EXPLORING THE EFFECTS OF TASK COMPLEXITY ON INTERACTION AND L2 PERFORMANCE WITH MATH WORD PROBLEMS IN FTF AND SCMC ENVIRONMENTS

A Dissertation

by

ZIHAN GENG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Zohreh R. Eslami
Committee Members,	Li-Jen Kuo
	Wen Luo
	Yeping Li
Head of Department,	Michael de Miranda

August 2021

Major Subject: Curriculum and Instruction

ABSTRACT

This dissertation investigated the effects of task complexity and modality on interaction and L2 performance in the mathematical content area. The participants included 82 college students, 41 English native speakers (NSs) and 41 English non-native speakers (NNSs). Forty-one NS-NNS dyads were formed and were then assigned to either the FTF group (N = 21) or the SCMC group (N = 20). The experimental tasks were two mathematical word problems, which required the same mathematical knowledge but varied in cognitive demands. Two levels of task complexity were operationalized along [+/- few steps] based on the Triadic Componential Framework. Self-ratings of task difficulty and retrospective duration judgment were adopted as the independent measures of task complexity. Each dyad discussed both word problems in FTF or online setting. After the discussion, the NNSs articulated the problem-solving process. The NS-NNS interaction and NNSs' L2 narration were recorded for data analysis. The interaction was evaluated based on the occurrence of language related episodes (LREs) and self-initiated repairs. NNSs' L2 oral narratives were assessed in terms of accuracy, syntactic complexity, lexical diversity, and fluency.

Results indicated that increasing task complexity did not lead to significant differences in the amount of LREs and self-initiated repairs across interaction modes. Nevertheless, the complex task yielded a higher rate of successful uptake than the simple task in both modes. Significantly more LREs and self-initiated repairs were yielded in the FTF mode than in the SCMC mode, irrespective of the cognitive condition. These findings suggested that both task complexity and modality played critical roles in facilitating noticing during task performance.

Results also showed that increasing task complexity led to greater accuracy in the SCMC mode, with no effect on accuracy in the FTF mode. Syntactic complexity was enhanced along increased cognitive demands in both interaction modes. No significant effects of task complexity on lexical variation across interaction modes. As for fluency, the only significance was found for the general repair fluency in the FTF mode, but not in the SCMC mode. These findings indicated that task complexity and modality could affect learners’ linguistic performance in mathematical problem-solving tasks in some areas but not in others.

ACKNOWLEDGMENTS

The completion of this dissertation is a long journey, which would not have been possible without the help of many people. First and foremost, I would like to express my deepest appreciation to my academic advisor, Dr. Zohreh R. Eslami, who has been guiding me throughout this dissertation project. Her continuous encouragement and support have given me confidence in myself to overcome self-doubts and to become a successful researcher in academia. Furthermore, my sincere gratitude goes to my committee members: Dr. Li-Jen Kuo, Dr. Wen Luo, and Dr. Yeping Li. Their invaluable insights and support have made this an inspiring experience for me.

I would like to thank Dr. Michael de Miranda, Dr. Beverly Irby, and the OGAPS for providing me financial support during my doctoral study. Without their generous support, this dissertation would not have been possible. Thanks also go to Dr. Patrick Slattery, Ms. Kara Bond, and Ms. Ambyr Rios, for their advice and assistance throughout this process. I also want to extend my appreciation to Dr. Monica Neshyba for her extensive support on my data collection and academic writing, and Dr. Matthew Etchells for his personal and professional advice that has inspired me during my journey of knowledge pursuit. Gratitude also goes to Jing Xu, Qihui Cui, and Jingjing Zhang, my best friends in life. Their friendship and encouragement helped me through the toughest time in this long journey.

Finally, thanks to my dear family for their unconditional love, support, and trust. Thank you, dad, for supporting me to pursue my dream overseas. Thank you, mom, for having faith in me and telling me things will be okay. Thank you, my brother and sister, for always being there for me. I could not have done this without all of you.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Zohreh R. Eslami (advisor) and Professor Wen Luo of the Department of Educational Psychology, and Professor Li-Jen Kuo and Professor Yeping Li of the Department of Teaching, Learning and Culture.

All work for the dissertation was completed independently by the student, under the advisement of Professor Zohreh R. Eslami of the Department of Educational Psychology.

Funding Sources

Graduate study was supported by the Dissertation Fellowship from Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Statement of the Problem

Over the past several decades, task-based language teaching (TBLT) has received considerable attention from researchers and educators in the field of second language acquisition (SLA). Unlike traditional language teaching approaches such as present-practice-produce (PPP), which mostly emphasizes separate linguistic items to be mastered (Van den Branden, 2006), TBLT views language as a means of communication and highlights learners' need to achieve effective communication in the target language (Ellis, 2003; Long, 2015). In TBLT, tasks serve as holistic units of communication and thus provide the basis for an entire curriculum. Given the imperative role of tasks in TBLT, how to design tasks that facilitate second language (L2) learning has become of great importance for both researchers and educators in the field of SLA. A variety of frameworks have provided the theoretical foundation for pedagogical task design.

