative task even more complex; in contrast, the students reported that they found the provision of supporting

ideas lowering the complexity of both genres, leading to a significant three-way interaction (post-hoc analysis results reported in the next paragraph). In addition, the results showed a significant interaction between genre and group on perceived task complexity ( $F(1, 104) = 9.06, p = .003, \eta_p^2 = .080$ ) and difficulty ( $F(1, 104) = 10.87, p = .001, \eta_p^2 = .095$ ). That is, the teachers perceived the argumentative genre more complex than the narrative, while the students found both genres similarly complex and difficult (see Figure 1). All significant interaction effects on perceived complexity and difficulty were medium in size, with  $\eta_p^2$  ranging from .08 to .11. Other categories (task anxiety, confidence, interest, and motivation) did not show any significant interactions.

For complexity and difficulty, which showed significant interactions, I performed post-hoc analyses separately for each group so that the effect of genre and idea support can be more clearly presented. As shown in Table 8, the manipulation of idea support actually led to significant changes in the students' perceptions of task complexity and difficulty, with medium effect sizes (complexity:  $F(1, 75) = 6.91, p = .010, \eta_p^2 = .084$ ; difficulty:  $F(1, 75) = 9.97, p = .002, \eta_p^2 = .117$ ). On the other hand, the students did not perceive that different genres impose significantly different levels of complexity and difficulty (complexity:  $F(1, 75) = 0.45, p = .51, \eta_p^2 = .006$ ; difficulty:  $F(1, 75) = 0.81, p = .37, \eta_p^2 = .011$ ). These perception-based findings potentially give support to the use of the idea provision condition as a cognitive complexity variable in written discourse and, more interestingly, refute the general assumption that narrative writing would be cognitively less demanding and less difficult to ESL students than argumentative writing.

Table 6.

*Descriptive Statistics for ESL Students' and Teachers' Perceptions of Writing Tasks*

Item	Group	Arg/-Support		Arg/+Support		Nar/-Support		Nar/+Support	
		<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI
Complexity	Student	5.55 (1.77)	[5.15, 5.96]	5.07 (1.73)	[4.67, 5.46]	5.59 (1.67)	[5.21, 5.97]	5.25 (1.65)	[4.87, 5.63]
	Teacher	6.23 (1.17)	[5.80, 6.67]	5.23 (2.24)	[4.40, 6.07]	4.47 (1.43)	[3.93, 5.00]	5.30 (1.82)	[4.62, 5.98]
Difficulty	Student	5.14 (1.76)	[4.74, 5.55]	4.50 (1.66)	[4.12, 4.88]	5.17 (1.84)	[4.75, 5.59]	4.82 (1.96)	[4.24, 5.56]
	Teacher	5.63 (1.38)	[5.12, 6.15]	5.27 (2.18)	[4.45, 6.08]	4.00 (1.44)	[3.46, 4.54]	4.90 (1.77)	[4.37, 5.26]
Anxiety	Student	4.95 (2.00)	[4.49, 5.40]	4.64 (2.04)	[4.18, 5.11]	4.97 (2.14)	[4.49, 5.46]	4.62 (1.83)	[4.20, 5.04]
	Teacher	5.37 (1.79)	[4.70, 6.04]	4.67 (1.81)	[3.99, 5.34]	4.10 (1.40)	[3.58, 4.62]	4.87 (1.78)	[4.20, 5.53]
Confidence	Student	4.84 (1.86)	[4.42, 5.27]	5.20 (1.74)	[4.80, 5.59]	5.39 (1.76)	[4.99, 5.80]	5.03 (1.80)	[4.61, 5.44]
	Teacher	5.70 (1.68)	[5.07, 6.33]	5.90 (1.79)	[5.23, 6.57]	6.57 (1.33)	[6.07, 7.06]	5.90 (1.63)	[5.29, 6.51]
Interest	Student	4.59 (1.90)	[4.16, 5.03]	5.20 (1.74)	[4.80, 5.59]	5.09 (1.67)	[4.71, 5.47]	5.53 (1.94)	[5.08, 5.97]
	Teacher	5.73 (1.72)	[5.09, 6.38]	5.87 (1.63)	[5.26, 6.48]	6.40 (1.77)	[5.74, 7.06]	6.27 (1.66)	[5.65, 6.89]
Motivation	Student	4.86 (2.00)	[4.40, 5.31]	5.18 (2.09)	[4.71, 5.66]	5.37 (1.87)	[4.94, 5.80]	5.47 (1.83)	[5.06, 5.89]
	Teacher	5.03 (1.94)	[4.31, 5.76]	5.33 (1.71)	[4.70, 5.97]	6.13 (1.85)	[5.44, 6.82]	5.87 (1.48)	[5.31, 6.42]

Table 7.

*Interaction Effects of Genre, Idea Support, and Group on Task Perceptions*

Item	Genre × Idea support × Group			Genre × Group			Idea support × Group			Genre × Idea support		
	<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power
Complexity	.003*	.082	.854	.003*	.080	.847	.284	.011	.187	.001*	.109	.942
Difficulty	.097	.026	.381	.001*	.095	.904	.015	.055	.687	.009	.064	.751
Anxiety	.032	.043	.576	.141	.021	.312	.315	.010	.170	.046	.038	.517
Confidence	.823	.001	.056	.520	.004	.098	.435	.006	.121	.014	.056	.696
Interest	.892	.001	.052	.727	.001	.064	.132	.022	.324	.534	.004	.095
Motivation	.639	.002	.075	.187	.017	.260	.530	.004	.096	.281	.011	.189

Note. \**p* values are significant with the Bonferroni correction ( $\alpha = .05/6$  or .0083).

Table 8.

*Post-hoc Analysis Results of Genre and Idea Support Effects for Each Group's Perceptions*

Item	Group	Genre			Idea support			Genre × Idea support		
		<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power
Complexity	Student	.506	.006	.101	.009*	.084	.737	.653	.003	.073
	Teacher	.005*	.240	.833	.771	.003	.059	< .001*	.492	.999
Difficulty	Student	.370	.011	.145	.002*	.117	.876	.379	.010	.141
	Teacher	.002*	.279	.899	.363	.029	.145	.005*	.238	.829

Note. \**p* values are significant with the Bonferroni correction ( $p < .05/2$  or .025).

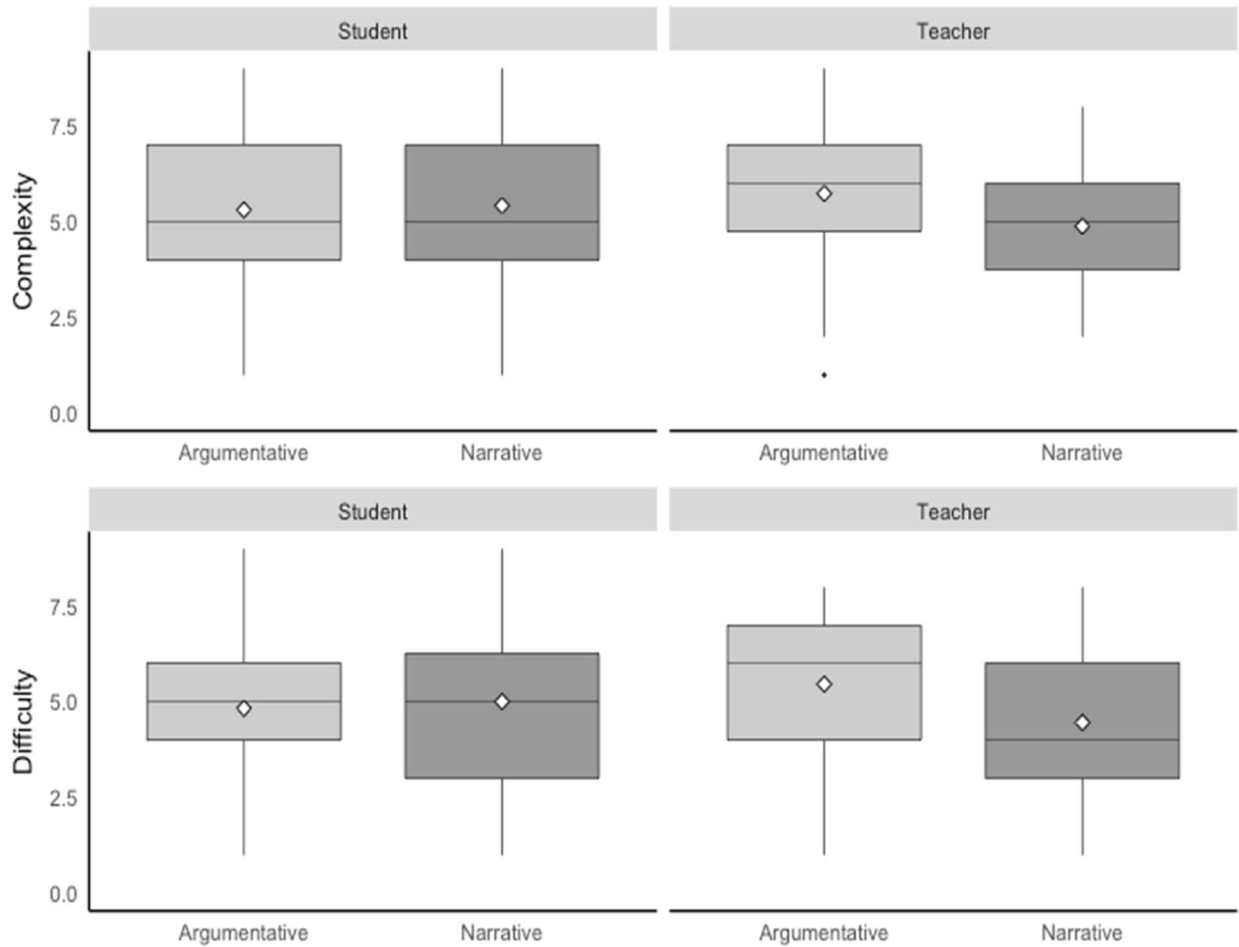


Figure 1. Students' and teachers' perceptions of task complexity and difficulty across genre conditions.

Results from the teachers showed an entirely different pattern. There were significant interactions between genre and idea support on their perceived complexity and difficulty, both with large effect sizes (complexity:  $F(1, 29) = 28.07, p < .001, \eta_p^2 = .49$ ; difficulty:  $F(1, 29) = 9.07, p = .005, \eta_p^2 = .24$ ). Specifically, the teachers predicted that the provision of supporting ideas would mitigate the complexity and difficulty of argument writing, but a similar manipulation on the narrative genre would increase the level of task complexity and difficulty (see Figure 2). Furthermore, contrary to the results from the students, the teachers expected that ESL students would have different levels of task complexity and difficulty across two genres

(i.e., argumentative tasks imposing greater complexity and difficulty on ESL students than narratives; complexity:  $F(1, 29) = 9.17, p = .005, \eta_p^2 = .24$ ; difficulty:  $F(1, 29) = 11.23, p = .002, \eta_p^2 = .28$ ). The wide gap between the students and teachers in their perceptions of task manipulation effects will be discussed in more detail in the next chapter.

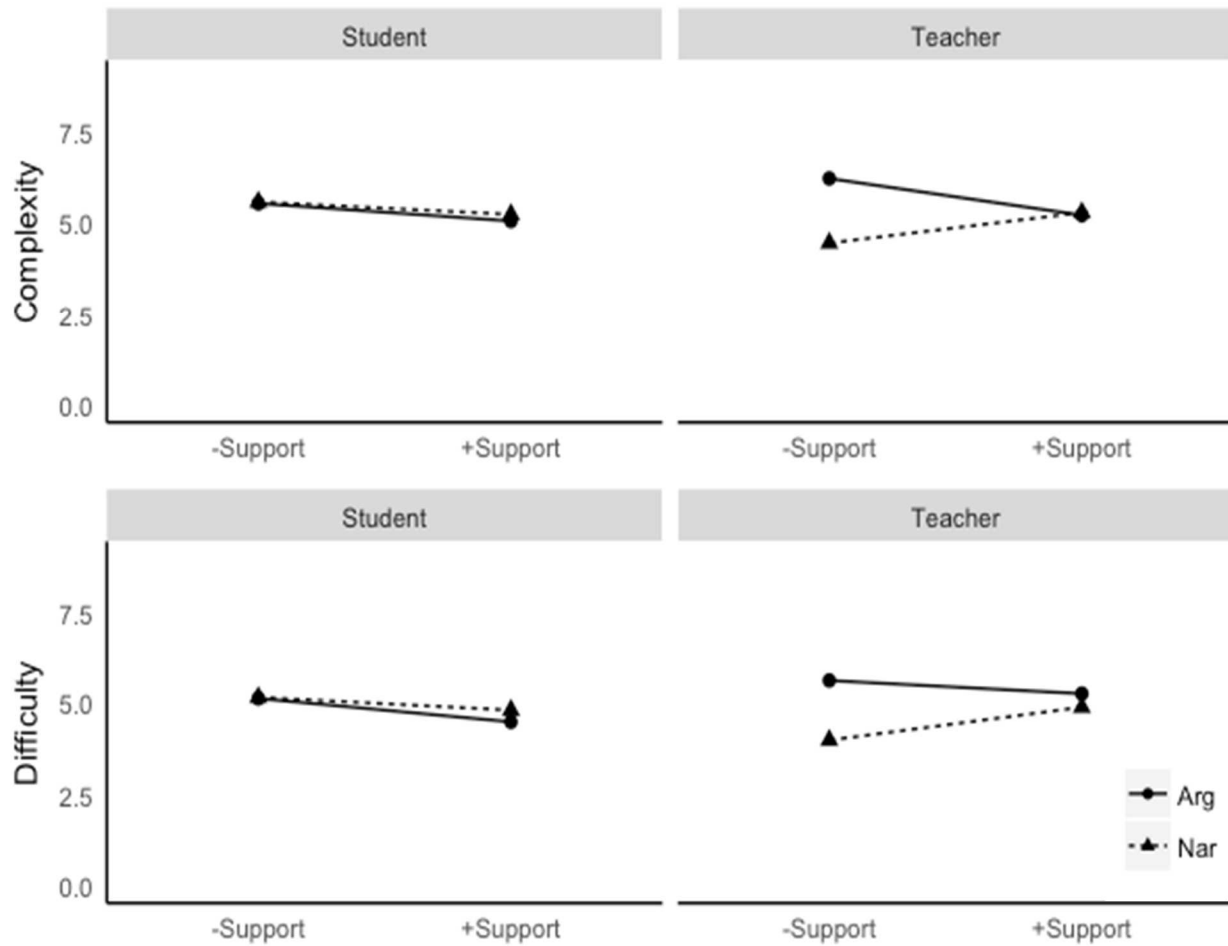


Figure 2. Interaction plots for perceived complexity and difficulty showing an interaction between genre and idea support only for teacher perceptions.

In terms of the main effect of each variable (see Table 9), the results showed that there were significant main effects for group on task confidence and interest (confidence:  $F(1, 104) = 13.63, p < .001, \eta_p^2 = .12$ ; interest:  $F(1, 104) = 15.76, p < .001, \eta_p^2 = .13$ ). Specifically, as Figure 3 shows, the teachers' expectations of ESL students' confidence and interest in the given tasks

were shown to be consistently higher than the actual confidence and interest levels expressed by the students. Moreover, the results showed significant main effects for genre on task interest and motivation (interest:  $F(1, 104) = 7.80, p = .006, \eta_p^2 = .07$ ; motivation:  $F(1, 104) = 15.17, p < .001, \eta_p^2 = .13$ ). In other words, both students and teachers viewed narrative writing more interesting than argumentative, making the students feel more strongly motivated to do the narrative tasks, compared to the argumentative tasks. Task anxiety was the category with no significant interaction or main effects.

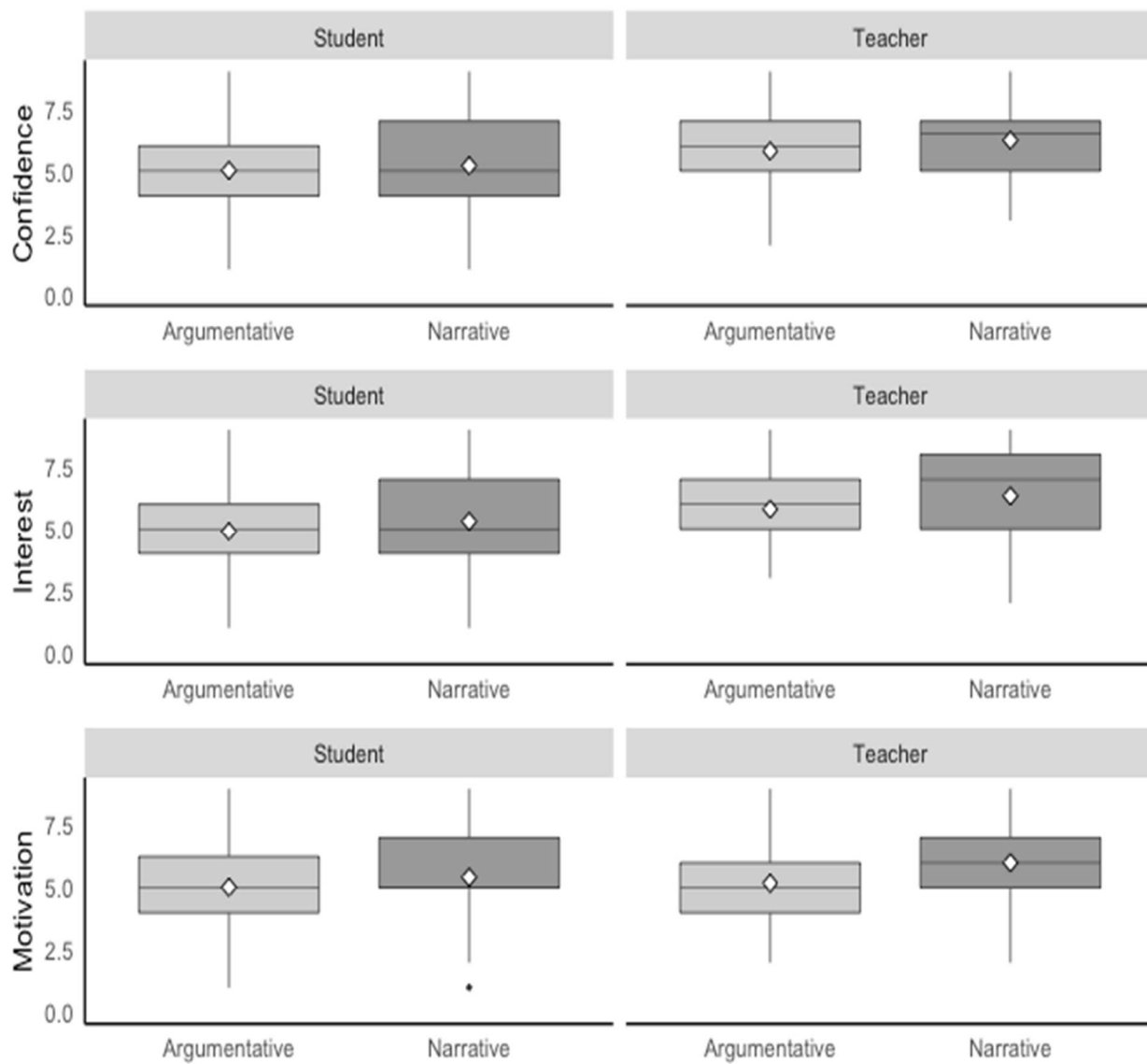


Figure 3. Students' and teachers' perceptions of task confidence, interest, and motivation across genre conditions.