Building upon Krashen's (1982, 1985) Input Hypothesis, which claims that comprehensible input is the only mechanism that induces language acquisition, Long (1983) argues that the best kind of comprehensible input is the one that has been interactionally modified. In his Interaction Hypothesis, Long (1996) proposes that comprehensible input can be enhanced when learners strive to resolve a communication breakdown by employing interactional modifications such as comprehension checks, clarification requests, and confirmation checks. The process of resolving conversational problems has been known as negotiation of meaning. Furthermore, Long asserts that during negotiation of meaning, learners may receive feedback regarding their problematic language, which pushes learners to attend to the linguistic items and reformulate their own utterances (i.e., output modification). As such, a well-designed task should offer opportunities for negotiation of meaning which then promotes L2 development.

The claim that output modification is beneficial for language acquisition is also tied into Swain's (1985) Comprehensible Output (CO) Hypothesis, which proposes that L2 production is necessary for learners to (a) form and test their hypothesis (i.e., interlanguage), (b) receive feedback from the interlocutor (during negotiation of meaning), and (c) pay attention to language gaps (between their interlanguage and L2). Swain contends that learners will not be able to realize the gap between what they want to say and what they are able to say until they are given opportunities to use their language in production.

Closely related to this is the Noticing Hypothesis proposed by Schmidt in the early 1990s. Schmidt (1995) claims that in order for acquisition to occur, learners need to consciously attend to linguistic items while they are communicating. He further argues that learners' noticing of the gap between their interlanguage and the L2 is an essential process in L2 acquisition. "SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be" (Schmidt, 2001, p. 2). Instances of noticing can be driven within learners or externally fostered. Scholars (e.g., Ellis, 2003; Long, 2015) have proposed both proactive and reactive approaches to elicit attention to linguistic features in a meaning-focused environment.

Ellis (2003) suggests the use of focused tasks to enhance noticing (i.e., a proactive approach). Focused tasks are work plans that are designed with the intention of eliciting pre-determined linguistic features during task performance. What distinguishes a focused task from a situational grammar exercise lies in whether learners are informed of the targeted linguistic features to be mastered prior to task performance. In the case of focused tasks, although the task is designed to direct learners' attention to specific linguistic features, the primary focus is still on content.

Long (1988, 2015) introduces a reactive approach – what he has called "focus on form" (FonF) - to direct learners' attention to linguistic features. Long contends that FonF "overtly draws students' attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on meaning or communication" (p. 45–46). In other words, FonF is incidental in nature, and it occurs in response to a communication breakdown. Long further argues that FonF is synchronized with learners' internal syllabus and allows for a slow and gradual process of L2 development.

To summarize, these existing conceptual frameworks have shed light on how to design pedagogical tasks that maximize language acquisition. It is suggested that tasks should be designed with a primary focus on meaning but should also allow for form-meaning mappings. Furthermore, task design features should be manipulated in ways that will promote negotiation of meaning and output modification. These proposals have led to a substantial body of empirical research examining the effects of different task characteristics on L2 learning. Researchers informed by the Interaction Hypothesis are interested to examine how opportunities for meaning negotiation can be provided and aim to identify the task dimensions that foster a higher amount of meaning negotiation (e.g., Kaivanpanah & Miri, 2017; Kim, 2009; Nuevo, 2006; Payant & Kim, 2017; Révész, 2011; Rouhshad et al., 2016; Yilmaz & Granena, 2010; Zabihi, 2020). Researchers interested in the role of output in L2 acquisition suggest that output provides learners with opportunities to notice gaps in their linguistic knowledge and reflect on their own and their interlocutors' use of language (e.g., Eslami & Kung, 2016; Gurzynski-Weiss & Baralt, 2015; Nuevo et al., 2011). Despite the considerable number of studies on TBLT, most of the studies are conducted focusing mainly on learners' basic interpersonal conversation skills (BICS) (Cummins, 2001). Studies examining the effects of task features on the development of cognitive academic language proficiency (CALP) remain to be scarce.

Academic language is associated with subject-matter learning and is commonly used in content-based instruction (CBI). Content-based instruction (CBI) refers to "an approach to language instruction that integrates the presentation of topics or tasks from subject matter classes (e.g., math, social studies) within the context of teaching a second or foreign language" (Crandall & Tucker, 1990, p. 187). The most common version of CBI is content and language integrated learning (CLIL), which was developed as a dual-focused educational approach with an additional language used for the learning of both content and language (Coyle et al., 2010). Recently, language educators and researchers have identified considerable commonalities between CLIL and TBLT, and proposed ways in which the two approaches can be integrated to facilitate content and language learning (García Mayo & Lazaro Ibarrola, 2015; Llinares & Dalton-Puffer, 2015; Lopes, 2020; Ortega, 2015; Shehadeh, 2018). In what follows, I will discuss the interface of CLIL and TBLT, and highlight the importance of examining task features in content-based language learning contexts.

As language teaching approaches, both CLIL and TBLT are based on the premise that language and meaning are inseparable (Ortega, 2015), although "meaning" is conceptualized differently by each approach. In CLIL, "meaning" represents subject-learning, while in TBLT "meaning" refers to experiential and goal-oriented learning. Another shared interest between CLIL and TBLT is that in both approaches, language learning is achieved through interaction and collaboration. Furthermore, learning activities in CLIL are mainly task-based and fulfill the criteria of pedagogical task design in TBLT (Ellis, 2003).

Nevertheless, each approach has its distinct features as well. Ortega (2015) points out three main differences between CLIL and TBLT. First, the two approaches usually target different age groups. While TBLT has a primary focus on college-level learners, CLIL is mostly applied in secondary education. Second, each approach has its own educational context. TBLT programs are mostly implemented in second language contexts, where learners are exposed to the target language outside of the classroom in a variety of ways. In contrast, CLIL programs are thriving in foreign language contexts, where teachers aim to create an input-rich environment that is lacking outside of the classroom. Finally, the goals of educational effectiveness are different. The ultimate goal of TBLT is to transfer learning from pedagogical tasks to authentic tasks in the real world. In CLIL, on the other hand, the goal is to achieve balanced gains in both language learning and content learning.

These divergences, however, do not necessarily signal or imply incompatibility between the two approaches. In fact, researchers and educators have proposed many benefits of integrating pedagogical tasks of TBLT to CLIL programs (Liu, 2019; Lopes, 2020; Lyster, 2015; Ortega, 2005; Shehadeh, 2018). For instance, Lopes (2020) claims that some issues in CLIL, such as learners paying less attention to language accuracy, can be managed by carefully planning tasks. In terms of content learning, Lopes (2020) and Meyer (2010) contend that the frequent negotiation of meaning generated in pedagogical tasks enables learners to process the subject knowledge to a greater extent and thus facilitate content learning. Moreover, at the methodological level, “TBLT provides the scaffolding needed for CLIL classes to strike a balance between the cognitive and linguistic demands” (Lopes, 2020, p. 8).

Despite the advantages of integrating TBLT to CLIL, the links between the two approaches have rarely been empirically explored until the recent few years. *System*’s (2015) special issue (Vol. 54) constitutes the first attempt at investigating research interfaces between CLIL and TBLT (Ortega, 2015). The studies included in this special issue offer empirical evidence regarding the opportunities and challenges of blending the two approaches. For example, Lyster (2015) provides an exemplary illustration of using form-focused tasks to draw learners' attention to language in CLIL contexts in Canada. Lyster conducted two studies, each of which employed a different task-based approach that crossed borders between content areas or target languages. The first study integrated a focus on French grammatical gender across 5th-grade (10-11 years old) French immersion students. The experimental group completed noticing activities in the language arts class, awareness activities in the social studies class, and practice activities in the science class. In contrast, the comparison group took the regular CLIL program without focused tasks. The second study involved literacy tasks with a focus on French derivational morphology. The participants were 2nd-grade (7-8 years old) French immersion students who completed the focused tasks in one language (e.g., French) and then in the other (e.g., English). The comparison group, on the other hand, attended the regular CLIL classes (i.e., French literacy). The results of both studies show that the experimental groups significantly outperformed the comparison groups on the language assessments. Based on the findings, Lyster claims that "Task would be pivotal in such a cross-curricular approach and would be designed to provide purposeful opportunities for strengthening connections between language and content learning" (p. 12). In the meanwhile, he points out two challenges educators may face when they attempt to integrate TBLT to CLIL: (a) language and content integrated task design, and (b) collaboration between teachers from different areas.

Given that pedagogical task design is critical in promoting language and content learning in academic contexts, and this area has yet been largely explored, the present study aims to contribute to this line of research by examining the effects of task characteristics on learner-learner interaction and L2 performance in a mathematical context. Among a variety of task features, the study specifically focuses on cognitive task complexity, which is closely associated with learner cognitive operations during task performance. In the following discussion, I will first review the theoretical underpinnings of cognitive task complexity, and then point out the research gaps in the area.

Research studies that investigate cognitive aspects of task design mainly rely on two theoretical frameworks: The Limited Capacity Hypothesis (Skehan 1998, 2009) and the Cognition Hypothesis (Robinson, 2011). Both hypotheses consider the engagement of cognitive resources during task performance. Nonetheless, the two models make somewhat distinct predictions regarding the effects of cognitive demands on internal processes during task performance.

Skehan's Limited Capacity Hypothesis assumes that learners' attentional capacity is limited, and thus there are trade-off effects between the content and linguistic forms. Additionally, he argues that there are also competitions among different linguistic aspects (i.e., complexity, accuracy, and fluency) of learners' L2 production. Therefore, a cognitively demanding task forces learners to engage much attention for content processing, which leads to little resources for linguistic forms. Furthermore, when these attentional resources are devoted to one prioritized linguistic aspect (e.g., accuracy), the performance of other linguistic aspects (e.g., complexity and fluency) would be negatively affected (Skehan, 2009).

Robinson's Cognition Hypothesis, on the other hand, proposes multiple pools of resources available for learners to draw attention from. He argues that competition only occurs when attentional resources are drawn from the same pool. In this sense, learners can pay simultaneous attention to content and linguistic forms, as well as different linguistic aspects of the target language. Robinson contends that increasing cognitive task complexity (a) promotes meaning negotiation, noticing of interactive corrective feedback, and modified L2 output; and (b) induces greater linguistic accuracy and complexity simultaneously (Robinson, 2011).