Table 9.

*Main Effects of Genre, Idea Support, and Group on Task Perceptions*

Item	Genre			Idea support			Group		
	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power
Complexity	.023	.049	.629	.109	.025	.361	.827	.001	.055
Difficulty	.022	.050	.638	.454	.005	.116	.875	.001	.053
Anxiety	.141	.021	.312	.412	.006	.129	.869	.001	.053
Confidence	.099	.026	.377	.409	.007	.130	< .001*	.116	.955
Interest	.006*	.070	.790	.132	.022	.324	< .001*	.132	.976
Motivation	< .001*	.127	.971	.464	.005	.113	.204	.016	.245

*Note.* \* $p$  values are significant with the Bonferroni correction ( $p < .05/6$  or .0083).

As a next step, I computed Pearson correlations to explore the relationship among the perception statements answered by the students. In Table 10, we can see that task complexity and task difficulty are positively related ( $.34 < rs < .62$ ), and the level of stress caused by each writing task had positive relationships with both task complexity ( $.30 < rs < .59$ ) and difficulty ( $.54 < rs < .64$ ). Additionally, there were positive relationships among task confidence, interest, and motivation ( $.22 < rs < .69$ ). In particular, the correlations between task interest and motivation were fairly strong ( $.53 < rs < .69$ ). These findings generally conform to our understanding of how various dimensions of task perceptions work.

Table 10.

*Correlations between Perception Items by Task Type*

Students ( $N = 76$ )	Difficulty	Anxiety	Confidence	Interest	Motivation
Arg/-Support					
Complexity	.497*	.389*	-.168	.207	.113
Difficulty	-	.594*	-.206	.182	.055
Anxiety		-	-.204	.026	-.092
Confidence			-	.394*	.440*
Interest				-	.687*
Arg/+Support					
Complexity	.582*	.587*	-.089	.213	.222
Difficulty	-	.638*	-.405*	-.062	-.061
Anxiety		-	-.412*	-.070	.001
Confidence			-	.217	.225
Interest				-	.567*
Nar/-Support					
Complexity	.340*	.299*	.141	.052	.027
Difficulty	-	.634*	-.236	.082	-.256
Anxiety		-	-.278	-.245	-.425*
Confidence			-	.327*	.369*
Interest				-	.530*
Nar/+Support					
Complexity	.619*	.543*	-.141	.029	.044
Difficulty	-	.536*	-.300*	-.135	-.176
Anxiety		-	-.251	-.138	-.116
Confidence			-	.339*	.364*
Interest				-	.683*

*Note.* \*correlations are significant at the alpha level of .01.

Nevertheless, there were some unexpected patterns revealed from this analysis. The increase in task complexity did not necessarily result in lower task interest or motivation, as evidenced by non-significant, but mostly positive, correlations of these dimensions with task complexity. Similarly, although task difficulty showed negative relationships with task confidence, task complexity did not necessarily correlate negatively with task confidence, suggesting that task complexity and difficulty tap different constructs and that a reasonable level of increase in task complexity does not harm learners' affective states. What we can infer from the correlation results is that an increased level of task complexity, when suitable for learners' developmental stage, can allow the learners to become more interested in a task.

To complement the result of task questionnaires, I examined the teachers' response to open-ended questions and found that, contrary to a general trend elicited from statistical analyses of the questionnaire data, two teachers were actually aware of the potential outcome of ESL students' learning experience on their knowledge and performance, such as construction of unbalanced genre schemas and greater difficulty with narrative writing:

*In terms of task difficulty, I don't think the two sets of prompts would be much different (although A2 [Arg/+Support] seems to be much easier than A1 [Arg/-Support] since it gives answers). Although narratives are often considered easier than argumentation, many ESL students often have a lot of experience of writing argumentative essays like A1 and A2 for their test prep. Depending on the extent to which they have been exposed to each genre of writing, for some students A1 and A2 can be easier than N1 and N2 [Nar/-Support and Nar/+Support]. (Participant ID: T115)*

*For the N1 and N2 tasks [Nar/-Support and Nar/+Support], I'd say that some students will be intimidated by the genre of the task, depending on whether or not they've had experience with this kind of writing. (Participant ID: T120)*

Some teachers also noted that the provision of idea support should be performed with caution. Two relevant excerpts are as follows:

*I thought the writing prompts were relevant to students' lives for the most part, but although the suggestions in N2 [Nar/+Support] and A2 [Arg/+Support] could help students by providing a starting point and/or some specific examples to draw on, they could also be frustrating if students hadn't encountered those specific situations.*

(Participant ID: T103)

*Both prompts [Nar/+Support and Arg/+Support] provided too much scaffolding, making it both too easy to do well at the task and too hard, since sometimes it is easier to come up with support for your own ideas than for another's. (Participant ID: T128)*

The excerpts above indicate ESL teachers' concerns about the adverse effect of supporting ideas, such as the possibility of their restriction on what students can think and write, particularly when the supporting ideas included in a prompt do not reflect students' life experience.

Additionally, teachers cautioned that providing too specific outlines would deprive students of the opportunity to generate their own unique ideas in writing, which is an important part of the writing skills generally targeted in L2 learning contexts. The excerpts below are some examples of such concerns:

*In general, I do not like prompts that lead the students too much (like N2 and A2) [Nar/+Support and Arg/+Support]. These prompts provide a road map for the supporting*

*points that makes a whole classroom of essays very repetitive to read. Is that intentional?*

(Participant ID: T107)

*Tasks N1 and N2 [two narrative prompts] were relatively simple and because they are narrative and personal, ESL students will perform well, I believe. Tasks A1 and A2 [two argumentative prompts] were more academic, but A2 [Arg/+Support] provided a brief outline of arguments which would be helpful to the test-takers. As an instructor, though, I think it would be better NOT to provide the outlines in A2 because part of the goal of the task is to see how well writers can generate and organize their own ideas. (Participant ID: T119)*

Last, as shown in the excerpts below, some teachers expressed the possibility of greater motivation for narrative writing, potentially arising from its less formulaic and more personalized characteristics.

*The first one [Nar/-Support] was open enough that better writers may be able to do something interesting and step outside of formulaic 5-paragraph essay writing.*

(Participant ID: T127)

*Motivation goes up when the student is in a “can-do” situation and is encouraged to communicate a message that they are personally invested in. (Participant ID: T111)*

The examination of these quotes selected from the teachers’ open-ended response enabled us to have a more in-depth understanding of the teachers’ perceptions of the tasks. All of the quantitative findings related to task perceptions are summarized in Table 11, which will be discussed together with text feature results in the next chapter.

Table 11.

*Summary of Task Perception Results*

	Students	Teachers	Correlation patterns
Complexity	<ul style="list-style-type: none"> <li>• Similar level of complexity for narrative and argument</li> <li>• Lower complexity for the tasks with idea support</li> </ul>	<ul style="list-style-type: none"> <li>• Higher complexity for argument than narrative</li> <li>• Idea support leads to higher complexity for narrative, but lower complexity for argument</li> </ul>	<ul style="list-style-type: none"> <li>• Positive correlations with task difficulty and stress</li> </ul>
Difficulty	<ul style="list-style-type: none"> <li>• Similar level of difficulty for narrative and argument</li> <li>• Lower difficulty for the tasks with idea support</li> </ul>	<ul style="list-style-type: none"> <li>• Higher difficulty for argument than narrative</li> <li>• Idea support leads to higher difficulty for narrative, but lower difficulty for argument</li> </ul>	<ul style="list-style-type: none"> <li>• Positive correlations with task complexity</li> <li>• Negative correlations with task confidence</li> </ul>
Anxiety	<ul style="list-style-type: none"> <li>• Similar level of anxiety for all tasks</li> </ul>	<ul style="list-style-type: none"> <li>• Similar level of anxiety for all tasks</li> </ul>	<ul style="list-style-type: none"> <li>• Positive correlations with task complexity and difficulty</li> <li>• Negative correlations with task confidence</li> </ul>
Confidence	<ul style="list-style-type: none"> <li>• Similar level of confidence for all tasks</li> </ul>	<ul style="list-style-type: none"> <li>• Similar level of confidence for all tasks</li> <li>• Higher task confidence from teachers than students</li> </ul>	<ul style="list-style-type: none"> <li>• Positive correlations with task interest and motivation</li> <li>• Negative correlations with task difficulty and stress</li> </ul>
Interest	<ul style="list-style-type: none"> <li>• Higher interest in narrative than argument</li> </ul>	<ul style="list-style-type: none"> <li>• Higher interest in narrative than argument</li> <li>• Higher interest from teachers than students</li> </ul>	<ul style="list-style-type: none"> <li>• Positive correlations with task confidence and motivation</li> </ul>
Motivation	<ul style="list-style-type: none"> <li>• Higher motivation for narrative than argument</li> </ul>	<ul style="list-style-type: none"> <li>• Higher motivation for narrative than argument</li> </ul>	<ul style="list-style-type: none"> <li>• Positive correlations with task confidence and interest</li> </ul>

## Textual Feature Changes across Task Types

The second research question addresses the effect of genre and idea support on various text features in ESL student writing (i.e., syntactic, lexical, discourse, and metadiscourse features), which is an attempt to (1) reveal how task manipulations lead students to use different features of language, (2) associate such linguistic changes with communicative functions of each genre, and (3) ultimately suggest a comprehensive picture of how genre-specific functions and learners' task perceptions influence their language use together and/or separately. To attain these aims, I examined 21 text features with respect to their changes across task types (see Table 12 for descriptive results). To illustrate the target measures briefly, there were ten measures tapping the construct of syntactic complexity:

- Unit-length measures: mean length of sentence (MLS) and mean length of clause (MLC)
- Subordination measures: clauses per T-unit (C/T), nominal clauses per 1,000 words (NOMC), adverbial clauses per 1,000 words (ADVC), and adjective clauses per 1,000 words (ADJC)
- Phrasal-level measures: coordinate phrases per clause (CP/C), complex nominals per clause (CN/C), average number of words before the main verb (left embedded), and average number of modifiers per noun phrase (modifiers/NP).

Two lexical measures were additionally targeted:

- Lexical diversity based on the vocd-D formula (D) and lexical sophistication based on average word frequency extracted from the CELEX corpus (WF; here, lower WF indicates greater lexical sophistication)

I examined five discourse measures:

- Two lexical cohesion measures: argument overlap between adjacent sentences (coreference cohesion) and semantic overlap between adjacent sentences (conceptual cohesion)
- Two connective density measures: causal connectives per 1,000 words (causal connective density) and temporal connectives per 1,000 words (temporal connective density)
- Nominalizations per 1,000 words (nominalization density)

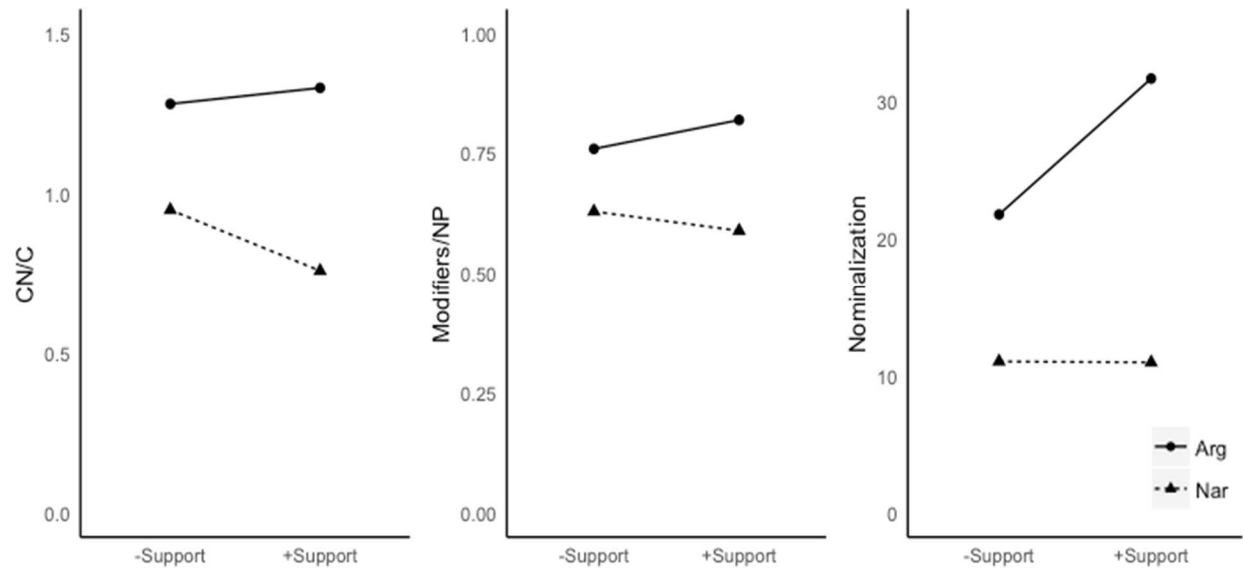
Finally, I included four metadiscourse measures:

- Number of hedges per 1,000 words (hedge density), number of boosters per 1,000 words (booster density), number of self-mentions per 1,000 words (self-mention density), and number of reader pronouns per 1,000 words (reader pronoun density)

Table 13 presents the results of two-way ANOVAs regarding how genre and idea support manipulations elicited different textual features. First, the result showed significant interaction between genre and idea support on various aspects of noun phrase complexification (CN/C:  $F(1, 75) = 22.02, p < .001, \eta_p^2 = .23$ ; Modifiers/NP:  $F(1, 75) = 21.14, p < .001, \eta_p^2 = .22$ ; Nominalization:  $F(1, 75) = 15.16, p < .001, \eta_p^2 = .17$ ). As presented in Figure 4, the result of post-hoc analyses (paired samples *t*-tests) suggested that the provision of idea support in argumentative writing had a tendency to lead to a significant increase (or increasing pattern with no statistical significance) in noun phrase complexity (CN/C:  $t(75) = -1.46, p = .150, d = -0.14$ ; Modifiers/NP:  $t(75) = -3.14, p = .002, d = -0.31$ ; Nominalization:  $t(75) = -4.43, p < .001, d = -0.96$ ), whereas the same manipulation in narrative writing likely resulted in a significant decrease (or a decreasing pattern with no statistical significance) in nominal complexity (CN/C:  $t(75) =$



5.22,  $p < .001$ ,  $d = 0.66$ ; Modifiers/NP:  $t(75) = 2.95$ ,  $p = .004$ ,  $d = 0.33$ ; Nominalization:  $t(75) = 0.05$ ,  $p = .96$ ,  $d = 0.01$ ).



*Figure 4.* Interaction plots for complex nominals per clause, modifiers per noun phrase, and nominalization density showing an interaction between genre and idea support conditions. It should be noted that the y-axes of the plots have different scales.

Table 12.

*Descriptive Statistics for Target Text Features by Task Type*

Measure	Arg/-Support		Arg/+Support		Nar/-Support		Nar/+Support	
	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI
Length of unit								
MLS	17.17 (4.03)	[16.24, 18.09]	18.16 (4.52)	[17.13, 19.19]	16.20 (4.09)	[15.27, 17.14]	16.00 (4.54)	[14.96, 17.04]
MLC	9.79 (1.67)	[9.41, 10.17]	9.74 (1.38)	[9.42, 10.06]	8.76 (1.57)	[8.40, 9.12]	8.20 (1.34)	[7.90, 8.51]
Subordination								
C/T	1.58 (0.28)	[1.51, 1.64]	1.67 (0.34)	[1.59, 1.74]	1.59 (0.30)	[1.52, 1.65]	1.67 (0.34)	[1.59, 1.75]
NOMC	8.55 (6.52)	[7.06, 10.04]	7.06 (5.92)	[5.71, 8.42]	10.27 (6.11)	[8.88, 11.67]	9.60 (5.90)	[8.25, 10.95]
ADVC	16.68 (8.99)	[14.62, 18.73]	16.36 (7.31)	[14.69, 18.04]	13.79 (6.60)	[12.28, 15.30]	12.65 (6.77)	[11.10, 14.20]
ADJC	10.32 (7.08)	[8.70, 11.93]	12.02 (6.52)	[10.53, 13.51]	8.31 (6.02)	[6.94, 9.69]	8.64 (6.21)	[7.23, 10.06]
Phrasal complexity								
CP/C	0.20 (0.14)	[0.17, 0.23]	0.19 (0.11)	[0.17, 0.22]	0.17 (0.12)	[0.14, 0.20]	0.15 (0.09)	[0.13, 0.17]
CN/C	1.28 (0.34)	[1.20, 1.35]	1.33 (0.33)	[1.26, 1.41]	0.95 (0.32)	[0.88, 1.02]	0.76 (0.20)	[0.72, 0.81]
Left embedded	5.05 (1.82)	[4.63, 5.47]	4.94 (1.77)	[4.53, 5.35]	4.15 (1.12)	[3.90, 4.41]	3.99 (1.27)	[3.70, 4.28]
Modifiers/NP	0.76 (0.16)	[0.72, 0.79]	0.82 (0.13)	[0.79, 0.85]	0.63 (0.12)	[0.60, 0.66]	0.59 (0.11)	[0.56, 0.61]
Lexical features								
D	75.36 (13.84)	[72.20, 78.52]	76.57 (16.35)	[72.83, 80.30]	78.14 (17.07)	[74.24, 82.04]	82.57 (15.11)	[79.12, 86.02]
WF	3.04 (0.09)	[3.02, 3.06]	3.08 (0.08)	[3.06, 3.09]	3.09 (0.07)	[3.08, 3.11]	3.11 (0.08)	[3.09, 3.12]
Discourse								
Argument overlap	0.64 (0.20)	[0.60, 0.69]	0.64 (0.17)	[0.60, 0.68]	0.67 (0.16)	[0.63, 0.71]	0.63 (0.16)	[0.60, 0.67]
Semantic overlap	0.24 (0.08)	[0.22, 0.26]	0.24 (0.07)	[0.23, 0.26]	0.21 (0.07)	[0.19, 0.23]	0.18 (0.05)	[0.17, 0.19]
Causal connective	37.76 (12.67)	[34.86, 40.65]	33.37 (10.86)	[30.89, 35.85]	34.16 (11.78)	[31.47, 36.85]	34.64 (11.95)	[31.91, 37.37]
Temporal connective	14.84 (9.22)	[12.73, 16.95]	15.08 (7.67)	[13.33, 16.83]	20.31 (9.07)	[18.25, 22.37]	26.12 (10.13)	[23.80, 28.43]
Nominalization	21.83 (12.70)	[18.93, 24.73]	31.71 (15.27)	[28.22, 35.20]	11.06 (8.13)	[9.20, 12.91]	10.99 (11.27)	[8.41, 13.56]
Metadiscourse								
Hedge	11.86 (7.42)	[10.16, 13.55]	17.57 (11.15)	[15.02, 20.11]	15.75 (8.90)	[13.72, 17.79]	17.92 (10.13)	[15.60, 20.23]
Booster	23.35 (13.69)	[20.23, 26.48]	21.88 (10.95)	[20.37, 24.38]	24.73 (9.62)	[22.53, 26.93]	24.86 (9.90)	[22.60, 27.12]
Self-mention	19.24 (19.12)	[14.88, 23.61]	18.03 (17.12)	[14.11, 21.94]	71.90 (29.61)	[65.13, 78.66]	68.93 (24.97)	[63.22, 74.63]
Reader pronoun	28.18 (20.70)	[23.45, 32.91]	29.58 (24.03)	[24.09, 25.07]	32.72 (22.08)	[27.78, 37.77]	34.82 (22.29)	[29.73, 39.92]

Table 13.