The biggest discrepancy between the two frameworks lies in whether learners can pay simultaneous attention to different linguistic aspects of L2 along increased cognitive demands. Despite having a great number of relevant research studies on cognitive task complexity (e.g., Baralt, 2013; Gilabert, 2007a, 2007b; Hardy & Moore, 2004; Ishikawa, 2007; Kim et al., 2015; Révész, 2011; Robinson, 2007b; Solon et al., 2017; Tavakoli, 2009; Yuan & Ellis, 2003), the existing findings fail to converge into a conclusive picture regarding the effects of different cognitive demands on L2 performance and learning. This lack of consensus is due to many factors, including learner factors, such as L2 proficiency and working memory capacity (e.g., Kim et al., 2015; Robinson, 2011), and methodological factors, such as task complexity manipulation and task modality (e.g., Baralt, 2013; Sasayama, 2016). While learner factors concern the generalization of the research findings, methodological factors indicate whether the study is a valid examination of cognitive operations. For task complexity studies, it is imperative to verify construct validity for the independent variables - whether the manipulations indeed induce desired changes in cognitive demands (Révész, 2011). Nevertheless, in many studies amount of task complexity expected by manipulations is assumed to be valid without sufficient evidence. In addition, there seems to be a lack of research on empirical studies that examine the effects of task complexity on L2 learning in content areas (e.g., math, science).

Another area that needs further investigation is task modality. While modality indicates both the interaction environment (i.e., face-to-face or computer-mediated communication), and the format of input and output production (i.e., oral or written), the scope of discussion in this dissertation will be limited to the first dimension. With the advancement of technology and distance learning, computer-assisted language learning (CALL) has gained substantial attention in recent decades. It has been empirically attested that computer-mediated communication (CMC) promotes meaning negotiation, noticing of feedback, and subsequent L2 development (Baralt, 2013; Eslami & Kung, 2016; Parlak & Ziegler, 2017; Payne & Whitney, 2002; Payne & Ross, 2005; Sachs & Suh, 2007; Smith, 2004, 2005, 2009, 2012). Recently, Ziegler (2016) conducted a meta-analysis of studies comparing the effectiveness of interaction in synchronous CMC (SCMC) and face-to-face (FTF) contexts on L2 development. The research included 14 studies that were completed between 1990 and 2012. The findings show some advantages for SCMC on overall L2 development, suggesting promising potential for interactive language learning in SCMC. However, as Ziegler points out, many important constructs could not be analyzed due to the lack of empirical data. Cognitive task complexity, for example, is one of these under-researched areas in the CMC context. Research into the relationship between task complexity and interaction-driven L2 learning, and the role of modality in this relationship will not only advance our understanding of learners' cognitive operations in different interactional environments, but also provide pedagogical implications for educational practices. This dissertation aims to contribute to this line of research.

To summarize, even though there have been a considerable number of studies examining the effects of task characteristics on L2 learning, very few of them has been associated with academic language learning. Recently, researchers and educators have called for the incorporation of pedagogical tasks in content-based language learning programs such as CLIL. Therefore, understanding how task design features influence L2 learning in academic content has become of great importance for both researchers and educators. Since task complexity has been viewed as a feature that contributes to the intrinsic cognitive demands of tasks (Robinson & Gilabert, 2007), and academic language development is also considered to be cognitively demanding (Cummins, 2001), this study aims to illuminate the effects of different cognitive task complexity on how learners engage and allocate their attentional resources while performing content-based tasks (i.e., mathematical tasks). In addition, given that technology is being widely used in education, the study also seeks to investigate the role of task modality (i.e., FTF vs. SCMC) in academic language learning. It is worth noting that this study examines L2 learning from a cognitive-interactionist perspective with the purpose of providing suggestions for optimal features of tasks being implemented in academic content areas. Therefore, the design of the study is different from that of most studies in CLIL, which usually are of a descriptive nature and conducted with intact classrooms. The study is experimental in nature and gets carried out in a laboratory setting, in which the individual learning processes are to be recorded and analyzed in depth.

1.2 The Study

The present study explores the effects of cognitive task complexity and task modality on interaction and L2 development as participants are involved in solving mathematical tasks. Native speaker (NS) - nonnative speaker (NNS) dyads were formed for conducting the tasks. The rationale for forming NS-NNS dyads is that mixed level dyads have been found to be more effective in eliciting meaning negotiation opportunities and corrective feedback, and modified output than equal proficiency dyads (Iwashita, 2001; Mackey et al., 2003; Storch, 2002). Learners have perceived NS-NNS interactions as more beneficial for L2 learning than NNS-NNS interactions (Wang et al., 2015).

In this study, interactions in the dyads were examined to find occurrences of incidental noticing, or incidental focus on form, which has been viewed as a critical step in L2 development (Schmidt, 2001, 2010). According to the Noticing Hypothesis, "input does not become intake for language acquisition unless it is noticed, that is, consciously registered" (Schmidt, 2010, p. 2). In order for language acquisition to occur, learners must notice the difference between the L2 input they are exposed to and their interlanguage (i.e., the language they are producing). Although noticing does not guarantee acquisition, it promotes learners to process linguistic forms in short-term memory, which tends to lead to further encoding in long-term memory.