*Inferential Statistics for Genre and Idea Support Effects on Textual Features*

Measure	Genre			Idea support			Genre $\times$ Idea support		
	<i>P</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power
Length of unit									
MLS	< .001*	.330	1.000	.260	.017	.202	.016	.075	.684
MLC	< .001*	.515	1.000	.038	.056	.548	.048	.051	.509
Subordination									
C/T	.785	.001	.058	.003	.109	.847	.933	.001	.051
NOMC	.002*	.120	.883	.134	.030	.321	.483	.007	.107
ADVC	< .001*	.161	.963	.431	.008	.123	.581	.004	.085
ADJC	< .001*	.215	.994	.107	.034	.364	.328	.013	.163
Phrasal complexity									
CP/C	.008	.090	.766	.550	.005	.091	.442	.008	.119
CN/C	< .001*	.767	1.000	.020	.070	.652	< .001*	.227	.996
Left embedded	< .001*	.369	1.000	.313	.014	.171	.714	.002	.065
Modifiers/NP	< .001*	.759	1.000	.368	.011	.146	< .001*	.220	.995
Lexical features									
D	.009	.088	.758	.040	.055	.541	.240	.018	.216
WF	< .001*	.302	1.000	< .001*	.162	.964	.063	.045	.462
Discourse									
Argument overlap	.489	.006	.106	.333	.012	.161	.396	.010	.134
Semantic overlap	< .001*	.407	1.000	.090	.038	.397	.006	.097	.801
Causal connective	.334	.012	.160	.104	.035	.369	.089	.038	.398
Temporal connective	< .001*	.447	1.000	.001*	.144	.938	.003	.114	.867
Nominalization	< .001*	.639	1.000	.001*	.139	.930	< .001*	.168	.970
Metadiscourse									
Hedge	.064	.045	.459	< .001*	.202	.991	.056	.048	.483
Booster	.090	.038	.396	.576	.004	.086	.447	.008	.117
Self-mention	< .001*	.824	1.000	.571	.004	.087	.822	.001	.056
Reader pronoun	.037	.056	.552	.445	.008	.118	.857	.001	.054

Note. \**p* values are significant with the Bonferroni correction ( $\alpha = .05/21$  or .0024).

Main effects of genre were prevalent for many of the textual measures with medium to large effect sizes ( $\eta_p^2$  from .12 to .82), while those of idea support existed only for a few measures. With regard to genre effects, the argumentative essays elicited significantly higher values of unit length (MLS:  $F(1, 75) = 36.92, p < .001, \eta_p^2 = .33$ ; MLC:  $F(1, 75) = 79.71, p < .001, \eta_p^2 = .52$ ), phrasal complexity (CN/C:  $F(1, 75) = 247.46, p < .001, \eta_p^2 = .77$ ; left embedded:  $F(1, 75) = 43.88, p < .001, \eta_p^2 = .37$ ; modifiers/NP:  $F(1, 75) = 236.71, p < .001, \eta_p^2 = .76$ ), and discourse measures (semantic overlap:  $F(1, 75) = 51.55, p < .001, \eta_p^2 = .41$ ; nominalization:  $F(1, 75) = 132.62, p < .001, \eta_p^2 = .64$ ) than the narrative essays. Of these measures with significant changes, the density of complex nominals (CN/C) was found to have the largest effect size ( $\eta_p^2 = .77$ ). The significant main effects of CN/C and modifiers/NP are illustrated in Figure 5.

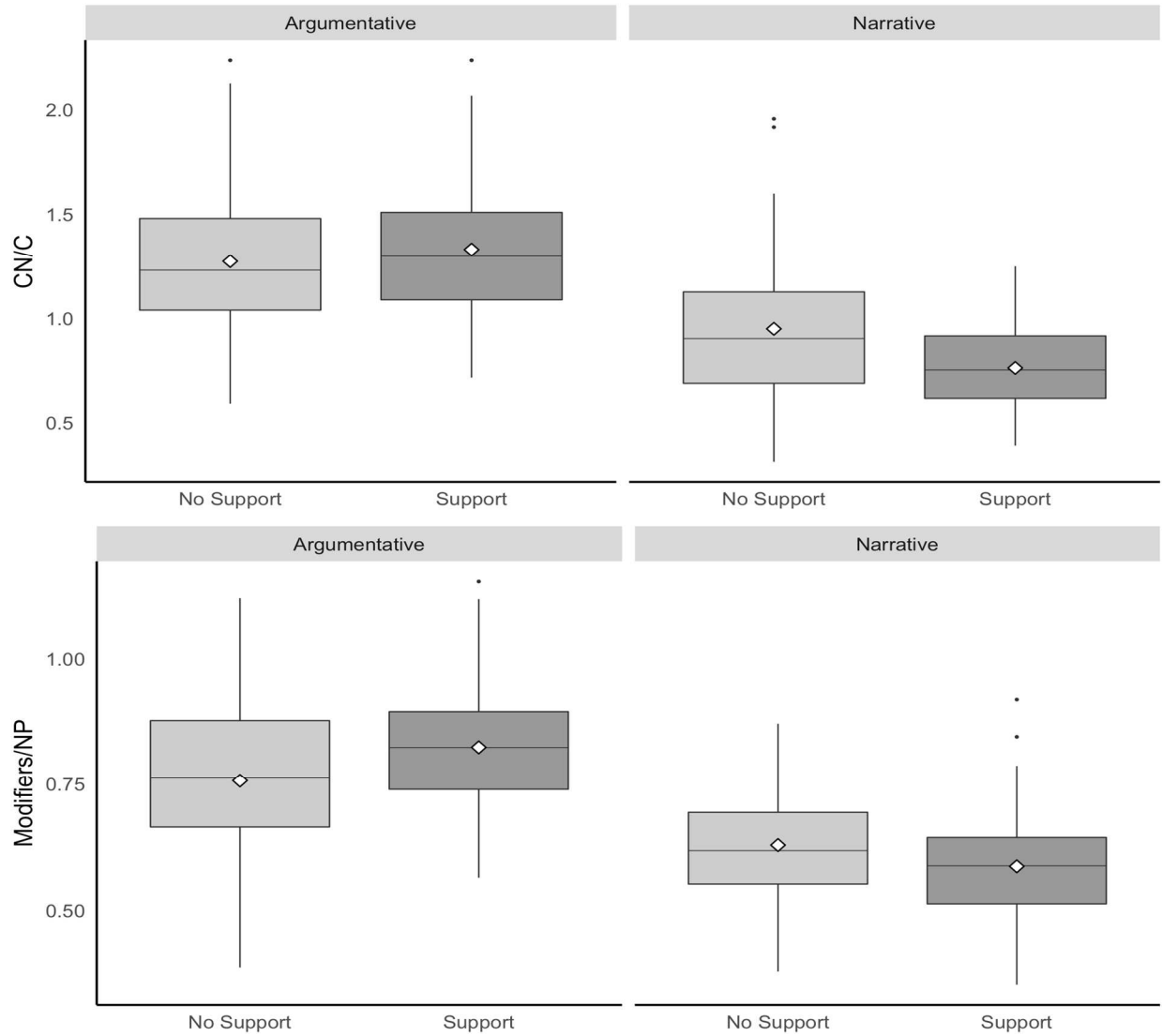


Figure 5. Complex nominals per clause and modifiers per noun phrase across genre and idea support conditions.

On the other hand, as displayed in Figure 6, the narratives showed significantly higher values in temporal connective density ( $F(1, 75) = 60.54, p < .001, \eta_p^2 = .45$ ) and self-mention density ( $F(1, 75) = 350.46, p < .001, \eta_p^2 = .82$ ) than the argumentative essays. This result of increased temporal connectives and self-mentions in narrative writing is not very surprising because they are important linguistic resources that writers use in narrating a personal story (Biber & Conrad, 2009).

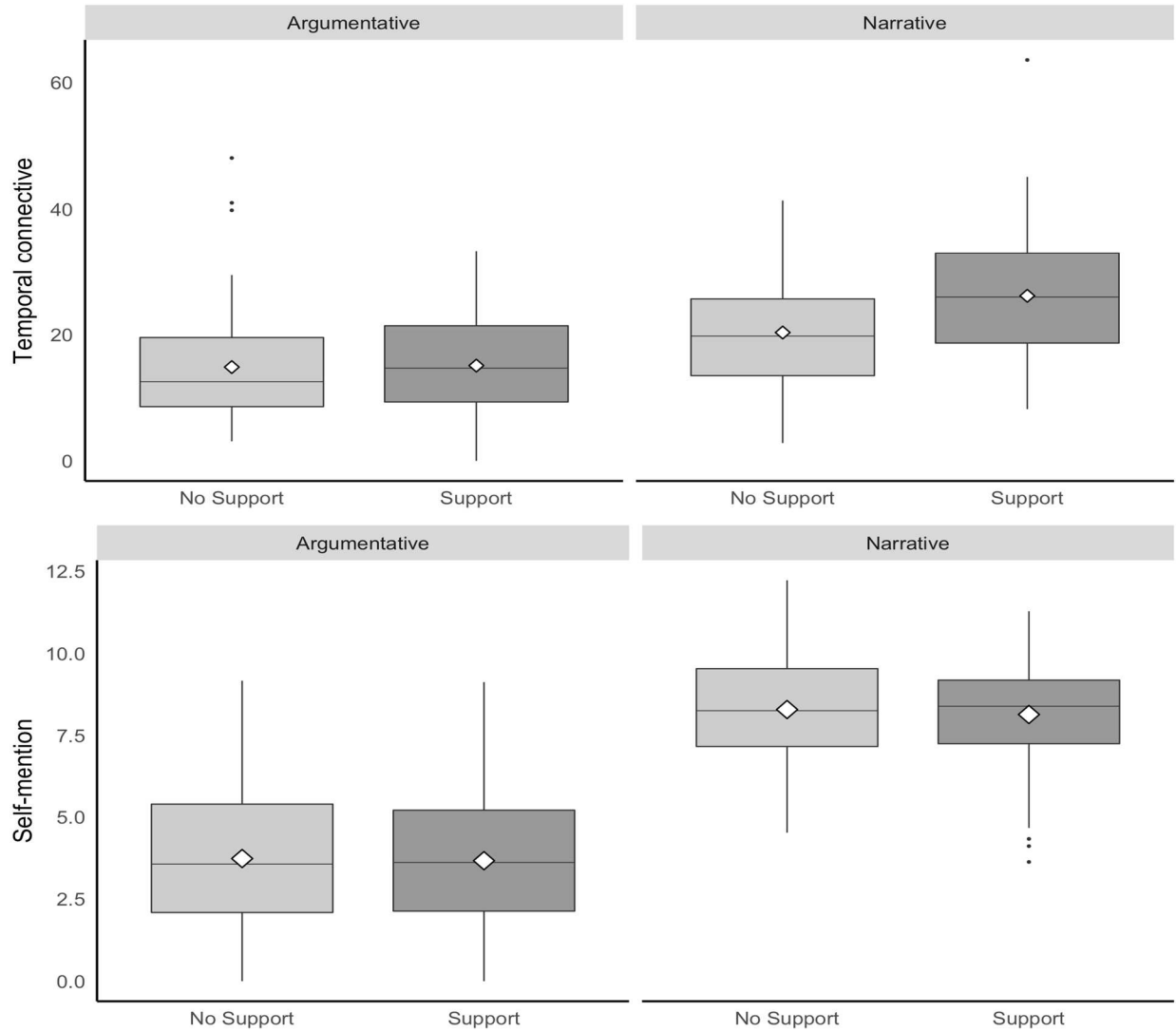


Figure 6. Temporal connective density and self-mention density across genre and idea support conditions.

The clauses per T-unit (C/T) measure, which had been extensively adopted as a typical measure of clausal subordination, was not shown to change across the two genres ( $F(1, 75) = 0.08, p = .79, \eta_p^2 = .001$ ), and this result is in line with the findings of previous research (e.g., Lu, 2011; Yoon & Polio, 2017). However, using more fine-grained measures of clausal subordination (i.e., nominal, adverbial, and adjectival clause density), I found that narrative writing is characterized by increased nominal clause density ( $F(1, 75) = 10.18, p = .002, \eta_p^2 = .12$ ) and argumentative writing by increased density of adverbial clauses ( $F(1, 75) = 14.38, p < .001, \eta_p^2$

= .16) and adjectival clauses ( $F(1, 75) = 20.49, p < .001, \eta_p^2 = .22$ ). This result is notable in that, unlike previous studies attending to phrasal measures and, accordingly, rejecting clausal subordination in relation to genre variation (except for Frear & Bitchener, 2015), the result clearly indicates that the use of more specific measures allows us to detect how different genres elicit different characteristics of clausal subordination in L2 writing (see Figure 7), which has been gone unnoticed in most previous research due to its reliance on a general subordination measure (Lambert & Kormos, 2014; Wolfe-Quintero et al., 1998).

There were some text features that varied significantly with the provision of idea support (WF:  $F(1, 75) = 14.53, p < .001, \eta_p^2 = .16$ ; nominalization:  $F(1, 75) = 12.12, p = .001, \eta_p^2 = .14$ ; temporal connective density:  $F(1, 75) = 12.57, p = .001, \eta_p^2 = .14$ ; hedge density:  $F(1, 75) = 19.03, p < .001, \eta_p^2 = .20$ ). For these measures, there was a general trend that the provision of idea support led to a significant increase in density. For example, the tasks with idea support elicited significantly more temporal connectives in learner writing than those without idea support. Also, the provision of idea support elicited more frequent lexical items (i.e., lower lexical sophistication), for which I will present possible explanations in the Discussion section.

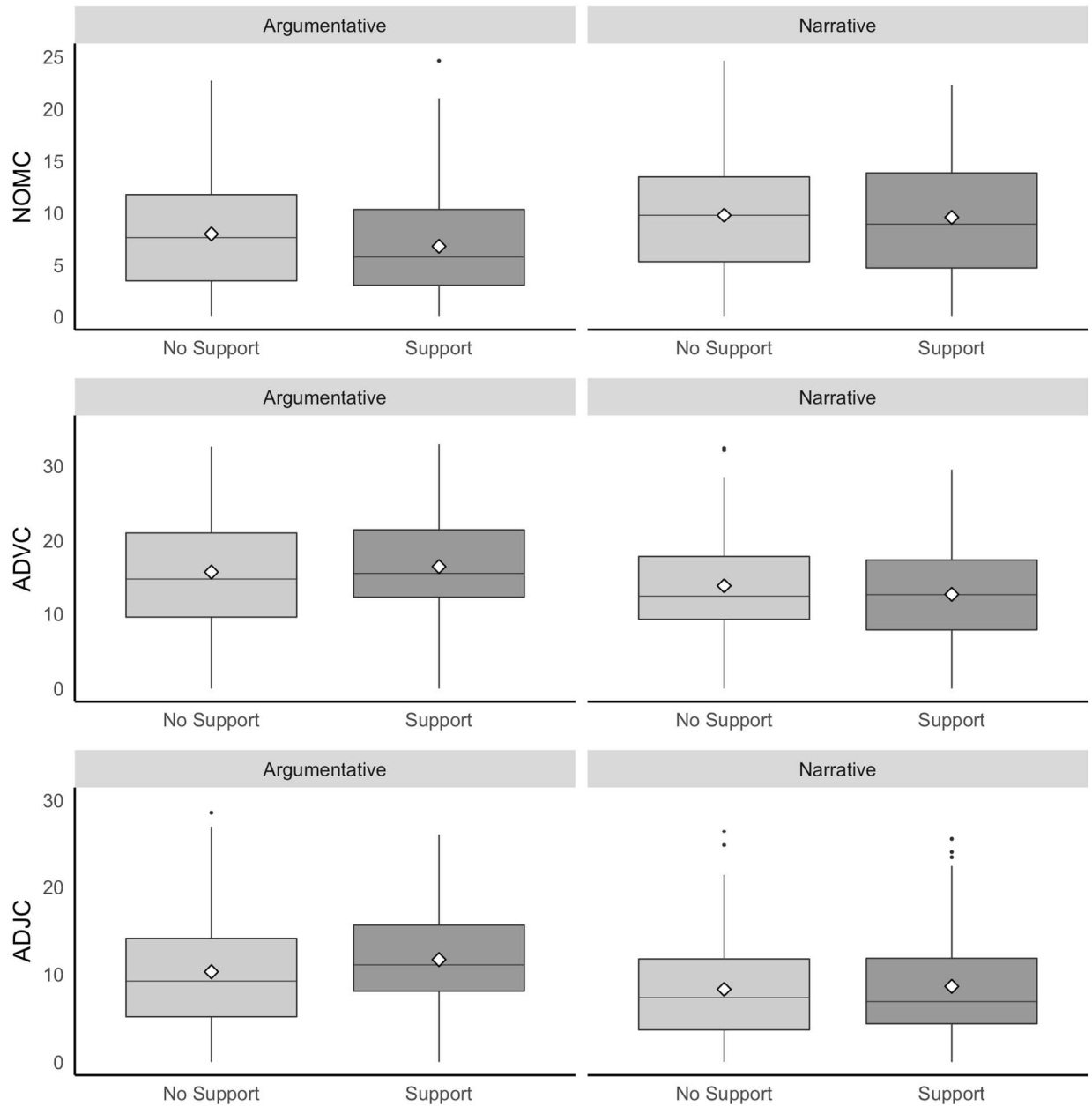


Figure 7. Nominal clause density, adverbial clause density, and adjectival clause density across genre and idea support conditions.



## **Interplay of L2 Proficiency and Task Manipulations Influencing Textual Features**

Next, I explored how the effects of genre and idea support on textual features vary with L2 proficiency in order to give insight into how genre and task manipulations need to be aligned with proficiency levels. For example, if the significant effect of genre exists only for the high-proficiency group's language (i.e., significant interaction between genre and proficiency), we can assume that low-proficiency students may not be fully capable of producing different language needed for different genres. Also, if idea support has significant effects on the low-proficiency group's language but not on that of the high-proficiency group (i.e., significant interaction between idea support and proficiency), we can assume that the manipulation of idea support may work greatly for the low-proficiency group because the cognitive complexity of the target tasks aligns well with their developmental stage. To test these hypotheses, as introduced in the Methods section, I used the high- and low-proficiency groups assigned based on cloze test performance (high-proficiency students who had cloze test scores equal or higher than 31 ( $n = 29$ ); low-proficiency students who had cloze test scores equal or lower than 25 ( $n = 28$ )) for three-way mixed ANOVAs (between-subjects variable: L2 proficiency; within-subjects variables: genre and idea support).

As shown in Table 14, the ANOVA result indicated that L2 proficiency exerted no significant main effect on any of the textual features. Additionally, there was no significant interaction that involves L2 proficiency, suggesting that the high- and low-proficiency groups constructed their essays with very similar linguistic resources. Although there were two text measures with notable three-way interactions (NOMC:  $F(1, 55) = 7.64, p = .008, \eta_p^2 = .12$ ; WF:  $F(1, 55) = 4.47, p = .039, \eta_p^2 = .08$ ) and one measure with the interaction between idea support

and proficiency (nominalization:  $F(1, 55) = 6.58, p = .013, \eta_p^2 = .11$ ), all of these measures were not statistically significant after the Bonferroni correction.