While noticing is an internal cognitive process and cannot be directly observed (Schmidt, 1993), it can be assessed through collaborative dialogues. Swain and Lapkin (2001) have suggested the use of language-related episodes (LREs) as a measure of incidental noticing of linguistic forms. LREs are short dialogues where learners "talk about the language they are producing, question their language use, or correct themselves or others" (Swain & Lapkin, 1998, p. 326). An LRE usually consists of three discourse moves: trigger, response, and uptake. A *trigger* is an erroneous utterance produced by the learner, which is followed by the interlocutor's *response* to the language problem (i.e., feedback). Learner's immediate response to the feedback is *uptake*. However, uptake is not an essential component of LRE. In real classroom interactions, a response is not always followed by an uptake.

To analyze the LREs, the study adapted Loewen's (2005) analysis of focus-on-form episodes (FFEs), which was defined as "a brief, spontaneous focus on a linguistic item within the context of a meaning-focused task" (p. 363). Based on this model, LREs can include the following features: type, linguistic focus, source, complexity, directness, emphasis, response, uptake, and successful uptake. Among these features, successful uptake has been identified as the strongest predictor of noticing and subsequent L2 learning (Egi, 2010; Loewen, 2005; Mackey et al., 2000; McDonough & Mackey, 2006). In a recent study, Gurzynski-Weiss & Baralt (2015) further examined whether the type of successful uptake (i.e., partial or full modified output) differentially indicates noticing in FTF and SCMC modes. The researchers utilized stimulated recall as the direct measure of noticing (Egi, 2010; Mackey et al., 2000), and related learners' noticing to the modified output type. The results showed that partial modified output (i.e., learners repeated only the item corrected in feedback) better predicted noticing of feedback than full modified output (i.e., learners repeated all feedback) in both FTF and SCMC modes. Based on the findings, the researchers concluded that producing partial modified output indicates "more focused attention allocation to the specific mismatch of interlanguage–target language forms, followed by a unitized encoding" (Gurzynski-Weiss & Baralt, 2015, p. 1410). Full modified output, on the other hand, can also indicate conceptually driven processing of the feedback but it sometimes functions as a social move of acknowledging receipt of the feedback, especially in the SCMC context.

In addition to LREs, instances of learners correcting their own linguistic errors without having triggers by their interlocutors (i.e., self-initiated repairs) are also utilized as a measure of noticing by researchers (Gilabert, 2007b; Lai & Zhao, 2006; Yuan & Ellis, 2003). Although there are self-corrections in LREs, they are mainly driven by interactional moves. In contrast, self-initiated repairs are spontaneously generated by learners as the result of learners' monitoring of their output (Kormos, 1999; Levelt, 1983). Recent research on self-initiated repairs has denoted that self-initiated repairs play an important role in prompting modified output (e.g., Sato & Takatsuka, 2016; Shehadeh, 2001). Following previous research, the present study involves self-initiated repairs as a measure of focus on form in addition to LREs. The present study aims to investigate the effects of cognitive task complexity and task modality on NS-NNS interaction-driven L2 learning opportunities in content-based learning (mathematics). The following research questions will be addressed in Chapter III:

1. How does cognitive task complexity affect interaction-driven L2 learning opportunities, as measured by LREs and self-initiated repairs, during the mathematical problem-solving process in the FTF mode?

2. How does cognitive task complexity affect interaction-driven L2 learning opportunities, as measured by LREs and self-initiated repairs, during the mathematical problem-solving process in text-based online chats?

3. How does task modality (i.e., FTF vs. SCMC) affect interaction-driven L2 learning opportunities, as measured by LREs and self-initiated repairs, across cognitive task complexity?

In addition to interaction, L2 performance is another prevailing area of task complexity research. Here, L2 performance research refers to the examination of linguistic features (i.e., overall complexity, accuracy, and fluency) of the language that learners produce during monologic or dialogic task performance. In general, existing literature supports Cognition Hypothesis claiming that increasing task complexity leads to improved linguistic performance. Nevertheless, whether the manipulation of task complexity promotes accuracy and complexity in L2 performance simultaneously remains inconclusive. Some studies provide empirical evidence showing that there is no competition in attentional resources between the two linguistic dimensions (e.g., Michel et al., 2007; Robinson, 1995). Others, however, detect trade-off effects between accuracy and complexity along increased task complexity (e.g., Gilabert, 2007a; Robinson, 2007; Sasayama & Izumi, 2012). The inconsistencies in findings may partially be due to measures of accuracy and complexity used in research. According to Robinson (2007b), global measures appear to be less successful in detecting task effects than specific measures. Researchers have called for a combination of global and specific measures of linguistic performance (Norris & Ortega, 2009; Révész, 2011; Robinson, 2007b; Robinson & Gilabert, 2007). In light of this, this dissertation supplements general indices with specific measures in the assessment of learners' L2 performance. The following research questions will be addressed in Chapter IV:

1. How does task complexity affect learners' oral language production in terms of syntactic complexity, accuracy, lexical complexity, and fluency (CALF) as they described the steps followed in solving the problem after their engagement in solving mathematical word problems with their interlocutors in the FTF mode?

2. How does task complexity affect learners' oral language production in terms of syntactic complexity, accuracy, lexical complexity, and fluency (CALF) as they described the steps followed in solving the problem after their engagement in solving mathematical word problems with their interlocutors in text-based SCMC?

3. Does the impact of task complexity on learners' language production as they described the steps followed in solving the problem after their engagement in solving mathematical word problems with their interlocutors in terms of CALF differ depending on the interactive mode (i.e., FTF vs. SCMC)?

4. Do different intended levels of task complexity lead to different levels of perceived task performance, as measured by self-perceived difficulty and time judgment?

t-**1.3 Definitions of Terms**

1.3.1 Task Complexity

In this dissertation, task complexity is defined as the cognitive demands that are solely dependent on the inherent design features of a pedagogical task (e.g., the number of elements involved). This is contrasted against "task difficulty", which is a consequence of learner factors and is dependent on the "resource pools" each learner owns (e.g., intelligence, aptitude, anxiety, motivation) (Robinson, 2001a).

1.3.2 Task Modality

The construct of task modality refers to the environment, in which the interaction takes place. Two modes were involved in this dissertation: face-to-face (FTF) and synchronous computer-mediated communication (SCMC).

1.3.3 Synchronous Computer-mediated Communication (SCMC)

In general, this term refers to real-time online interactions. In this dissertation, the definition of SCMC is restricted to text-based online chats (i.e., instant messaging) in Google Hangout.

1.3.4 Incidental Noticing /Focus on Form

The two terms - incidental noticing and incidental focus on form – are used interchangeably in this dissertation. Both refer to learners' incidental focus on linguistic items while they are negotiating meaning with their interlocutors (Ellis, 2001).

1.3.5 Uptake

The construct of uptake refers to learners' immediate response to interlocutors' corrective feedback (Lyster & Ranta, 2013). In this dissertation, this term includes simple acknowledgments of the feedback (e.g., yes), and modified output (both accurate and inaccurate).

1.4 Organization of the Dissertation

This dissertation is organized into five chapters. The first chapter (this chapter) provides a general introduction to TBLT, discusses the theoretical foundation for pedagogical task design, and the interface between TBLT and CLIL. This chapter then discusses the cognitive dimensions of task-based research, and addresses areas that need further investigation. The next chapter (Chapter II) reviews the history of using pedagogical tasks in L2 teaching and learning, theoretical frameworks of task complexity, and the relevant research studies conducted in recent decades. Chapter II also discusses the characteristics of mathematical language and highlights the importance of examining subject-specific task design in TBLT. Finally, the chapter reviews task-based research conducted in SCMC contexts and highlights research gaps with regard to task complexity research in SCMC contexts. Chapter III and Chapter IV report on two independent studies that examine the effects of cognitive task complexity on task performance in content areas (mathematics). The study in Chapter III investigated how different levels of task complexity influenced NS-NNS L2 learning opportunities while engaged in discussions to solve math word problems in FTF and SCMC interactions. The study in Chapter IV explored the impact of task complexity on learners' mathematical language production as they are engaged in solving math problems with their interlocutors in FTF and SCMC environments, respectively. Finally, Chapter V concludes this dissertation by summarizing the findings, acknowledging the study limitations, highlighting implications of the findings in the field, and providing suggestions for future research.

CHAPTER II

BACKGROUND LITERATURE

This chapter reviews current literature associated with the use of tasks in L2 teaching and learning, and highlights the role of cognitive task complexity in task sequencing. It reviews empirical studies examining the relationship between cognitive task complexity and L2 learning, and the independent measures of task complexity used in the current literature. Furthermore, this chapter discusses the interface of task-based and content-based language teaching approaches and points out the need to explore the effects of different task design features on language development in content areas. Finally, this chapter elaborates on the use of tasks in synchronous computer-mediated communication (SCMC). In the concluding section, I cover areas that need further exploration and the aims of the present study.

2.1 Tasks in Second Language Acquisition

Tasks have been the focus of second language acquisition (SLA) since the 1980s. From the perspective of SLA research, the samples of language use elicited by tasks could be utilized to investigate L2 learning process and test SLA hypotheses. For example, a substantial body of research has utilized tasks to test Long's (1996) Interaction Hypothesis, which highlights the essential role of meaning negotiation in L2 acquisition (e.g., Iwashita, 2003; Mackey, 1999). From the perspective of language pedagogy, tasks not only facilitate L2 learning but also provide evidence of learners' L2 competence (i.e., task-based assessment) (Ellis, 2003). The facilitative role of tasks in L2 learning has been revealed in many studies in terms of (a) engaging learners in goal-oriented communication and fostering form-meaning connections (e.g., de la Fuente, 2006; Fujii & Mackey, 2009); (b) promoting incidental L2 acquisition (e.g., Ellis & Shintani, 2013); and (c) providing opportunities for learners to practice language and notice the gap between the target language and their interlanguage (e.g., Swain, 1995). With regard to language assessment, tasks could be used for eliciting learners' L2 performance. As Ellis (2003) points out, "tasks typically call for real-time production and therefore elicit learners' use of implicit knowledge" (p.137). Given the imperative role of tasks in SLA research and language pedagogy, it is critical to obtain an in-depth understanding of what a "task" is and how task design features influence L2 learning.