Table 15 presents the summary of the statistical analyses for the second research question (i.e., task type and L2 proficiency that led to a significant increase in textual features). To delve into the motivation for language changes across task types, I compared the results of text feature changes (Table 13) with the results of students' task perceptions (the *Students* column of Table 11). An interesting finding elicited from this comparison is that the majority of text feature changes across task types had little to do with how the students judged the writing tasks in terms of their task complexity or difficulty, clearly challenging a widely held assumption in task-based writing research. Specifically, in many previous studies, the validity of task manipulations (e.g., whether the addition of cognitive demands in a writing prompt actually leads to an increase in the cognitive burden associated with writing production) has been tested with regard to significant changes in linguistic measures, mostly those tapping the constructs of linguistic complexity or accuracy. However, in this study, while the addition of idea support, which was intended to lower students' cognitive pressure, actually led to a significant decrease in students' perceived task complexity and difficulty, this effective manipulation of task complexity did not push the students to complete the tasks with different linguistic resources.

Table 14.

*Interaction and Main Effects of L2 Proficiency on Textual Features*

Item	Genre $\times$ Idea support $\times$ Level			Genre $\times$ Level			Idea support $\times$ Level			Level		
	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power
Length of unit												
MLS	.460	.010	.113	.276	.022	.191	.875	.001	.053	.545	.007	.092
MLC	.545	.007	.092	.052	.067	.496	.257	.023	.203	.172	.034	.275
Subordination												
C/T	.819	.001	.056	.134	.040	.321	.745	.002	.062	.962	.001	.050
NOMC	.008	.122	.774	.472	.009	.110	.351	.016	.152	.845	.001	.054
ADVC	.788	.001	.058	.942	.001	.051	.404	.013	.131	.422	.012	.125
ADJC	.521	.008	.097	.426	.012	.124	.224	.027	.227	.327	.017	.163
Phrasal complexity												
CP/C	.976	.001	.050	.301	.019	.176	.092	.051	.392	.776	.001	.059
CN/C	.721	.002	.064	.428	.011	.123	.826	.001	.055	.297	.020	.179
Left embedded	.646	.004	.074	.174	.033	.273	.967	.001	.050	.634	.004	.076
Modifiers/NP	.958	.001	.050	.121	.043	.340	.976	.001	.050	.303	.019	.176
Lexical features												
D	.186	.032	.261	.101	.048	.374	.913	.001	.051	.206	.029	.242
WF	.039	.075	.547	.931	.001	.051	.642	.004	.074	.716	.002	.065
Discourse												
Argument overlap	.119	.044	.344	.736	.002	.063	.915	.001	.051	.933	.001	.051
Semantic overlap	.912	.001	.051	.326	.018	.164	.463	.010	.112	.536	.007	.094
Causal connective	.080	.055	.417	.090	.051	.396	.858	.001	.054	.793	.001	.058
Temporal connective	.750	.002	.061	.057	.064	.481	.905	.001	.052	.689	.003	.068
Nominalization	.407	.013	.130	.679	.003	.069	.013	.107	.712	.594	.005	.082
Metadiscourse												
Hedge	.794	.001	.058	.557	.006	.089	.269	.022	.196	.944	.001	.051
Booster	.776	.001	.059	.293	.020	.181	.646	.004	.074	.973	.001	.050
Self-mention	.729	.002	.064	.497	.008	.103	.820	.001	.056	.512	.008	.099
Reader pronoun	.852	.001	.054	.710	.003	.066	.505	.008	.101	.703	.003	.066

*Note.* Level = L2 proficiency level

Table 15.

*Summary of Task Manipulation and L2 Proficiency Conditions with Significantly Higher Values of Textual Features*

Construct	Genre	Idea support	L2 proficiency
Length of production	Argument	-	-
Nominal clause	Narrative	-	-
Adverbial clause	Argument	-	-
Adjectival clause	Argument	-	-
Noun phrase complexity	Argument	-	-
Lexical sophistication	Argument	No support	-
Conceptual cohesion	Argument	-	-
Connectives	Narrative	With support	-
Metadiscourse	Narrative	With support	-

Conversely, genre variation, which was shown to have little influence on students' perceptions of task complexity and difficulty, led the students to use widely different language in writing. This finding suggests the necessity of disentangling the effects of task manipulation on students' perceptions from those on their language production because different levels of cognitive burden elicited from writing tasks do not necessarily result in the formulation of different linguistic constructions, potentially due to the characteristics of the written mode that allows for a series of planning and revising (Hayes, 1996; Hayes & Flower, 1980). We also need to understand that writers modify their language to fulfill different rhetorical functions in different genres (e.g., Gilquin & Paquot, 2008; Ravid, 2005; Yasuda, 2011), pointing to the need to separate between task complexity and linguistic complexity in writing. In this respect, the findings of this study that showed extensive genre effects on L2 learners' language can be explained as the outcome of their attempt to accomplish genre-specific functions.

To further test the relationship between task complexity and linguistic complexity (or the influence of task complexity on linguistic complexity), for each task type, I computed Pearson correlations of perceived task complexity with various text features that tap linguistic complexity dimensions. As shown in Table 16, the result of this analysis indicated very limited relationships between ESL writers' perceptions of task complexity and their linguistic performance.

Table 16.

*Correlations of Perceived Task Complexity with Linguistic Complexity Features*

Linguistic features	Arg/-Support	Arg/+Support	Nar/-Support	Nar/+Support
	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>
Length of unit				
MLS	.217	.090	-.205	-.103
MLC	.146	.177	-.221	-.046
Subordination				
C/T	.072	.009	-.099	-.062
NOMC	-.081	-.051	-.096	.113
ADVC	.120	-.064	-.174	.057
ADJC	.081	-.126	-.098	-.079
Phrasal complexity				
CP/C	.064	.030	-.162	-.077
CN/C	.208	.107	-.189	.013
Left embedded	.123	.093	-.269*	.217
Modifiers/NP	.193	.149	-.167	.016
Lexical features				
D	-.212	-.018	-.168	.077
WF	.063	-.176	.106	-.171

*Note.* \*correlations are significant at the alpha level of .05.

While refuting the assumption of a close link between perceived task complexity and linguistic complexity, I found it necessary to suggest a detailed functional interpretation of genre-

specific linguistic features for more convincing arguments. To this end, I conducted a qualitative analysis of some textual features that showed clear genre variation. Of many syntactic complexity measures, nominal complexity (complex nominals per clause and modifiers per noun phrase) was found to change to the largest extent across the two genres. These notable within-subjects changes can be interpreted in terms of how ESL students' language use in written discourse reflects their selection of linguistic resources to fulfill different communicative functions. The example excerpts extracted from the two essays composed by the same writer are presented below (full essays in Appendix D). The underlined parts of the excerpts indicate complex nominals based on the scheme used for the validation of automated processing tools (Polio & Yoon, in preparation).

*The chief reason to support my idea is that an adequate foreign language is beneficial to enlarge social network. It's very common for student who study on abroad that the living level depends on the language level. In this society, the social network is very important for having a successful life. Taking my own example, I have good level of English. So I can find many internships in MSU, which are very useful for me to know many brilliant students and to enlarge social network. Hence, that can lay a fundament for my future career.* (Arg/+Support, Participant ID: S4)

*About two month ago, I was in the airplane from Beijing to Detroit. A waitress came to me and said "Sir, would you want something to drink." I was so happy, because at this time I was extremely thirsty. And I replied that "Sure, I want orange juice. Please add some ice." Then, I found the waitress was very unhappy. She said "Sir, if you want ass, please add your own ass." Eventually, I realized that my pronunciation was wrong. That I pronounced a wrong vowel sound led the waitress to misunderstand my meaning. I*

*immediately apologized to this waitress and explained my real meaning. To be honest, I felt really embarrassed in that situation. But at least I corrected a wrong pronunciation.*

(Nar/+Support, Participant ID: S4)

The first excerpt is from an argumentative essay, and the second one is from a narrative essay. As you can see from the excerpts, an ESL writer's use of complex nominals varied greatly across the two genres, clearly indicating that the use of complex noun phrases concerns an issue beyond language development but rather relates to the selection of appropriate linguistics resources in different rhetorical situations.

Additionally, narrative essays were characterized by increased use of temporal connectives and personal pronouns that are necessary for the coherent organization of a personal story. It has been widely acknowledged that the extensive use of first person pronouns allows writers to clearly denote their position as a main character in their personal story, and the use of temporal connectives contributes to linking events in chronological order. The following are the example excerpts with these points highlighted (see the D2 part of Appendix D for full essays).

*Besides understanding culture, speaking a foreign language has lots of other benefits, for instance, you will be provided a greater job opportunities related to international business. This opportunity is valuable since there are huge markets in other countries. Those who can speak many languages have earned a lot of money from international business. Moreover, by having a good command of a foreign language, you gain more fun from various activities such as traveling or watch foreign TV programs. You can enjoy different kind of view and broaden your horizons. This is a very cool experience that definitely worth a try.* (Arg/+Support, Participant ID: S45)

***One month ago**, I started my new life in America. Everything went well **at first**, and I was quite satisfied with my new circumstance here. The air was clean and fresh, and the sky was pure blue. I can seldom enjoy this kind of environment in my hometown. I was in good mood, and well-prepared to start my study life here, **until** that day I went to my first Mathematics class. I found my classroom easily and took a seat there. I was nervous since I was unfamiliar with the American teaching style, but I was confident too because my mathematics had **always** been very good in China. **When** the professor started talking, I was astonished that he spoke too fast for me to follow. (Nar/+Support, Participant ID: S45)*

These patterns presented in the excerpts clearly represent linguistic features prevalent in the entire essays (e.g., only one first person pronoun and two temporal connectives used in the entire argumentative essay).

Below are two example excerpts intended to show how various types of dependent clauses appear differently in argumentative and narrative essays (nominal clauses in bold, adverbial clauses double underlined, and adjectival clauses underlined). In interpreting these excerpts, I focus only on nominal and adjectival clauses that have fairly contrastive functions.

*With the globalization in Asia, a increasingly amount of countries are seeking the opportunities of cooperating with China, so the people who have the ability to speak other languages have more chances to participate in international events. In the meantime, the rise of international companies gives people more job opportunities, and most of the jobs they provide a relatively high income... On the other hand, you travel experience can be fantastic if you can understand the language that the country use. (Arg/+Support, Participant ID: S47)*



*I remember **that I tried to ask somebody for the right path by using English, because my friend said it's okay to say English to them, they'll understand. But soon I found out **that my biggest issue is not speaking correct English to them, but I can't understand what they reply in English.** Then I had to read their gesture, and a nice lady even used electronic dictionary in her phone to translate her word into English. Fortunately, most of them can understand **what you said in English.** All I have to do is that to get used to their Korean-style English, and I did it.** (Nar/+Support, Participant ID: S47)*

When I attended to the occurrences of nominal clauses, it was observed that narrative writing tends to include many *stative mental verb + nominal clause* constructions (verbs including *find out, remember, and understand*), whereas the excerpt from argumentative writing does not contain any nominal clause (only one case in the entire essay; see the D3 part of Appendix D). Given the major function of mental verbs for describing states and actions experienced by humans (Biber, Johansson, Leech, Conrad, & Finegan, 1999), this finding of increased nominal clauses in narrative writing can be interpreted as high-level ESL writers' attempt to describe their experience in an accurate way.

On the other hand, as illustrated by the excerpts above, argumentative essays likely contain more adjectival clauses (e.g., *people who have the ability to speak other languages and jobs they provide*). This pattern of increased postmodifying adjective clauses and complex noun phrases is known to allow the meaning of an academic text to be more compressed and denser, thus making its knowledge transfer and argumentation more effective (Biber & Gray, 2011; Halliday, 1993; Parkinson & Musgrave, 2014). This finding of higher adjectival clauses in argumentation, therefore, can be explained as ESL students' effort to convey a complex meaning from condensed nominal expressions for more convincing arguments.

### Essay Score Changes across Task Types

The third research question involved how ESL students' writing scores vary across genres and idea provision conditions. The two expert raters scored all essays using the revised analytic rubric introduced in the Method section, and their averaged scores were used. For the essays with seriously discrepant scores (subscale scores differing by 3 or more), a third rater assigned new scores, and the average of two close scores was used. Table 17 presents descriptive statistics for the essay scores analyzed in this study. Each of the rubric categories had a full score of 20 (except for mechanics whose full score was 10). The total score in Table 17 indicates the sum of the five rubric categories (full score = 90).

Table 18 presents the results of two-way ANOVAs with genre and idea support as within-subjects variables. As shown in Figure 8, the result indicated significant interaction effects between genre and idea support on content ( $F(1, 75) = 11.16, p = .001, \eta_p^2 = .13$ ), organization ( $F(1, 75) = 7.82, p = .007, \eta_p^2 = .09$ ), and language use scores ( $F(1, 75) = 7.51, p = .008, \eta_p^2 = .09$ ), jointly leading to a significant interaction between genre and idea support on the essays' total scores ( $F(1, 75) = 9.47, p = .003, \eta_p^2 = .11$ ). Specifically, significantly higher scores (or such pattern with no statistical significance) were given to the three rubric categories (content, organization, and language use) for the argumentative prompt with idea support (content:  $t(75) = -2.82, p = .006, d = -0.32$ ; organization:  $t(75) = -2.81, p = .006, d = -0.33$ ; language use:  $t(75) = -1.60, p = .11, d = -0.19$ ) and for the narrative prompt without idea support (content:  $t(75) = 2.09, p = .04, d = 0.24$ ; organization:  $t(75) = 1.70, p = .09, d = 0.20$ ; language use:  $t(75) = 2.36, p = .02, d = 0.27$ ). That is, the condition of idea support entailed a positive impact on the quality of argumentative writing but negatively affected the quality of narrative writing, particularly with regard to idea development (the largest effect size for the content category).

Table 17.

*Descriptive Statistics for Essay Scores by Genre and Idea Support*

Category (full score)	Arg/-Support		Arg/+Support		Nar/-Support		Nar/+Support	
	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI
Content (20)	12.63 (2.21)	[12.12, 13.13]	13.23 (2.22)	[12.72, 13.74]	14.15 (2.11)	[13.67, 14.63]	13.53 (2.24)	[13.01, 14.04]
Organization (20)	12.65 (2.28)	[12.12, 13.17]	13.26 (2.01)	[12.80, 13.72]	14.36 (1.92)	[13.92, 14.79]	13.88 (2.01)	[13.42, 14.34]
Vocabulary (20)	13.53 (1.54)	[13.17, 13.88]	13.70 (1.38)	[13.38, 14.01]	14.11 (1.39)	[13.79, 14.42]	13.63 (1.50)	[13.28, 13.97]
Language use (20)	13.53 (1.76)	[13.13, 13.93]	13.84 (1.45)	[13.51, 14.17]	13.88 (1.69)	[13.50, 14.27]	13.39 (1.51)	[13.04, 13.73]
Mechanics (10)	7.43 (1.26)	[7.14, 7.72]	7.34 (1.19)	[7.06, 7.36]	7.44 (1.25)	[7.15, 7.73]	7.33 (0.96)	[7.11, 7.54]
Total score (90)	59.75 (7.53)	[58.03, 61.47]	61.36 (6.89)	[59.78, 62.93]	63.93 (6.51)	[62.45, 65.42]	61.75 (7.04)	[60.14, 63.35]

Table 18.

*Inferential Statistics for Genre and Idea Support Effects on Essay Scores*

Category	Genre			Idea support			Genre × Idea support		
	<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power	<i>p</i>	$\eta_p^2$	Observed power
Content	< .001*	.182	.981	.957	.001	.050	.001*	.130	.909
Organization	< .001*	.303	1.000	.662	.003	.072	.007*	.094	.788
Vocabulary	.079	.041	.420	.210	.021	.239	.027	.064	.608
Language use	.756	.001	.061	.507	.006	.101	.008*	.091	.772
Mechanics	.999	.001	.050	.300	.014	.178	.925	.001	.051
Total score	< .001*	.133	.918	.584	.004	.084	.003*	.112	.859

Note. \**p* values are significant with the Bonferroni correction (alpha = .05/6 or .0083).

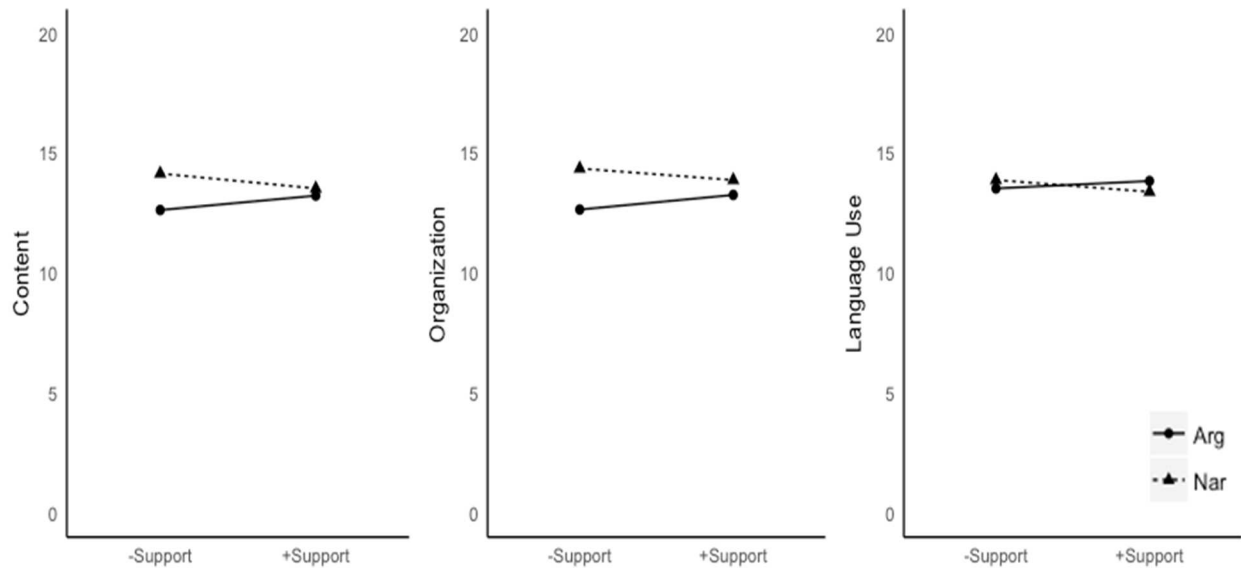


Figure 8. Interaction plots for content, organization, and language use scores showing an interaction between genre and idea support conditions.

This result can be seen as evidence for beneficial effects of supporting ideas on the perceptions and production of argumentative writing because such provided ideas would enable the students to focus on developing more detailed ideas and coherent organization. On the other hand, in the personal narrative genre, the provision of supporting ideas potentially restricts L2 learners to a limited range of storylines provided in the prompt rather than helps them develop fully developed stories, resulting in lower scores on the narrative essays composed with supporting ideas.

Additionally, it was found that the students obtained significantly higher content and organization scores on their narratives than argumentative essays (content:  $F(1, 75) = 16.65, p < .001, \eta_p^2 = .18$ ; organization:  $F(1, 75) = 32.65, p < .001, \eta_p^2 = .30$ ), with a particularly large effect on organization scores (see Figure 9). This result that showed a clear genre effect on discourse-level writing scores can be interpreted either as the outcome of an actual difference in essay quality across the two genres or as the outcome of the difficulty of assigning comparable

scores on discourse-level categories due to raters' different levels of strictness with regard to genre (Hamp-Lyons & Mathias, 1994). However, there was no significant influence of genre on any of the sentence-level writing scores (vocabulary, language use, and mechanics), which also contrasts with the syntactic complexity finding that showed prevalent genre effects.

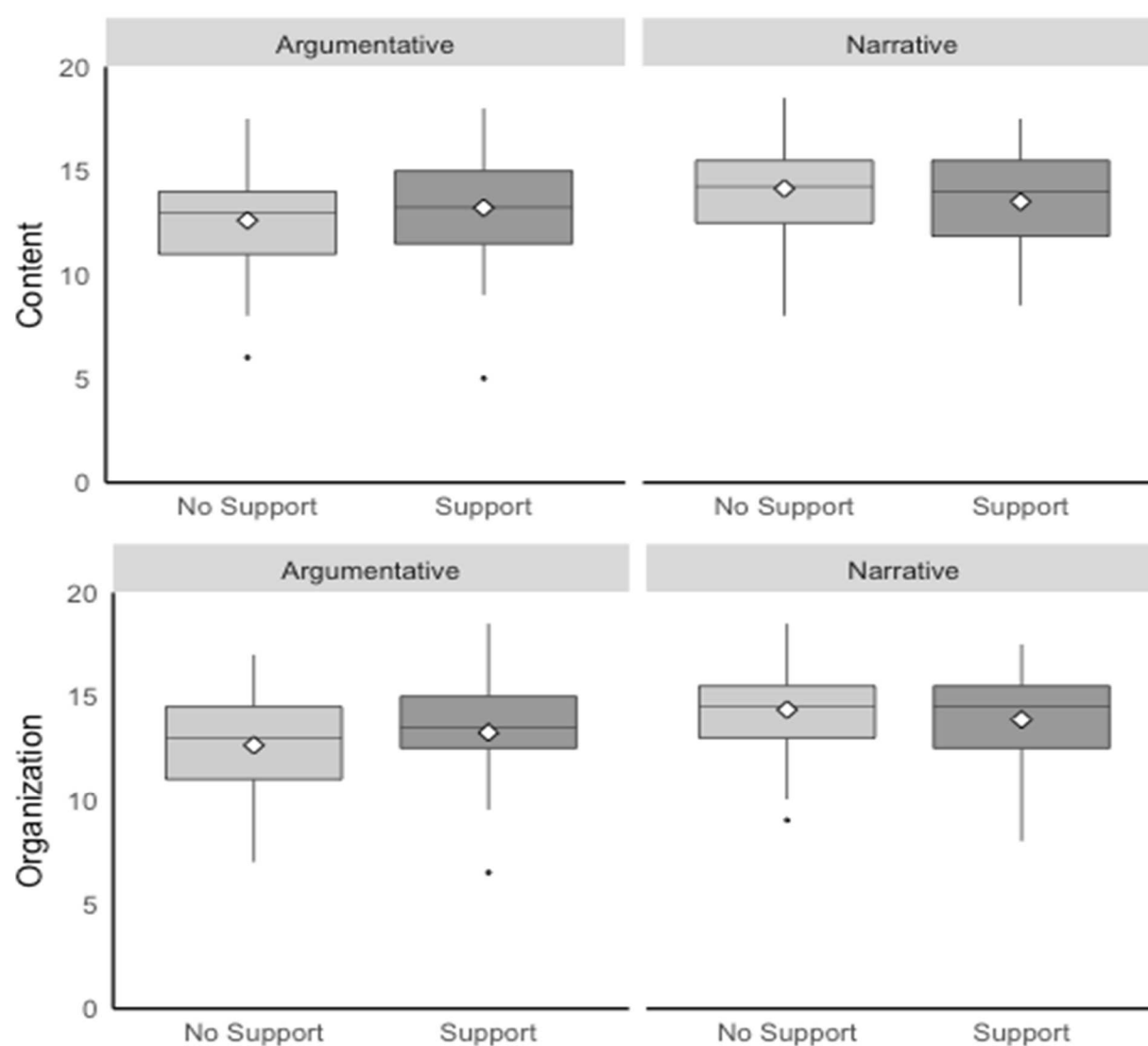


Figure 9. Content and organization scores across genre and idea support conditions.

The result showed that none of the rubric categories had a significant main effect of idea support, indicating that, despite a clear function of idea support in relieving L2 learners' cognitive burden, the existence of supporting ideas in the prompts did not necessarily lead to

different essay scores. Taking into account the result of essay scores and that of learner perceptions together, I suggest that the students' subjective judgments of writing tasks do not correspond with the quality of their essays assessed by expert raters.

### **Interplay of L2 proficiency and Task Manipulations Influencing Essay Scores**

Thus far, I have demonstrated that genre and idea support had interaction effects on various dimensions of essay quality (content, organization, and language use). Additionally, I showed that genre exerted a significant effect on discourse-level essay quality (content and organization), while idea support has no significant effect on any of the rubric categories. To explore the potential interplay of L2 proficiency and task manipulations, I computed three-way mixed ANOVAs with L2 proficiency as a between-subjects variable, as well as genre and idea support as within-subjects variables (Table 19 for descriptive statistics). The alpha level was set with the Bonferroni adjustment ( $\alpha = .05/6$  or  $.0083$ ). As shown in Table 20, the result indicated that the high-proficiency students received significantly higher scores on sentence-level rubric categories than the low-proficiency students (vocabulary:  $F(1, 55) = 9.75, p = .003, \eta_p^2 = .15$ ; language use:  $F(1, 55) = 8.78, p = .004, \eta_p^2 = .14$ ), while the effect of L2 proficiency on the content and organization categories approached statistical significance (content:  $F(1, 55) = 6.49, p = .014, \eta_p^2 = .11$ ; organization:  $F(1, 55) = 6.49, p = .014, \eta_p^2 = .11$ ). Figure 10 illustrates specific patterns of L2 proficiency effects on vocabulary and language use scores across task types.

Table 19.

*Descriptive Statistics for Essay Scores by L2 Proficiency, Genre, and Idea Support*

Category (full score)	Level	Arg/-Support		Arg/+Support		Nar/-Support		Nar/+Support	
		<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI	<i>M</i> ( <i>SD</i> )	95% CI
Content (20)	High	13.22 (1.93)	[12.49, 13.96]	13.95 (1.88)	[13.23, 14.66]	14.40 (2.21)	[13.56, 15.24]	14.22 (1.77)	[13.55, 14.90]
	Low	12.09 (2.77)	[11.02, 13.16]	12.86 (2.75)	[11.79, 13.92]	13.80 (1.93)	[13.06, 14.55]	12.63 (2.47)	[11.67, 13.58]
Organization (20)	High	13.19 (2.01)	[12.43, 13.95]	13.74 (1.79)	[13.06, 14.42]	14.62 (1.89)	[13.90, 15.34]	14.64 (1.61)	[14.03, 15.25]
	Low	11.95 (2.87)	[10.84, 13.06]	13.04 (2.43)	[12.09, 13.98]	14.02 (1.98)	[13.25, 14.79]	12.95 (2.29)	[12.06, 13.84]
Vocabulary (20)	High	14.07 (1.27)	[13.59, 14.55]	14.03 (1.16)	[13.59, 14.50]	14.59 (1.42)	[14.05, 15.13]	14.17 (1.27)	[13.69, 14.66]
	Low	13.07 (1.80)	[12.37, 13.77]	13.61 (1.83)	[12.90, 14.32]	13.75 (1.17)	[13.30, 14.21]	12.96 (1.70)	[12.31, 13.62]
Language use (20)	High	14.21 (1.64)	[13.58, 14.83]	14.26 (1.45)	[13.71, 14.81]	14.41 (1.28)	[13.93, 14.90]	13.74 (1.37)	[13.22, 14.26]
	Low	12.79 (1.80)	[12.09, 13.48]	13.54 (1.62)	[12.91, 14.16]	13.68 (1.71)	[13.02, 14.34]	12.98 (1.75)	[12.30, 13.66]
Mechanics (10)	High	7.58 (1.38)	[7.05, 8.10]	7.51 (1.31)	[7.00, 8.01]	7.52 (1.23)	[7.05, 7.98]	7.71 (0.90)	[7.36, 8.05]
	Low	7.21 (1.30)	[6.70, 7.71]	7.20 (1.16)	[6.75, 7.65]	7.52 (1.24)	[7.04, 8.00]	7.09 (0.97)	[6.72, 7.46]
Total score (90)	High	62.27 (7.05)	[59.59, 64.95]	63.49 (6.03)	[61.20, 65.78]	65.53 (6.28)	[63.15, 67.92]	64.48 (5.51)	[62.39, 66.58]
	Low	57.10 (8.80)	[53.69, 60.51]	60.23 (8.69)	[56.86, 63.60]	62.77 (6.23)	[60.35, 65.18]	58.61 (8.20)	[55.43, 61.79]

Table 20.

*Interaction and Main Effects of L2 Proficiency on Textual Features*

Category	Genre $\times$ Idea support $\times$ Level			Genre $\times$ Level			Idea support $\times$ Level			Level		
	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power	$p$	$\eta_p^2$	Observed power
Content	.182	.032	.264	.976	.001	.050	.292	.020	.182	.014	.106	.707
Organization	.065	.061	.456	.740	.002	.062	.455	.010	.115	.014	.106	.707
Vocabulary	.135	.040	.320	.361	.015	.148	.744	.002	.062	.003*	.151	.866
Language use	.274	.022	.192	.341	.017	.157	.308	.019	.173	.004*	.138	.829
Mechanics	.133	.041	.323	.890	.001	.052	.241	.025	.214	.189	.031	.257
Total score	.063	.062	.463	.949	.001	.050	.654	.004	.073	.004*	.139	.833

*Note.* Level = L2 proficiency level; \* $p$  values are significant with the Bonferroni correction (alpha = .05/6 or .0083).



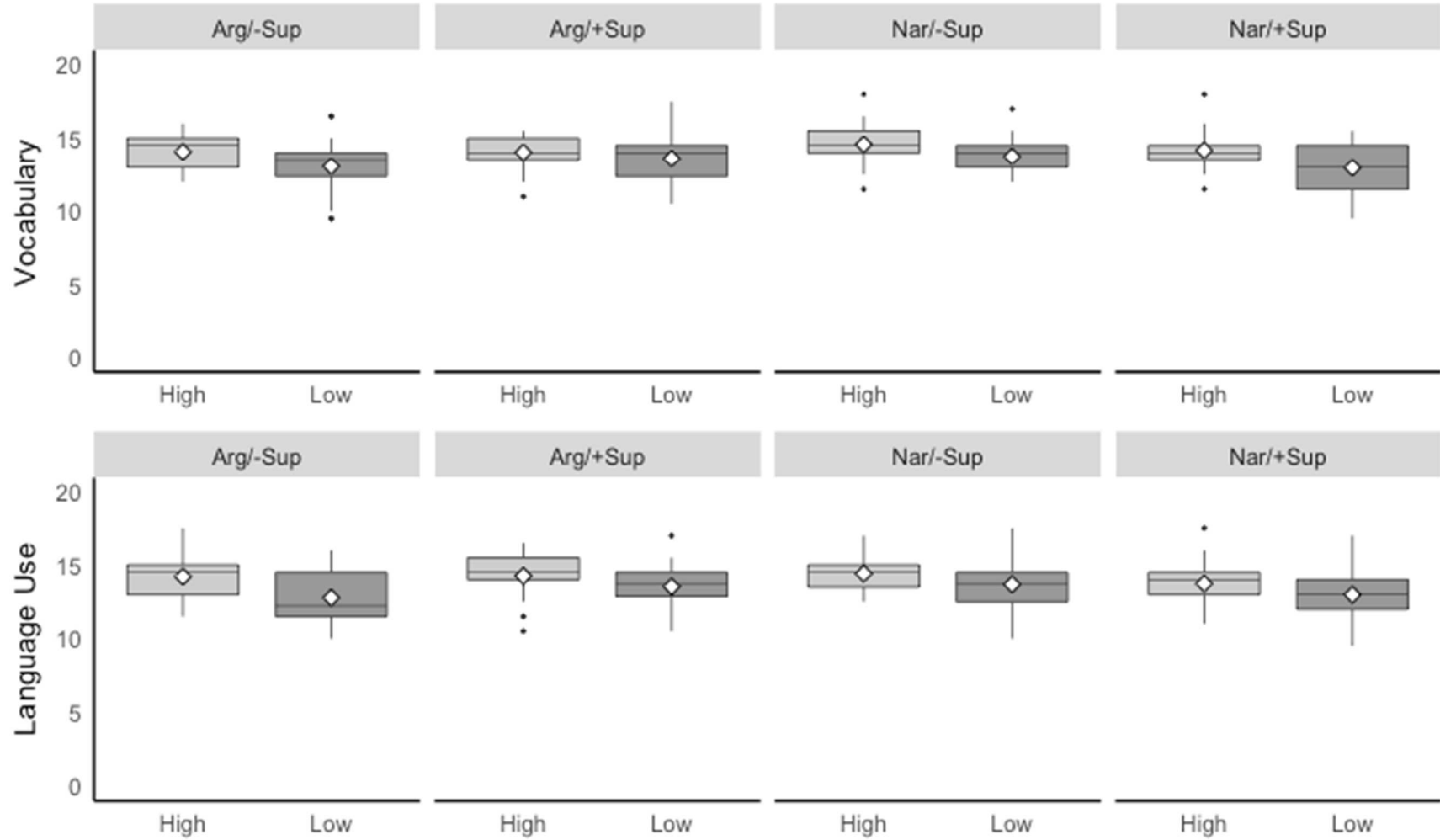


Figure 10. Vocabulary and language use scores across task types and L2 proficiency.

Of particular note here is that the high- and low-proficiency groups had statistically different essay scores (greater difference in sentence-level rubric categories), whereas the two groups did not differ in their use of linguistic resources in writing. This finding potentially indicates that the quality of language use and vocabulary involves qualitative dimensions that cannot be fully captured through quantity-based textual features. That is, while there is no group difference in their use of textual features, the high-proficiency group may still have better command of the target language in fulfilling the goal of a writing task. In contrast, the result showed no significant interaction that involves L2 proficiency (see Table 20), suggesting that the impact of task manipulations on essay quality is likely to be consistent regardless of L2 proficiency (or at least within the proficiency level range targeted in this study).

## **CHAPTER 5.**

### **DISCUSSION**

#### **ESL Students' and Teachers' Perceptions of Writing Tasks**

This study aimed to add to the limited amount of research into the perceptions and production of various L2 writing tasks. The results of the questionnaires indicated that there is a gap between students' and teachers' perceptions of the writing tasks adopted in this study. As shown in Table 8, the most notable difference between the two groups involved the cognitive complexity and difficulty imposed by each of the two genres. Specifically, although the teachers predicted that ESL students would have greater cognitive pressure and difficulty in composing the argumentative genre than the narrative, the students found both genres causing a similar level of complexity and difficulty. In fact, the teachers' expectations of genre-specific cognitive demands imposed by argumentative and narrative tasks reflect how L2 researchers have explained their findings that involved multiple genres (e.g., higher cognitive demands of non-narrative writing than narrative; Ruiz-Funes, 2014, 2015; Yang, 2014), which merits further discussion.

It has been a widely accepted belief that L2 students would find the argumentative genre more cognitively demanding than the narrative because the former necessitates students' higher-order reasoning and interpretation that goes beyond knowledge telling (Bereiter & Scardamalia, 1987). It may be true that reasoning skills needed to fulfill argumentative tasks are more difficult to obtain than those needed for narrative tasks and that such argumentation skills require more conceptual processes of writers. This may be why young writers, who have not fully developed a mature cognitive system, have greater difficulty completing argumentative or expository tasks than narratives, as shown in much L1 writing research (e.g., Berman, 2008; Engelhard et al.,

1992; Ravid, 2005). However, this prediction does not seem to be in line with ESL students' actual perceptions of argumentative and narrative tasks, potentially because of their extensive experience with argumentation as a primary genre in academic settings (Christie, 1997; Johns, 1995; Mei, 2006).

More specifically, I argue that the same prediction about a genre-cognition connection should not be made to adult L2 learners who have extensive academic writing experience and are equipped with a full-fledged cognitive system. Cognitive models of writing processes (Hayes, 1996; Hayes & Flower, 1980) emphasize the mediating effects of genre schemas, task schemas, and other long-term memory factors (e.g., topic awareness) on working memory pressures during writing. Therefore, a potential explanation is that the majority of adult L2 writers who had much experience in preparing for a standardized L2 writing test are likely to possess well-established genre schemas for argumentation, and accordingly the use of these genre schemas probably leads to a reduced processing burden during argumentative writing despite the inherent, higher-level cognitive loads of this particular genre.

Unlike the lack of genre effects on the students' perceptions of task complexity and difficulty, the finding of this study showed that the idea support condition led to a significant change in the level of perceived complexity and difficulty for both genres. In line with Révész et al.'s (in press) results using argumentative tasks, the current finding from the argumentative and narrative genres can be seen as additional evidence of idea support as a valid task manipulation in written discourse. With this finding as a starting point, future studies would be able to explore how to maximize the intended impact of idea support manipulations in various writing tasks and test the applicability of other task variables to the written modality (e.g., exploring the function

of the number of elements in writing based on the Triadic Componential Framework; Robinson, 2001b, 2007).

Regarding the role of idea support for different genres, in their open-ended responses, the teachers expressed concerns about the potentially adverse effect of idea support on narrative writing performance. This point was also confirmed by the teachers' response to the task perception questionnaire that indicated a significant interaction between genre and idea support on task complexity and difficulty (i.e., teachers' expectations that the provision of idea support in argumentative writing would decrease its cognitive complexity and difficulty, whereas the same manipulation would increase the complexity and difficulty of narrative writing). That is, considering the nature of personal narratives, having students draw on specific storylines provided by a task developer can cause detrimental effects on their writing performance because the given stories can be largely irrelevant to students' experience (Hinkel, 2002; Lo & F. Hyland, 2007). Therefore, considering the present result that showed a negative effect of supporting ideas on narrative writing scores (RQ 3) as well as the previous findings that showed the elicitation of increased syntactic complexity and better performance from a topic more closely related to students' lives (Hinkel, 2002; Yoon, 2017b), I argue that all information constituting a writing prompt (e.g., topic, task, and supporting ideas) should be relevant to writers' experience in order to elicit their best performance.

An additional point to discuss from the perception result is the level of task interest and motivation across the two genres. The results showed that both students and teachers found the narrative genre involving more interest-sparking features than the argumentative. In this regard, Zhang (2013) stated that "many ESL learners' personal written narratives are embodiments of their dreams and aspirations" (p. 447), which implies that personal narrative writing is a medium

that enables students to communicate their experience in written discourse. Also, because narrative writing is full of culture- and language-specific characteristics (Berman & Slobin, 1994; Kang, 2005), an instructional focus on narrative writing will allow ESL students to learn how to use their linguistic and cultural resources in organizing their personal thoughts.

In terms of relationships between task perception items (see Table 10), the result showed that, although significantly correlated, task complexity and difficulty operate as two different constructs (Révész et al., 2016), as demonstrated by the positive relationship of task complexity with task interest and motivation, which did not hold true for task difficulty in most cases. I view this finding as evidence pointing to the importance of developing a task appropriately challenging to the target student population. For example, if a writing task is too simple to students, they will not be fully engaged in the task and have lower motivation for completing the task successfully. Likewise, Xu (2003) suggested the use of moderately challenging tasks as one of the ways to increase L2 learning motivation. Melendy (2008) also showed that approximately 50% of the undergraduate student participants selected the most challenging writing task when asked to select one out of the three task options to complete for assessment purposes. Given these findings, going beyond the well-known sequencing of simple-to-complex tasks in a language curriculum (Robinson, 2010), our next step is to build a framework for designing appropriately challenging tasks for students at various proficiency levels (and for those with different educational backgrounds).