2.1.1 Defining "Tasks"

Tasks have been used by both researchers and teacher educators for different purposes. Depending on the degree of authenticity, tasks could be broadly classified into two categories: target tasks and pedagogical tasks (Nunan, 2004). Target tasks refer to real-world tasks beyond the classroom, for example, "borrowing a book from a library", "making an appointment with a doctor", etc. Long (1985) defines a target task as "a piece of work undertaken for oneself or for others, freely or for some reward…In other words, by 'task' is meant the hundred and one things people do in everyday life, at work, at play and in between" (p. 89).

On the other hand, pedagogical tasks are tasks designed to help learners develop the competencies needed to perform target tasks (Long, 2005). What distinguishes a pedagogical task from an "exercise" or a "drill", is that the former emphasizes a non-linguistic outcome while the latter does not. Widdowson (1998) contends that the fundamental difference between "task" and "exercise" lies in how learners perceive the role of linguistic skills while completing the activity. A "task" is contrived "to induce learners to use language they can learn from" (Widdowson, 1998, p. 714). An "exercise", in contrast, requires learners to focus on semantic meaning only (Widdowson, 1998).

Different definitions of pedagogical tasks have been provided by researchers and educators. Table 1 presents some of the definitions from the literature. Within the examined literature, there is a consensus among the researchers and educators that a pedagogical task should have a primary focus on meaning and a non-linguistic outcome, based on which the teacher could determine whether or not the task is completed successfully (e.g., Ellis, 2003; Richards et al., 1985; Skehan, 1996; Willis, 1996). Some researchers and educators believe that a working procedure is also necessary for a pedagogical task as it “allows teachers to control and regulate that process” (Prabhu, 1987, p. 17). Another critical feature of a pedagogical task is that a task should be designed for eliciting real-world language use (Ellis, 2003; Skehan, 1996). As Ellis (2003) elucidates, a task may not replicate a real-world activity, but it should involve “real-world processes of language use” (p. 10).

Table 1

Definitions of a Pedagogical Task

Literature	Definition
Richards et al., (1985)	A task is “an activity or action which is carried out as the result of processing or understanding language, i.e. as a response…A task usually requires the teacher to specify what will be regarded as successful completion of the task” (p. 289).
Breen (1987)	A task is “any structured language learning endeavor which has a particular objective, appropriate content, a specified working procedure, and a range of outcomes for those who undertake the task” (p. 23).

Table 1 Continued

Literature	Definition
Prabhu (1987)	A task is "an activity which requires learner to arrive at an outcome from giving information through some process of thought and which allows teachers to control and regulate that process" (p. 17).
Skehan (1996)	A task is an activity "which has meaning as their primary focus. Success in tasks is evaluated in terms of achievement of an outcome, and takes generally bear some resemblance to real-life language use" (p. 20).
Willis (1996)	A task is "a goal-oriented activity in which learners use language to achieve a real outcome" (p. 2).
Lee (2000)	A task is "(1) a classroom activity or exercise that has (a) an objective attainable only by the interaction among participants, (b) a mechanism for structuring and sequencing interaction, and (c) a focus on meaning exchange; (2) a language learning endeavor that requires learners to comprehend, manipulate, and/or produce the target language as they perform some set of workplans" (p. 32).
Bygate et al., (2001)	A task is "an activity which requires learners to use language, with emphasis on meaning, to attain an objective" (p. 11).
Ellis (2003)	A task is "a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed…A task is intended to result in language use that bears a resemblance, direct or indirect to the way language is used in the real world" (p. 16).

Table 1 Continued

Literature	Definition
Nunan (2004)	A task is "a piece of classroom work that involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form" (p. 4).
Tavakoli & Foster (2011)	A task is "anything that classroom language learners do when focusing their attention primarily on what they want to say to others or what others are trying to say to them" (p. 39).

Among these definitions, Prabhu's (1987) definition concerns the cognitive processes learners experience during task performance. He points out that a task should involve "some process of thought". This claim comes from his earlier work in the Bangalore Communicational Teaching Project (CTP), which aimed to develop a new English teaching approach to replace the traditional "structural" teaching (Brumfit, 1984). Instead of changing the syllabus content, Prabhu decided to create a series of problem-solving tasks to promote language learning. By "some process of thought", Prabhu meant that the task should pose certain cognitive challenge on learners, which pushes them to analyze and evaluate the given information to construct new information through communication. An example of these tasks is making a trip schedule based on prices and timetables. With regard to how these tasks should be sequenced, Prabhu (1987) proposed that the same type of tasks should be ordered by "a commonsense judgment of increasing complexity" (p. 40). That is, the later tasks are supposed to be similar but more complex and cognitively challenging than the earlier ones. Prabhu's project perhaps is the first large-scale attempt to design and sequence pedagogical tasks with a primary focus on the mental challenge rather than linguistic difficulty. Nevertheless, this approach to task sequencing is relatively intuitive, based on teacher's subjective decisions about the cognitive challenges posed to learners, and thus difficult to be systematically applied in the classroom. Following Prabhu's proposal, SLA researchers have offered a number of rationales for task sequencing.