### **Effects of Task Type on Textual Features**

The second goal of this study was to explore various textual features with a focus on how they vary across task types. I first examined the effect of genre and idea support on the language use of all student participants and, then, further analyzed how such task type effects interact with

the students' L2 proficiency. The major finding of these analyses is that the language produced by ESL students differed widely across the two genres, while their language differed to a limited extent across the idea support conditions. This confirms some of the previous findings and, at the same time, refutes several assumptions that have existed in the field of task-based writing research.

First, supporting the findings of previous research (e.g., Lu, 2011; Qin & Uccelli, 2016; Way et al., 2000; Yoon & Polio, 2017), I argue that genre indeed functions as a task variable that elicits different linguistic features from L2 learners. Specifically, it was confirmed that the argumentative genre leads students to produce syntactically more complex language, while the narrative allows them to produce more temporal connectives and first person pronouns. In this regard, I showed the argumentative and narrative excerpts that were composed by one writer but were characterized by notably different linguistic structures, suggesting evidence of the writer's understanding of register flexibility and capability of communicating different meanings across the two genres. We can infer from this finding that, for example, temporal connectives and personal pronouns need to be targeted as linguistic resources for coherent narrative writing.

In addition, using the fine-grained measures of subordination (nominal, adverbial, and adjectival clause density), I found that the argumentative essays indicated greater adverbial and adjectival clause density, while the narratives showed greater nominal clause density. The present finding that contrasts with the previous findings of genre effects on clausal syntactic complexity (e.g., Lu, 2011; Yoon & Polio, 2017) points to the importance of adopting more specific measures when exploring genre effects (or generally task type effects) on clausal subordination (i.e., Frear & Bitchener, 2015). Also, as we observed from the examination of the essay excerpts, researchers need to interpret genre-specific language structures with regard to their

communicative functions necessary or useful for that particular genre (or task) because one of the important functions of language tasks is to elicit task natural, useful, and essential structures from L2 learners (Loschky & Bley-Vroman, 1993). For example, L2 learners with adequate competence in grammar will be prompted to use more nominal clauses and temporal connectives in the narrative task, while using more adverbial and adjectival clause structures in the argumentative task, because different language structures are useful for the completion of different genres.

The findings of the present study have indicated that ESL students at high intermediate or low advanced proficiency seem to have sufficient genre awareness and understand the need to write differently in different contexts. Particularly, I have shown how rhetorical functions associated with each genre leads to a range of genre-specific linguistic features, demonstrating the importance of focusing on what meaning writers attempt to communicate in their writing rather than on how the different cognitive demands of writing tasks lead to changes in language use. That is, as Berman and Slobin (1994) suggested, “the development of grammar cannot be profitably considered without attention to the psycholinguistic and communicative demands of the production of connective discourse” (p. 2). This argument for the connection between rhetorical functions and linguistic features is further strengthened by the result that showed no genre effects on perceived task complexity and difficulty.

As I discussed above, unlike prevalent effects of genre, the provision of idea support influenced learners’ language use to a limited extent. Specifically, the result showed a significant increase in a few textual features in the idea support condition (e.g., temporal connective, nominalization, and hedge density), while lexical sophistication was significantly lower (i.e., higher word frequency) in the essays composed with supporting ideas. There are several



interpretations of this finding, each of which is discussed here in terms of their viability. The first explanation involves priming effects on language use, which has been investigated extensively with a focus on oral language development (see McDonough & Trofimovich, 2011). For example, when given a prompt that includes many low-frequency words, L2 learners who are likely to borrow some words included in the prompt due to their limited lexical repertoire would compose an essay that contains more low-frequency words. When checked for the current prompts, however, this explanation did not hold true because there was no particular pattern of higher or lower levels of word frequency between the +Support and -Support prompts (average word frequency of all words used in each prompt: Arg/-Support: 2.85; Arg/+Support: 2.81; Nar/-Support: 3.00; Nar/+Support: 3.00).

Another possibility is the influence of essay length on lexical sophistication. It has been argued that various dimensions of linguistic complexity, accuracy, and fluency tend to be in competition due to limited cognitive resources (Skehan, 1998, 2009; Skehan & Foster, 2001). While some researchers argued for the positive relationship between linguistic complexity and accuracy (Robinson, 2001a, 2005, 2007), it is conceivable that an essay full of sophisticated words would be relatively shorter than that full of simple words when composed under the same time constraint. Therefore, a potential scenario is that if the idea support condition in fact encouraged students to write more within a given time, their greater attention to fluency (i.e., completing a lengthier essay) might have led to lower lexical sophistication. I tested this hypothesis by examining text length (total word count) for each task type as well as the relationship between text length and word frequency. This analysis showed inconsistent patterns of change in text length with regard to the idea support condition (text length in words: Arg/-Support: 289.09; Arg/+Support: 302.80; Nar/-Support: 310.67; Nar/+Support: 301.39), offering

no evidence for this hypothesis. Similarly, the correlation result showed the lack of relationships between text length and word frequency (Arg/-Support:  $r = .11$ ; Arg/+Support:  $r = .03$ ; Nar/-Support:  $r = .04$ ; Nar/+Support:  $r = .08$ ), rejecting the feasibility of this explanation.

By refuting these two interpretations, I was assured that the decrease in lexical sophistication (i.e., higher average word frequency) in the idea support condition (i.e., less complex tasks as indicated in the student perception result) might be evidence of a significant impact of task complexity on lexical complexity. That is, of various dimensions of linguistic complexity, lexical sophistication is probably the only area that gives reliable support to the connection between cognitive complexity and linguistic complexity. This explanation is in line with most previous task-based studies that explored the effect of idea support (e.g., Kormos, 2011; Ong & Zhang, 2010; Révész et al., in press). For example, Ong and Zhang found that the increase in cognitive complexity through idea support led to greater lexical complexity but little change in fluency. Additionally, Révész et al. and Kormos indicated significant effects of idea support on lexical complexity but not on the majority of other complexity or accuracy measures.

Given such consistent findings of the association between task complexity and lexical complexity in written discourse (Kormos, 2011; Ong & Zhang, 2010; Révész et al., in press), I tentatively argue that the major area on which the cognitive burden of a writing task exerts an influence is the extent to which ESL student use sophisticated lexical items. Specifically, when given a more cognitively demanding task in which students need to come up with more specific and relevant ideas, they would direct a greater amount of their attentional resources to using more sophisticated words. In contrast, the majority of syntactic complexity dimensions that are exploited to fulfill various communicative functions would not be greatly influenced by the cognitive demand of a task in the written mode.

Last, regarding a significant interaction between genre and idea support on nominal features, I could infer that, with more cognitive resources made available from idea support, L2 learners might be able to provide more packed information by increasing the use of complex noun phrases and nominalizations in argumentative writing, which contributed to making convincing arguments. In contrast, the greater amount of cognitive resources, which could be used for more intriguing narration, led L2 learners to focus even less on nominal features because complex noun phrases and nominalizations make the text more informational and, accordingly, less interpersonal (Halliday & Mathiessen, 1999).

### **Effects of Task Type on Essay Quality**

Regarding the effect of task type on essay scores, this study showed several important findings. First, while the provision of supporting ideas resulted in higher argumentative essay scores, the same task manipulation led to lower narrative writing scores. This interaction effect between genre and idea support on quality scores existed for three rubric categories (content, organization, and language use), with the largest effect on content scores. If we equate an essay score assigned by expert raters with the quality of an essay, this finding can be interpreted as varying roles of idea support in assisting L2 writers across genres.

One possible scenario is that, given some ideas to use as supporting points, ESL students might have had a lower cognitive burden for completing the argumentative task, which enabled them to allocate their greater cognitive and attentional resources to other writing areas related to language construction and essay structure. However, when given several possible plots that needed to be incorporated for narrative writing, students might have felt forced to use them rather than narrate their own stories, potentially leading to the narration of a less relevant story and, consequently, to lower essay scores. This point was expressed in the teachers' responses to

open-ended survey questions; specifically, as shown in some excerpts above, several teachers indicated that the incorporation of supporting ideas that are not relevant to ESL students' experience could be a challenging task to them.

This finding can also be explained as the outcome of varying areas that ESL students find challenging in composing different genres. For example, if ESL students can improve the quality of their argumentative writing with some supporting ideas provided in a prompt, it can be inferred that the area of students' difficulty involves coming up with logical, convincing ideas in the argumentative genre and, accordingly, that idea development needs greater pedagogical attention when teaching argumentation. In contrast, adult students may already have sufficient ideas and experiences to use as storyline resources for the personal narrative task. Accordingly, the support that the students probably need for narrative writing is register-specific linguistic expressions that they can rely on when turning their experience at the conceptual level into the language needed to complete the narration.

I suggested in the Literature Review section that, considering the emphasis of standardized L2 writing tests on argumentation (Qin & Karabacak, 2010), adult ESL students are likely to have experienced narrative tasks much less than many researchers and teachers have expected. Because the students received higher scores on the narrative tasks than on the argumentative tasks, the finding of this study does not fully support this reasoning that points to the need for more instructional focus on L2 narrative writing. However, the perception results indicated that there is a wide gap between the students and teachers in how they view different genres, and the students did not see the argumentative tasks more cognitively demanding than the narrative, despite the potentially increased reasoning for argumentation.

Based on these findings, I suggest that the writing instruction intended to teach narrative-related linguistic resources (e.g., particle phrasal verbs, locative elements, and temporal connections) would contribute to expanding ESL students' genre conventions and improving their general L2 writing proficiency. One of the potential reasons that ESL students have difficulty fulfilling L2 narrative writing is the need to use many particle verbs in expressing the path and manner of motion in the narrative of English, a satellite-framed language (see Berman & Slobin, 1994). Particularly, ESL students who use L1s that tend to express manner and path in verbs (i.e., verb-framed languages such as Hebrew, Japanese, Korean, Spanish, and Turkish, although still under debate) can find it very challenging to use various types of particle verbs appropriately (Slobin, 2004; Talmy, 1985, 2000). This argument, however, is somewhat speculative; thus, exploring the effect of such instruction on ESL students' perceptions and production of narrative writing will advance our understanding of the development of narrative writing skills.

Another major finding related to essay quality is that ESL students received significantly higher scores on the narratives than the argumentative essays, offering confirmatory evidence against the generalizability of writing scores across different genres (e.g., Bouwer et al., 2015; Way et al., 2000). Particularly, the results indicated significantly higher content and organization scores on the narrative genre than the argumentative. One of the possible explanations for this finding is that students were expected to follow more rigid top-down organization rules for argumentative writing (Berman, 2008) and, as a result, argumentative essays that did not meet such organizational expectations were likely to gain lower scores. Narrative writing typically involves a linear structure, which is less salient than argumentative writing's hierarchical, top-down structure (i.e., main ideas first, followed by supporting information) (Van Dijk & Kintsch,

1983). These genre-specific expectations, from a perspective of rater effects, might encourage raters to be more lenient for the narrative genre with regard to content and organization, suggesting the need to have raters better understand such potential genre effects on their rating behavior and to train them to evaluate different writing tasks more reliably.

Much L1 and L2 research has interpreted their finding of higher narrative writing scores than those of non-narrative genres as the outcome of higher cognitive demands of non-narrative genres (e.g., Bouwer et al., 2015; Crowhurst, 1980; Engelhard et al., 1992; Kegley, 1986; Way et al., 2000). However, in this study, I do not attribute a significant genre difference in text quality scores to the cognitive demands required by different genres because task perception results showed no difference in task complexity or difficulty between argumentative and narrative tasks. Instead, considering other dimensions of task perceptions, I suggest the potential role of students' interest and motivation in eliciting different score between the two genres. The perception results from the student participants indicated significantly higher interest and motivation for narrative writing than for argumentative, which might have led them to devote more attention to the narrative genre. It has been extensively documented that task interest and attitudes exert a significant impact on writing performance (e.g., Graham, Berninger, & Fan, 2007; Knudson, 1995; Lo & F. Hyland, 2007; Zimmerman & Bandura, 1994), and the finding of this study can be seen as empirical evidence partly supporting this claim, although it still needs more reliable data controlled for genre and rater effects.

The last point to discuss in this chapter is the interaction of L2 proficiency and task type on essay scores. Previously, I showed that none of 21 text features was significantly influenced by L2 proficiency, which was interpreted as the consequence of either a narrow proficiency range or the incorrect alignment of writing tasks with target learner characteristics. However, the

finding of essay scores showed significant main effects of L2 proficiency on vocabulary and language use scores. That is, ESL students' sentence-level writing scores (vocabulary and language use categories) better reflected their L2 proficiency than discourse-level scores (content and organization categories) did. This finding can be interpreted mainly in two ways. First, this finding can be seen as the effect of rating behaviors. In this regard, Rezaei and Lovorn (2010) revealed raters' greater sensitivity to syntactic and mechanical features than to content or rhetorical features, meaning, for example, that a subtle difference in the quality of sentence-level features between essays can lead to changes in their scores. Accordingly, despite a narrow range of L2 proficiency levels among the student participants, they still had significantly different scores on their use of syntactic structures and lexical items.

Another possible interpretation is that L2 proficiency approximated by cloze test scores might tap sentence-level writing skills, better reflecting the development of sentence-level writing skills. There has been much debate on whether cloze tests are capable to assess both sentence- and discourse-level competence or they can only assess sentence-level competence (see Tremblay, 2011); thus, I acknowledge the possibility that different patterns might have been obtained with different measures of L2 proficiency. This issue can be resolved by using a more objective, standardized method to assess L2 proficiency, or by replicating this study with ESL students at a much lower proficiency level. I particularly assume that the latter will allow us to obtain a more comprehensive picture of proficiency effects on the performance of different writing tasks. In the following section, I will discuss implications of the present study and directions for future research.

## **CHAPTER 6.**

### **CONCLUSION**

#### **Theoretical and Research Implications**

This study offers several important implications for L2 writing research. First, as the perception result showed, there is a possibility that some interpretations based on long-standing beliefs do not accurately depict the motivation behind what have been empirically observed. The presumption that I intended to explore and challenge involved the genre-cognition connection in L2 writing research. Thus far, many L2 researchers have explained their findings of cross-genre language and score differences as arising from the difference in cognitive pressure between genres (e.g., argumentative tasks as cognitively more complex than narrative tasks), and this practice has been widely accepted in the field because many have believed that linguistic features are dependent on cognitive processes due to humans' limited cognitive resources and the majority of previous research has produced very consistent findings of higher linguistic complexity and lower essay scores in the argumentative genre than the narrative. However, as evidenced by the findings of the present study, L2 learners' perceptions of the complexity and difficulty of writing tasks have little to do with linguistic features or quality scores of their essays. Specifically, it was found that the majority of textual features are a manifestation of a set of communicative functions demanded by each genre, while lexical sophistication is one of a few areas that were shown to differ according to the cognitive complexity of writing tasks.

Therefore, task-based writing researchers should not set out to investigate their research questions with the presumption of task-specific challenges and task manipulation effects because, for example, their prediction of task manipulation effects would not always match students' actual perceptions of different tasks. A possible way of addressing this issue is to



conduct task-based research in two separate stages: (1) testing students' perceptions of task manipulation effects for various dimensions and (2) investigating the effect of confirmed task manipulations on students' language use. In doing so, researchers will be able to better understand how to gain intended task manipulation effects and interpret their findings in more flexible and accurate ways.

Additionally, in the field of TBLT research, there has been a tendency to explore changes in traditional linguistic complexity and accuracy measures in an attempt to infer the cognitive demands of different tasks. This trend in task-based studies might have come from their focus on the validation of competing cognition hypotheses (Robinson, 2005, 2007; Skehan, 1998, 2009). However, by exploring a comprehensive range of linguistic features at different levels, this study identified some findings that had gone unnoticed in previous research. For example, I found how hedging expressions differ across the idea support conditions, and how various cohesion markers and connectives vary across the two genres. More interestingly, the present finding indicated how important it is to employ more fine-grained dependent clauses as target measures, instead of traditional subordination ratio measures, in identifying more specific patterns across task types. Therefore, I recommend that future research into task manipulation effects on linguistic features need to explore linguistic features at various levels to obtain a more comprehensive picture of how some task features elicit different language use and promote development.

While this study showed confirmatory evidence for the function of supporting ideas in reducing cognitive pressure, there is still an issue of how specific such supporting ideas should be in a prompt. On this point, Huot (1990) argued that a moderate level of specificity that clearly informs audience and purpose would greatly benefit students' writing production, while Brossell (1983) and Smith et al. (1985) suggested that there is the potential that writing prompts with too

specific information will cause adverse effects on students' writing. Similarly, some teacher participants in the present study cautioned that too much supporting ideas could deprive students of the opportunity to develop their own ideas. As a next step, future research can explore how different levels of specificity and amount in supporting ideas exert different effects on students' perceptions and language production.

### **Pedagogical and Assessment Implications**

This study offers implications for L2 writing pedagogy that generally involve how teachers need to understand and implement different genres in L2 writing classes. Considering the present finding that revealed a wide gap between teachers' and students' task perceptions, I suggest that it is important for teachers to have a better awareness of potential genre effects on students' task perceptions and language production. For this purpose, teachers may need some training to increase their knowledge of how various areas of task features create different outcomes. As a result of such training, they will be able to design and select writing tasks appropriate for their students. In the case of choosing target tasks, while considering students' L2 proficiency as a primary factor, teachers also should take into account students' task interest or motivation because such motivational variables have been found to influence students' writing performance (e.g., Graham et al., 2007; Zimmerman & Bandura, 1994). One way of achieving this goal is to conduct task-based needs analysis at the beginning of a semester (Long, 2015) and then select a range of target tasks that will be covered over the course of the semester.

Furthermore, due to a widespread belief that argumentative writing, a cognitively challenging task, is most suitable for testing purposes, L2 writing teachers tend to focus on developing students' skills for argumentative writing; accordingly, they have paid relatively less attention to other genres such as narrative or descriptive writing. Similarly, it is likely that

teachers assume that they do not need any more instruction for narrative writing when their students show sufficient skills for argumentative writing because of their conception of narrative as a simpler task than argumentative writing. However, based on the finding of this study, I argue that teachers should not make an a priori decision on how tasks will work and what to include in a curriculum. Interestingly, several parts of the present findings pointed to the need for giving greater instructional emphasis on narrative writing. First, it was found that ESL students tend to see the narrative genre as more interesting and motivating than the argumentative. Aside from the cultural or affective benefits of narrative tasks (e.g., Berman & Slobin, 1994; Kang, 2005; Zhang, 2013), this study also suggested an additional justification for the inclusion of narrative writing in the ESL classroom, namely, the lack of schemas for effective narrative writing.

One of the unexpected findings of this study was a significant interaction between genre and idea support on discourse-level writing scores. This result in fact arose from students' significantly lower narrative scores when they were given the prompt with supporting ideas (see Table 17). As discussed above, the most probable explanation for this finding is supporting ideas' unexpected restrictions on the scope of personal stories that need to be used for interesting narrative construction. An important implication of this finding is that the provision of supporting ideas for a personal narrative task should be avoided in order to give students opportunities to better learn how to turn their experience into a well-organized narrative essay. For lower-level students who need additional support, teachers can instead provide a list of relevant particle verbs that students can use for narrative writing, while offering some idea support for argumentative writing.

Based on their previous test-taking experience, ESL and EFL students are likely to expect that they will be given argumentative tasks in the context of standardized writing assessment. In fact, despite some attempts to implement multiple writing tasks in one language test, there is still a tendency to rely on a single task of argumentation in various proficiency and placement test settings, mostly for practical reasons. However, different linguistic features and task performances across different genres have informed us that test developers need to provide at least more than one genre to obtain a more comprehensive picture of test-takers' writing proficiency. Similarly, calling for the necessity of targeting multiple genres (or modes of discourse in her study), Kegley (1986) argued "the practitioner should be cognizant of the limitations of using a single mode of discourse for making decisions about overall student writing competency for either groups of, or individual, students" (p. 154).

While following this suggestion may cause some concerns related to the constraints of time and cost (e.g., more time for test implementation and increased cost for rating), test developers can avail themselves of an automated language processing technology that has gone through much advancement over the past decade. While the scores produced by such automated systems may not fully reflect the complexity of writing proficiency, they can be used with scores from human raters. In this process of incorporating computational techniques into essay scoring, it would be extremely important for researchers and test developers to have a clear and ever-evolving understanding of the variations in linguistic, discourse, and metadiscourse features across written genres (and even across sub-genres) to obtain valid and reliable scores.

### **Limitations and Future Research**

In this study, I explored genre and task complexity effects on students' perceptions and language production systematically and provided meaningful suggestions on how researchers

and teachers need to understand genre and idea support as distinct task variables. Nevertheless, there are several limitations that need to be addressed in future research in order to further advance this line of research. First, the target of this study was limited to the independent writing task under the time constraint of 30 minutes. Although this reflects a strictly controlled design of the present study, given the increasing trend of integrating other skills materials in assessing writing (Plakans, 2010; Plakans & Gebril, 2013), the exploration of L2 learners' performance across genres in the format of integrated writing will offer valuable information on more authentic writing skills.

Additionally, the student participants of this study were the ESL students enrolled in high-level courses at the English language program. We can expect that L2 learners at this proficiency level might have acquired sufficient genre awareness, leading to clear genre effects on language production. Although I attempted to examine the relationship between L2 proficiency and task type effects, I acknowledge that dividing the student participants into two groups based on their cloze test scores might have resulted in reduced power (Plonsky & Oswald, in press) and that the gap between the two proficiency groups was not large enough to confirm the generalizability of the findings to lower proficiency students. Therefore, future research needs to be followed in order to test how beginning-level students' perceptions and production of the writing tasks differ from the current findings from high intermediate students, offering a more complete picture of task type effects in written discourse.

Finally, in an attempt to control for topic effects, I used the shared topic of foreign language use for all writing tasks targeted in this study. While it was a proper decision to design and use such an approachable topic, it might also be the case that this topic is quite common to many L2 learners, and there is a possibility that some of the participants might have experienced

a similar writing prompt before. For example, in their various projects of L1 genre differences, Berman and her colleagues have used interpersonal conflict as a shared topic (e.g., Berman, 2008; Berman & Katzenberger, 2004; Berman & Nir-Sagiv, 2004, 2007), which can be considered less common than foreign language use for many adult L2 learners; using such topics might have elicited somewhat different patterns. Therefore, exploring similar research questions using a different (less common and more challenging) topic will provide information on whether the present findings are generalizable to uncommon or complex topics.

## APPENDICES

## **Appendix A.**

### **Writing Prompts**

#### **Argumentative 1 (Arg/-Support)**

Situation:

You attended a seminar and the main theme was that using a foreign language fluently has become necessary in this globalized era.

Writing task:

Write an essay about whether you agree or disagree with the statement about the necessity of foreign language abilities. Support your position with reasons. Be sure to fully develop your essay by including clear explanations and logical supporting ideas.

#### **Argumentative 2 (Arg/+Support)**

Situation:

You attended a seminar and the main theme was that the ability to speak a foreign language raises the possibility of having a successful life.

Writing task:

Write an essay about whether you agree or disagree with the statement about the relationship between foreign language abilities and success. Support your position using the reasons provided below. Be sure to fully develop your essay by including clear explanations and logical supporting ideas.

Agree/Support to argue for the position

- Better understanding of cultural differences and other ethnic groups
- Greater job opportunities related to international business
- Possibilities for fun activities such as traveling or watching foreign TV programs

Disagree/Support to argue against the position

- Other qualities (such as self-confidence) more important than foreign language skills
- Foreign language skills not necessary for many great jobs
- A huge investment of time and effort for language learning that could be used for other skill development



### **Narrative1 (Nar/-Support)**

Situation:

Your friend has plans to learn a foreign language but is afraid it might be useless to spend the time learning a language. You have successfully learned a foreign language and use it often. You want to show your friend that language learning and use can be interesting by telling him/her about your positive experience.

Writing task:

Tell a story about ONE of your positive experiences related to foreign language use. Be sure to fully develop your story by including specific details.

### **Narrative 2 (Nar/+Support)**

Situation:

Your friend is planning a trip to a foreign country. While excited about this trip, your friend is worried about how to communicate with people using a foreign language. You have greater foreign language experience, so your friend wants to know some of the possible difficulties she may have while interacting with foreigners.

Writing task:

Tell a story about ONE of your difficult experiences related to interactions using a foreign language. When developing your ideas, you can refer to the storylines below and use any of them to facilitate your writing. Be sure to fully develop your story by including specific details.

Example storylines

- You visited a public place in a foreign country. When you were talking to a foreigner, he/she corrected your language constantly, making you feel offended. Then...
- You were talking to a foreigner. While interacting with him/her, you experienced some cultural differences that made you feel uncomfortable. Then...
- You had to fix a problem or sign a contract using a foreign language. For such purposes, you expressed your ideas to a native speaker of the language, but it caused a misunderstanding, leading to a serious accident. Then...

## Appendix B.

### Revised Analytic Scoring Rubric

	Content		Organization		Vocabulary		Language Use	Score /2	Mechanics
20	Thorough and logical development of thesis Substantive and detailed No irrelevant information Interesting A substantial number of words for amount of time given	20	Excellent overall organization Clear thesis statement Substantive introduction and conclusion Excellent use of transition word	20	Very sophisticated vocabulary Excellent choice of words with no errors Excellent range of vocabulary Idiomatic and near native-like vocabulary	20	No major errors in word order or complex structures No errors that interfere with comprehension Only occasional errors in morphology	20	Appropriate layout with indented paragraphs No spelling errors No punctuation errors
16		16	Excellent connections between paragraphs Unity within every paragraph	16	Academic register	16	Frequent use of complex sentences Excellent sentence variety	16	
15	Good and logical development of thesis Fairly substantive and detailed Almost no irrelevant information Somewhat interesting An adequate number of words for the amount of time given	15	Good overall organization Clear thesis statement Good introduction and conclusion Good use of transition words Good connections between paragraphs Unity within most paragraphs	15	Somewhat sophisticated vocabulary Attempts, even if not completely successful, at sophisticated vocabulary Good choice of words with some errors that don't obscure meaning	15	Occasional errors in awkward order or complex structures Almost no errors that interfere with comprehension Attempts, even if not completely successful, at a variety of complex structures	15	Appropriate layout with indented paragraphs No more than a few spelling errors in less frequent vocabulary No more than a few punctuation errors
11		11		11	Adequate range of vocabulary but some repetition Approaching academic register	11	Some errors in morphology Frequent use of complex sentences Good sentence variety	11	
10	Some development of thesis Not much substance or detail Some irrelevant information Somewhat uninteresting Limited number of words for the amount of time given	10	Some general coherent organization Minimal thesis statement or main idea Minimal introduction and conclusion Occasional use of transitions words	10	Unsophisticated vocabulary Limited word choice with some errors obscuring meaning Repetitive choice of words No resemblance to academic register	10	Errors in word order or complex structures Some errors that interfere with comprehension Frequent errors in morphology Minimal use of complex sentences	10	Appropriate layout with most paragraphs indented Some spelling errors in less frequent and more frequent vocabulary Several punctuation errors
6		6	Some disjointed connections between paragraphs Some paragraphs may lack unity	6		6	Little sentence variety	6	
5	No development of thesis No substance or details Substantial amount of irrelevant information Completely uninteresting Very few words for the amount of time given	5	No coherent organization No thesis statement or main idea No introduction and conclusion No use of transition words	5	Very simple vocabulary Severe errors in word choice that often obscure meaning No variety in word choice No resemblance to academic register	5	Serious errors in word order or complex structures Frequent errors that interfere with comprehension Many error in morphology Almost no attempt at complex sentences	5	No attempt to arrange essay into paragraphs Several spelling errors even in frequent vocabulary Many punctuation errors
0		0	Disjointed connections between paragraphs Paragraphs lack unity	0		0	No sentence variety	0	

## Appendix C.

### Cloze Test

Name: \_\_\_\_\_

Class: \_\_\_\_\_

#### DIRECTIONS:

1. Read the passage quickly to get the general meaning.
2. Write only **one word in each blank** in the column to the right. Contractions (e.g., can't) are considered one word.
3. Check your answers.

NOTE: Spelling will not count against you as long as the scorer can read the word.

#### EXAMPLE:

I met my friend who took a final exam yesterday.

He told me that he is satisfied \_\_\_\_\_ his performance.

Answer: *with*

You have 30 minutes to complete the cloze test.

#### MAN AND HIS PROGRESS

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. \_\_\_\_1\_\_\_\_ ever restless brain has used the \_\_\_\_2\_\_\_\_ and the wisdom of his ancestors \_\_\_\_3\_\_\_\_ improve his way of life. Since \_\_\_\_4\_\_\_\_ is able to walk and run \_\_\_\_5\_\_\_\_ his feet, his hands have always \_\_\_\_6\_\_\_\_ free to carry and to use \_\_\_\_7\_\_\_\_. Man's hands have served him well \_\_\_\_8\_\_\_\_ his life on earth. His development, \_\_\_\_9\_\_\_\_ can be divided into three major \_\_\_\_10\_\_\_\_, is marked by several different ways \_\_\_\_11\_\_\_\_ life.

Up to 10,000 years ago, \_\_\_\_12\_\_\_\_ human beings lived by hunting and \_\_\_\_13\_\_\_\_. They also picked berries and fruits, \_\_\_\_14\_\_\_\_ dug for various edible roots. Most \_\_\_\_15\_\_\_\_, the men were the hunters, and \_\_\_\_16\_\_\_\_ women acted as food gatherers. Since \_\_\_\_17\_\_\_\_ women were busy with the children, \_\_\_\_18\_\_\_\_ men handled the tools. In a \_\_\_\_19\_\_\_\_ hand, a dead branch became a \_\_\_\_20\_\_\_\_ to knock down fruit or to \_\_\_\_21\_\_\_\_ for tasty roots. Sometimes, an animal \_\_\_\_22\_\_\_\_ served as a club, and a \_\_\_\_23\_\_\_\_ piece of stone, fitting comfortably into \_\_\_\_24\_\_\_\_ hand, could be used to break \_\_\_\_25\_\_\_\_ or to throw at an animal. \_\_\_\_26\_\_\_\_ stone was chipped against another until \_\_\_\_27\_\_\_\_ had a sharp edge. The primitive \_\_\_\_28\_\_\_\_ who first thought of putting a \_\_\_\_29\_\_\_\_ stone at the end of a \_\_\_\_30\_\_\_\_ made a brilliant discovery: he \_\_\_\_31\_\_\_\_ joined two things to make a \_\_\_\_32\_\_\_\_

useful tool, the spear. Flint, found \_\_\_\_33\_\_\_\_ many rocks, became a common cutting \_\_\_\_34\_\_\_\_ in the Paleolithic period of man's \_\_\_\_35\_\_\_\_. Since no wood or bone tools \_\_\_\_36\_\_\_\_ survived, we know of this man \_\_\_\_37\_\_\_\_ his stone implements, with which he \_\_\_\_38\_\_\_\_ kill animals, cut up the meat, \_\_\_\_39\_\_\_\_ scrape the skins, as well as \_\_\_\_40\_\_\_\_ pictures on the walls of the \_\_\_\_41\_\_\_\_ where he lived during the winter.

\_\_\_\_42\_\_\_\_ the warmer seasons, man wandered on \_\_\_\_43\_\_\_\_ steppes of Europe without a fixed \_\_\_\_44\_\_\_\_, always foraging for food. Perhaps the \_\_\_\_45\_\_\_\_ carried nuts and berries in shells \_\_\_\_46\_\_\_\_ skins or even in light, woven \_\_\_\_47\_\_\_\_. Wherever they camped, the primitive people \_\_\_\_48\_\_\_\_ fires by striking flint for sparks \_\_\_\_49\_\_\_\_ using dried seeds, moss, and rotten \_\_\_\_50\_\_\_\_ for tinder. With fires that he kindled himself, man could keep wild animals away and could cook those that he killed, as well as provide warmth and light for himself.

### Cloze Test Answers

Exact answer	Acceptable answers
1 his	man's, our, the
2 knowledge	accomplishments, culture, cunning, examples, experience(s), hands, ideas, information, ingenuity, instinct, intelligence, mistakes, nature, power, skill(s), talent, teaching, technique, thought, will, wit, words, work
3 to	
4 man	he
5 on	upon, using, with
6 been	felt, hung, remained
7 tools	adequately, carefully, conventionally, creatively, diligently, efficiently, freely, implements, objects, productively, readily, them, things, weapons
8 during	all, for, improving, in, through, throughout, with
9 which	also, basically, conveniently, easily, historically, however, often, since, that, thus
10 periods	areas, categories, divisions, eras, facets, groups, parts, phrases, sections, stages, steps, topics, trends

11	of	for, in, through, towards
12	all	early, hungry, many, most, only, primitive, the, these
13	fishing	farming, foraging, gathering, killing, scavenging, scrounging, sleeping, trapping
14	and	or, often, some, they
15	often	always, emphatically, important, nights, normally, of, times, tribes
16	the	all, house, many, most, older, their, younger
17	the	all, many, married, most, often, older, primate, these
18	the	all, constructive, many, most, older, primate, tough, younger
19	man's	able, big, closed, coordinated, creative, deft, empty, free, human('s), hunter's, learned, needed, needy, person's, right, single, skilled, skillful, small, strong, trained
20	tool	club, device, instrument, pole, rod, spear, stick, weapon
21	dig	burrow, excavate, probe, search, test
22	bone	arm, easily, foot, head, hide, horn, leg, skull, tail, tusk
23	sharp	big, chipped, fashioned, flat, hard, heavy, large, rough, round, shaped, sizeable, small, smooth, soft, solid, strong, thin
24	the	a, his, man's, one('s)
25	nuts	apart, bark, bones, branches, coconuts, down, firewood, food, fruit, heads, ice, items, meat, objects, open, rocks, shells, sticks, stone, things, tinder, wood
26	one	a, each, flat, flint, glass, hard, obsidian, shale, softer, some, the, then, this
27	it	each, one, they
28	man	being, creature, human, hunter, men, owner, people, person
29	sharp	glass, hard, jagged, large, lime, pointed, sharpened, small
30	stick	bone, branch, club, log, pole, rod, shaft
31	had	accidentally, cleverly, clumsily, conveniently, creatively, dexterously, double, easily, first, ingeniously, securely, simply, soon, suddenly, tastefully, then, tightly
32	very	bad, extremely, good, hunter's, incredibly, intelligent, long, modern, most, necessarily, new, portentously, quite, really, tremendously
33	in	all, among, amongst, by, inside, on, that, using, within

34	tool	device, edge, implement, instrument, item, material, method, object, piece, practice, stone, utensil age, ancestry, discoveries, era, evolution, existence, exploration, history,
35	development	life, time
36	have	actually, apparently, ever
37	by	and, for, from, had, made, through, used, using
38	could	did, would
39	and	carefully, help, or, skillfully, then, would
40	draw	carve, create, drawing, engrave, hang, paint, painting, place, sketch, some, the
41	cave(s)	animals, place(s), room
42	in	and, during, with
43	the	across, aimless, all, barren, dry, flat, high, in, long, many, plain, stone, through, to, toward, unknown, various
44	home	appetite, camp, course, destination, destiny, diet, direction, domain, foundation, habitat, income, knowledge, location, lunch, map, meal, path, pattern, place, plan, route, supplement, supply, time, weapon
45	women	
46	or	and, animal, animal's, covered, in, like, of, on, their, using, with
47	baskets	bags, blankets, chests, cloth(s), clothes, fabric, garments, hides, material, nets, pouches, sacks
48	made	began, built, lighted, lit, produced, set, started, used
49	and	also, by, occasionally, or, then, together, while
50	wood	bark, branches, dung, forage, grass, leaves, lumber, roots, skin, timber, tree(s)

## Appendix D.

### Example Essays

For greater clarity, I corrected all spelling errors contained in the example essays and included an indentation at the beginning of each paragraph. I did not change any grammatical or lexical errors.

#### D1. Two essays for the analysis of complex nominals

For the following two essays (composed by S4), I underlined complex nominals based on the scheme used for Polio & Yoon, in preparation (e.g., noun phrases with multiple premodifiers, noun phrases with postmodifiers, noun clauses, and infinitives and gerunds in the subject position).

- Arg/+Support (Participant ID: S4)

*Recent years, we had witnessed the rapid globalization in our world. Naturally, we have increasing chance to use a foreign language. And the ability to speak a foreign language has increasingly close relation with the possibility of having a successful life, because an adequate skills of a foreign language is beneficial to enlarge our social network and get more good chances for future career.*

*Admittedly, some people who never need to go abroad think the ability to speak a foreign has nothing to do with the possibility of having a successful life. And getting the skills of a foreign language would waste a lot of time. However, it is viewed from another angle, getting a foreign language represents a general trend. If you want avoid to lose chances in the future, you need to handle a foreign language.*

*The chief reason to support my idea is that an adequate foreign language is beneficial to enlarge social network. It's very common for student who study on abroad that the living level depends on the language level. In this society, the social network is very important for having a successful life. Taking my own example, I have good level of English. So I can find many internships in MSU, which are very useful for me to know many brilliant students and to enlarge social network. Hence, that can lay a fundament for my future career.*

*The another reason that should be take into account is that it is good for getting more good chances in your future career. For instant, my sister, who graduated from UCLA, is a senior manager in IBM. She can get this opportunity because of her good English skill.*

- Nar/+Support (Participant ID: S4)

*It is very common for foreign language user to have some difficult experiences related to interactions. Of course, I also have a very embarrassing experience about making communication with English speaker. This experience is still vivid. And I never forget it in my life. In the following, I would like to share my embarrassing story to you and also want you not to worry too much.*

*About two month ago, I was in the airplane from Beijing to Detroit. A waitress came to me and said “Sir, would you want something to drink.” I was so happy, because at this time I was extremely thirsty. And I replied that “Sure, I want orange juice. Please add some ice.” Then, I found the waitress was very unhappy. She said “Sir, if you want ass, please add your own ass.” Eventually, I realized that my pronunciation was wrong. That I pronounced a wrong vowel sound led the waitress to misunderstand my meaning. I immediately apologized to this waitress and explained my real meaning. To be honest, I felt really embarrassed in that situation. But at least I corrected a wrong pronunciation.*

*All in all, even though you will face a lot challenge to use foreign language, you are supposed to be brave. Practicing a lot can make you adequate.*

## **D2. Two essays for the analysis of temporal connectives and first person pronouns**

For the following two essays (composed by S45), I put temporal connectives in bold and underlined first person pronouns.

- Arg/+Support (Participant ID: S45)

*Many people are in favor of the idea that speaking a foreign language raises the possibility of being successful. As far as I am concerned, this statement is very reasonable, since ability of speaking a foreign language is a huge advantage and it is significant in many ways.*

***First of all**, speaking a foreign language indicates a better understanding of cultural differences and other ethnic groups. Language is like a key to the gate of communication, **once** you have the ability of communicating, you can chat with people and understand their thoughts. It's easy to live with local people and get used to their culture and lifestyle with the ability of speaking their language.*



*Besides understanding culture, speaking a foreign language has lots of other benefits, for instance, you will be provided a greater job opportunities related to international business. This opportunity is valuable since there are huge markets in other countries. Those who can speak many languages have earned a lot of money from international business.*

*Moreover, by having a good command of a foreign language, you gain more fun from various activities such as traveling or watch foreign TV programs. You can enjoy different kind of view and broaden your horizons. This is a very cool experience that definitely worth a try.*

*Speaking a foreign language is so beneficial that it is almost necessary if you want to be successful in this globalized era. We should attach significance to learning a foreign language and enjoy the great benefits brought by that.*

- Nar/+Support (Participant ID: S45)

*Studying overseas is a wonderful experience. I can see and feel different culture and make foreign friend. However, there can be as many difficulties as the benefits as well. I had many difficulties **when I first** came to America, and I had to confront them. It was really a tough experience for me.*

***One month ago**, I started my new life in America. Everything went well **at first**, and I was quite satisfied with my new circumstance here. The air was clean and fresh, and the sky was pure blue. I can seldom enjoy this kind of environment in my hometown. I was in good mood, and well-prepared to start my study life here, **until** that day I went to my first Mathematics class. I found my classroom easily and took a seat there. I was nervous since I was unfamiliar with the American teaching style, but I was confident too because my mathematics had **always** been very good in China. **When** the professor started talking, I was astonished that he spoke too fast for me to follow. I couldn't even understand what the homework assignments were. All my confidence were destroyed and I felt self-abashed. The professor was nice and humorous, but I just couldn't understand the jokes. I was worried about my future here and I was really stressed.*

***After** spending a month studying here, I finally get used to the speed that my professor talks. It was a tough time **at first**, but **once** you make up your mind to confront it to overcome it, nothing will stop you, and you will be fine **at last**. So don't be nervous and afraid, my friend, there will be difficulties, but that's not a big deal. It's better to get prepared for the vocabulary **before** you go abroad. That will make you feel more comfortable.*

### D3. Two essays for the analysis of dependent clauses

For the following two essays (composed by S47), I put nominal clauses in bold, double underlined adverbial clauses, and underlined adjective clauses.

- Arg/+Support (Participant ID: 47)

*Mastering another language can make people successful. I agree with the statement that being capable of another language can make people successful.*

*With the globalization in Asia, a increasingly amount of countries are seeking the opportunities of cooperating with China, so the people who have the ability to speak other languages have more chances to participate in international events. In the meantime, the rise of international companies gives people more job opportunities, and most of the jobs they provide a relatively high income.*

*Maybe money is the common standard of a successful life. Being able to speak other languages, however, can give people more benefits than just material life. Learning another language let people know **what is like in another side of the globe**, they can also learn more about cultural differences. In the earlier period of Qing dynasty, China refuse to communicate with other countries, and that led to a severe consequence, which is he left over in education, technology and so on.*

*Therefore, language is the key to another world. On one hand, learning language can give you multiple ways to perceive. It can also help you have a better understand of other countries' culture. On the other hand, you travel experience can be fantastic if you can understand the language that the country use.*

*In conclusion, mastering other languages can give people much more amazing experiences than they ever have, the tendency of globalization make it like a requirement if you want to be successful.*

- Nar/+Support (Participant ID: 47)

*Two years ago, I have gone to the South Korea to visit a friend, but the interesting thing is **that I can't say a single Korean word**. Also my phone couldn't work in there. So this trip is more like an adventure and a really amazing one.*

*I remember **that I tried to ask somebody for the right path by using English**, because my friend said it's okay to say English to them, they'll understand. But soon I found out **that my biggest issue is not speaking correct English to them**, but I can't understand **what they reply in English**. Then I had to read their gesture, and a nice lady even used electronic dictionary in her phone to translate her word into English. Fortunately, most of them can understand **what you said in English**. All I have to do is that to get used to their Korean-style English, and I did it.*

*Since you're going to have a trip to another country, I personally think **what kind of language** they can speak is a very important thing you need to know, such as Dubai, Korea, or most countries in Europe, most of them can understand English and even can talk to you in English. In this case, you don't have to worry too much about language. In the mean time, you have to know **people may speak English**, but they have accent, like Japan or India, so make sure you prepare for this. You even can install an app in your cellphone for translating.*

*Another thing is **that respect their cultures**, there's different manners in different countries, you can search that online to make sure **you won't be too rude**.*

*Finally, I hope **you enjoy it**, travelling to another country is really a great experience.*

## REFERENCES

## REFERENCES

- Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication, 31*, 151–183.
- Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Journal, 100*, 320–240.
- Beauvais, C., Olive, T., & Passerault, J. M. (2011). Why are some texts good and others not? Relationship between text quality and management of the writing processes. *Journal of Educational Psychology, 103*, 415–428.
- Beers, S., & Nagy, W. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing, 22*, 185–200.
- Beers, S., & Nagy, W. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing, 24*, 183–202.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.
- Berman, R. A. (2008). The psycholinguistics of developing text construction. *Journal of Child Language, 35*, 735–771.
- Berman, R. A., & Katzenberger, I. (2004). Form and function in introducing narrative and expository texts: A developmental perspective. *Discourse Processes, 38*, 57–94.
- Berman, R. A., & Nir-Sagiv, B. (2004). Linguistic indicators of inter-genre differentiation in later language development. *Journal of Child Language, 31*, 339–380.
- Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes 43*, 79–120.
- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale: Erlbaum.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.
- Biber, D. (2006a). *A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D. (2006b). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*, 97–116.

- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge, UK: Cambridge University Press.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9, 2–20.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45, 5–35.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Longman.
- Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, 31, 83–100.
- Brossell, G. (1983). Rhetorical specification in essay topics. *College English*, 45, 165–173.
- Brown, J. D. (1978). *Correlational study of four methods for scoring cloze tests*. MA Thesis, University of California at Los Angeles.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, 64, 311–317.
- Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. *Written Communications*, 8, 533–556.
- Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181–202). Cambridge: Cambridge University Press.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42–65.
- Butler, Y.G., & Iino, M. (2005). Current Japanese reforms in English language education: the 2003 ‘Action Plan’. *Language Policy*, 4, 25–45.
- Byun, K., Chu, H., Kim, M., Park, I., Kim, S., & Jung, J. (2011). English-medium teaching in Korean higher education: Policy debates and reality. *Higher Education*, 62, 431–449.
- Common Core State Standards (CCSS). (2017). *English language arts standards*. Retrieved from <http://www.corestandards.org/ELA-Literacy/>
- Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35–54). Norwood, NJ: Ablex.
- Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25,

- Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18*, 80–98.
- Chow, A. W., & Mok-Cheung, A. H. (2004). English language teaching in Hong Kong SAR: Tradition, translation and transformation. In W. K. Ho & R. Y. L. Wong (Eds.), *English language teaching in East Asia today* (pp. 150–177). Singapore: Eastern Universities Press.
- Christie, F. (1997). Curriculum macrogenres as forms of initiation into a culture. In F. Christie & J. R. Martin (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp. 134–160). New York, NY: Continuum.
- Connor-Linton, J., & Polio, C. (2014). Comparing perspectives on L2 writing: Multiple analyses of a common corpus. *Journal of Second Language Writing, 26*, 1–9.
- Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System, 41*, 965–981.
- Crossley, S. A., Kyle, C., & McNamara, D. S. (in press). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*.
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*, 115–135.
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*, 66–79.
- Crossley, S. A., Salsbury, T., McNamara, D. S. & Jarvis, S. (2010). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing, 28*(4), 561–580.
- Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language, 26*, 92–113.
- Crowhurst, M. (1980). Syntactic complexity and teachers' quality ratings of narrations and arguments. *Research in the Teaching of English, 14*, 223–231.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies on Second Language Acquisition, 26*, 59–84.
- Engelhard, G., Gordon, B., & Gabrielson, S. (1992). The influences of mode of discourse, experiential demand, and gender on the quality of student writing. *Research in the*

*Teaching of English*, 26, 315–336.

- Foltz, P. W. (2007). Discourse coherence and LSA. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 167–184). Mahwah, NJ: Lawrence Erlbaum.
- Fotos, S. S., (1991). The cloze test as an integrative measure of EFL proficiency: A substitute for essays on college entrance examinations? *Language Learning*, 41, 313–336.
- Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing*, 30, 45–57.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum. □
- Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching*, 45, 215–240.
- Ginsburg, H. P., & Opper, S. (1988). *Piaget's theory of intellectual development* (3<sup>rd</sup> ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing*. New York: Longman.
- Graham, S., Berninger, V. W., & Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary Educational Psychology*, 32, 516–536.
- Halliday, M. A. K. (1993). Some grammatical problems in scientific English. In M. A. K. Halliday & J. R. Martin (Eds.), *Writing science* (pp. 2–21). London: The Falmer Press.
- Halliday, M. A. K., & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition*. London: Cassell.
- Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication*, 17, 307–352.
- Hayes, J. R. (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. R., & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, 23, 135–149.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.



- Hickmann, M. (2003). *Children's discourse: Person, space, and time across languages*. Cambridge: Cambridge University Press.
- Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah, NJ: Lawrence Erlbaum.
- Hinofotis, F. B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. In J. W. Oller, Jr., & K. Perkins (Eds.), *Research in language testing* (pp. 121–128). Rowley, MA: Newbury House.
- Hong, H., & Cao, F. (2014). Interactional metadiscourse in young EFL learner writing: A corpus-based study. *Interactional Journal of Corpus Linguistics*, 19, 201–224.
- 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: Investigating complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam/Philadelphia: John Benjamins.
- Huot, B. (1990). Literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60, 237–263.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7, 173–192.
- Hyland, K. (2008). Disciplinary voices: Interactions in research writing. *English Text Construction*, 1, 5–22.
- Institute of International Education (IIE). (2016). *Open doors 2016*. Retrieved from <http://www.iie.org/Research-and-Publications/Open-Doors/Data/Fast-Facts#.WKDq0rYrJo4>
- Ishikawa, T. (2007). The effect of manipulating task complexity along the [+/- Here- and-Now] dimension of L2 written narrative discourse. In G. M. M. del Pilar (Ed.), *Investigating tasks in formal language learning* (pp. 136–156). Clevedon, UK: Multilingual Matters.
- Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, 63, 330–367.
- Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, 12, 377–403.
- Jeffery, J. V. (2009). Construct of writing proficiency in U.S. state and national writing assessments: Exploring variability. *Assessing Writing*, 14, 3–24.
- Jeon, M. (2009). Globalization and native English speakers in English programme in Korea (EPIK). *Language, Culture and Curriculum*, 22, 231–243.

- Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing Writing*, 31, 113–125.
- Johansson, R., Wengelin, Å., Johansson, V., & Holmqvist K., (2010). Looking at the keyboard or the monitor: Relationship with text production processes. *Reading and Writing*, 23, 835–851.
- Johns, A. M. (1995). Teaching classroom and authentic genres: Initiating students into academic cultures and discourses. In D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 277–293). Norwood, NJ: Ablex.
- Johnson, M. D., Mercado, L., & Acevedo, A., (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, 21, 264–282.
- Kang, J. Y. (2005). Written narratives as an index of L2 competence in Korean EFL learners. *Journal of Second Language Writing*, 14, 259–279.
- Kegley, P. H. (1986). The effect of mode discourse on student writing performance: Implications for policy. *Educational Evaluation and Policy Analysis*, 8, 147–154.
- 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. 57–72). Mahwah, NJ: Lawrence Erlbaum.
- Kikuchi, K. (2006). Perspectives: Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 27, 77–96.
- Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, 423–430.
- Knudson, R. E. (1995). Writing experiences, attitudes, and achievement of first to sixth graders. *Journal of Educational Research*, 89, 90–97.
- Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20, 148–161.
- Kormos, J. (2014). Differences across modalities of performance: An investigation of linguistic and discourse complexity in narrative tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 193–216). Amsterdam: John Benjamins.
- Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, 62, 439–472.
- Kuiken, F., Mos, M. & Vedder, I. (2005). Cognitive task complexity and second language writing performance. In S. Foster-Cohen, M.P. García Mayo, & J. Cenoz (Eds.), *Eurosla Yearbook*. Vol. 5 (pp. 195–222). Amsterdam: John Benjamins.

- Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics*, 45, 261–284. □
- Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17, 48–60.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284. □
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16, 307–332.
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30, 358–392.
- Lo, J., & Hyland, F. (2007). Enhancing students' engagement and motivation in writing: The case of primary students in Hong Kong. *Journal of Second Language Writing*, 16, 219–237.
- Long, M. (2015). *Second language acquisition and task-based language teaching*. Oxford, UK: John Wiley & Sons.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Philadelphia: Multilingual Matters.
- Lu, X. (2010). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics*, 14, 3–28.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45, 36–62.
- Malicka, A., & Levkina, M. (2012). Measuring task complexity: Does L2 proficiency matter? In A. Shehadeh & C. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 43–66). Amsterdam: John Benjamins.
- Malvern, D. D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, UK: Palgrave Macmillan.
- Manchón, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing: A study of foreign language writers. *Language Learning*, 57, 549–593.
- Matsuda, P. K. (2015). Identity in written discourse. *Annual Review of Applied Linguistics*, 35, 140–159.
- Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, 29, 3–15.
- McCarthy, P. M., & Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of

- sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42, 381–392.
- McCutchen, D., & Perfetti, C. A. (1982). Coherence and connectedness in the development of discourse production. *Text-Interdisciplinary Journal for the Study of Discourse*, 2, 113–140.
- McDonough, K., & Trofimovich, P. (2011). *Using priming methods in second language research*. New York, NY: Routledge.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27, 57–86.
- McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.
- Mei, W. S. (2006). Creating a contrastive rhetorical stance: Investigating the strategy of problematization in students' argumentation. *RELC*, 37, 329–353.
- Melendy, G. A. (2008). Motivating writers: The power of choice. *The Asian EFL Journal*, 10, 187–198.
- Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578.
- Ojima, M. (2006). Concept mapping as pre-task planning: A case study of three Japanese ESL writers. *System*, 34, 566–585.
- Olinghouse, N. G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of Educational Psychology*, 101, 37–50.
- Oller, J. W., & Conrad, C. A. (1971). The cloze technique and ESL proficiency. *Language Learning*, 21, 185–195.
- Ong, J. (2013). Discovery of ideas in second language writing task environment. *System*, 41, 529–542.
- Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers? *Journal of Second Language Writing*, 23, 17–30.
- Ong, J., & Zhang L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, 19, 218–233.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24, 492–518.

- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48–59.
- Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum.
- Plakans, L. (2010). Independent vs. integrated writing tasks: A comparison of task representation. *TESOL Quarterly*, 44, 185–194.
- Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22, 217–230.
- Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, 36, 73–97.
- Plonsky, L., & Oswald, F. L. (in press). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*.
- Polio, C. (2013). Revising a writing rubric based on raters' comments: Does it result in a more reliable and valid assessment? *Midwest Association of Language Testers*, East Lansing, MI.
- Polio, C., & Yoon, H. (2016). Task and genre differences in L2 writing research. Invited colloquium (Colloquium title: Researching written task complexity in diverse contexts organized by Lawrence Zhang) presented at *American Association for Applied Linguistics (AAAL) 2016*, Orlando, FL.
- Polio, C., & Yoon, H. (under review). The use of two automated tools to examine ESL learners' syntactic complexity across two genres. *International Journal of Applied Linguistics* [Special issue: Perspectives and challenges for research on grammatical complexity in SLA: The case of variation].
- Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, 38, 444–456.
- Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, 33, 3–17.
- Quinlan, T., Loncke, M., Leijten, M., & Van Waes, L. (2012). Coordinating the cognitive processes of writing: The role of the Monitor. *Written Communication*, 29, 345–368.
- Ravid, D. (2005). Emergence of linguistic complexity in later language development: Evidence from expository text construction. In D. Ravid & H. B. Shyldkrot (Eds.), *Perspectives on language and language development: Essays in honor of Ruth A. Berman* (pp. 337–356). London: Kluwer Academic.

- Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, 31, 437–470.
- Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics*, 35, 87–92.
- Révész, A., Kourтали, N., & Mazgutova, D. (in press). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*.
- Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition*, 38, 703–737.
- Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning*, 64, 615–650.
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15, 18–39.
- Robinson, P. (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57.
- Robinson, P. (2001b). 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, UK: Cambridge University Press.
- Robinson, P. (2003). The cognition hypothesis of adult, task-based language learning. *Second Language Studies*, 21, 45–107.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, 43, 1–32.
- Robinson, P. (2007). 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*, 45, 193–213.
- Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Scola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243–268). Amsterdam, The Netherlands: John Benjamins.
- Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, 61 (Suppl. 1), 1–36.

- Ruiz-Funes, M. (2014). Task complexity and linguistic performance in advanced college-level foreign language writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 163–192). Amsterdam: John Benjamins.
- Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing*, 28, 1–19.
- Sakamoto, M. (2012). Moving towards effective English language teaching in Japan: Issues and challenges. *Journal of Multilingual and Multicultural Development*, 33, 409–420.
- Shim, R. J., & Baik, M. J. (2004). English education in South Korea. In W. K. Ho & R. Y. L. Wong (Eds.), *English language teaching in East Asia today* (pp. 241–261). Singapore: Eastern Universities Press.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, 30, 510–532.
- Skehan, P., & Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task based learning. *Language Teaching Research*, 1, 185–211.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge: Cambridge University Press.
- Slobin, D. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (Eds.), *Relating events in narrative, volume 2: Typological and contextual perspectives* (pp. 219–257). Mahwah, NJ: Lawrence Erlbaum.
- Smith, W. L., Hull, G. A., Land, R. E., Moore, M. T., Ball, C., Dunham, D. E., Hickey, L. S., & Ruzich, C. W. (1985). Some effects of varying the structure of a topic on college students' writing. *Written Communication*, 2, 73–89.
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and lexical description, volume 3: Grammatical categories and the lexicon* (pp. 36–149). Cambridge: Cambridge University Press.
- Talmy, L. (2000). *Toward a cognitive semantics, volume 1: Concept structuring systems*. Cambridge: MIT Press.
- Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 217–236). Amsterdam: John Benjamins.
- Tedick, D. J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on

- performance. *English for Specific Purposes*, 9, 123–143.
- Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. *Proceedings of HLT-NAACL 2003*, 252–259.
- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition*, 33, 339–372.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 318–333.
- Way, P., Joiner, E. G., & Seaman, M. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal*, 84, 171–184.
- Wengelin, Å., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods*, 41, 337–351.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.
- Wu, X. (2003). Intrinsic motivation and young language learners: The impact of the classroom environment. *System*, 31, 501–517.
- Yang, W. (2014). *Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of L2 writing*. Doctoral dissertation. Retrieved from: [http://scholarworks.gsu.edu/alesl\\_diss/29](http://scholarworks.gsu.edu/alesl_diss/29)
- Yang, W. Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53–67.
- Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*, 23, 31–48.
- Yoon, H. (2017a). Textual voice elements and voice strength in EFL argumentative writing. *Assessing Writing*, 32, 72–84.
- Yoon, H. (2017b). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, 66, 130–141.



- Yoon, H., & Polio, C. (2017). ESL students' linguistic development in two written genres. *TESOL Quarterly*.
- Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24, 1–27.
- Zhang, L. J. (2013). Second language writing as and for second language learning. *Journal of Second Language Writing*, 22, 446–447.
- Zhao, C. G. (2012). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing*, 30, 201–230.
- Zhao, C. G., & Llosa, L. (2008). Voice in high-stakes L1 academic writing assessment: Implications for L2 writing instruction. *Assessing Writing*, 13, 153–170.
- Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, 31, 845–862.