2.1.2 Pedagogical Task Sequencing

In 1992, Long and Crookes proposed the use of a pre-determined syllabus in TBLT. The researchers argued that needs analyses should be conducted and serve as the basis for pedagogical tasks, and these tasks should be sequenced by their "complexity". As Long and Crookes (1992) point out:

> Pedagogical tasks are then derived from the task types and sequenced to form the task-based syllabus. It is the pedagogical task that teachers and students actually work on in the classroom. They will be increasingly complex approximations to the target tasks, which motivated their inclusion. Simplicity and complexity will not result from application of traditional linguistic grading criteria, however, but reside in some aspects of the tasks. (pp. 44-45)

While the researchers did not provide a clear definition regarding the construct of task complexity, they listed a number of task design features that are believed to be influential in manipulating task complexity, including the number of steps involved, the kind of language required, and the planning time provided. Two competing accounts of how task complexity may affect attentional allocation and the amount of focus on L2 constructions during task performance include Skehan's (1998) "Trade-Off Hypothesis" and Robinson's (2003)

“Cognition Hypothesis”. Skehan (1998) believes people have limited attentional capacity, and hence, for L2 learners, there is a continual competition for attentional resources between content and language. With competing and limited attentional resources, learners are more likely to prioritize content over language. Skehan (1998) highlighted the important role of task complexity, what he calls “task difficulty”, in task design and sequencing, and further distinguished three factors that may affect the complexity level of a task: code complexity, cognitive complexity, and communicative stress. First, code complexity refers to the linguistic demands of a task, involving syntactic, semantic, and vocabulary variety and complexity. That is, more complex tasks require more advanced sentence structures and a greater variety of vocabulary. Second, cognitive complexity includes cognitive familiarity and cognitive processing. Cognitive familiarity, on the one hand, indicates the extent to which learners’ schematic knowledge is associated with the topic, the discourse genre, and the procedure of the task. For example, writing an academic essay on political policies is considered as a more complex task compared to writing an unofficial email to a friend. Cognitive processing, on the other hand, entails the degree of information processing required in the task (e.g., comprehending, analyzing, evaluating, constructing). Finally, communicative stress is concerned with implementation factors, such as the number of participants and time limits. In sum, Skehan does not propose a set of criteria for task sequencing, but rather, provides a schema that teachers can use to analyze tasks based on the students’ needs and task difficulty.

Robinson (2001a) has suggested a somewhat different notion of task complexity and argued that cognitive task complexity should be the primary criterion for task sequencing. Robinson’s Cognition Hypothesis is built upon Wickens’ (1992, 2007) theoretical framework and puts less emphasis on cognitive constraints and differentiates between separate resource domains along three dichotomous categories: processing stages (perception vs. response), modality (auditory perception/vocal response vs. visual perception/manual response), and codes of processing (verbal vs. spatial), each of them being in charge for a different aspect of task performance. Based on this perspective, the relative ease or difficulty of a task depends on the competition within these resource pools during task performance.

Robinson also distinguishes cognitive task complexity from two other factors: task difficulty and task condition. Task complexity is defined as "the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner" (Robinson, 2001a, p. 29). It is solely dependent on the inherent design features of the task (e.g., the number of elements involved). Task difficulty is a consequence of learner factors and is dependent on the "resource pools" each learner owns (e.g., intelligence, aptitude, anxiety, motivation). Due to different variables related to individual differences, it is impossible to accurately anticipate learners' perceptions of task difficulty before task performance. Therefore, a syllabus based on task difficulty could be challenging. Lastly, task condition refers to the interactive demands of the task, which is influenced by participation factors (e.g., the interaction mode, the context of the task). Since task conditions are usually determined by task implementers (i.e., classroom teachers), it plays a small role in syllabus design.

The considerations of task complexity in task design and task sequencing have led to different predictions about how task complexity actually influences task performance and L2 learning. Building upon different theoretical backgrounds, Robinson (2001a, 2001b) and Skehan (1998) proposed the Cognition Hypothesis and the Trade-Off Hypothesis, respectively.

2.1.3 Robinson's Cognition Hypothesis

In the Cognition Hypothesis, Robinson (2001a, 2001b, 2005, 2007a, 2007b, 2011) argues that increasing the task cognitive complexity along certain dimensions will (a) result in more accurate and complex, but less fluent L2 production, and (b) promote interaction and learners' attention to linguistic forms, which facilitates noticing and uptake of input. These predictions are based on Wickens' (1989) multiple resources model. According to Wickens, we have multiple attentional resource pools. For example, when we are listening to music, we are using our auditory resources. When we are driving, we are using our spatial resources. Wickens claims that there is no competition between these resource pools, and competition only exists when we draw attention from the same resource pool. For instance, it is easier for us to listen to music (auditory) while driving (spatial) than reading a book (visual) while watching TV (visual). Building upon this model, Robinson (2003) proposes that learners' attentional resources are not in competition while processing the content and the linguistic forms of a task. As shown in Figure 1, increasing cognitive demands of a task would engage more cognitive resources leading to more attention to the linguistic forms in the target language. Increasing cognitive task complexity would also more likely induce comprehension difficulty, which leads to internal (i.e., self-monitoring) and external feedback (e.g., confirmation checks, clarification requests). As Swain and Lapkin (1995) claim, such internal or external feedback can trigger a syntactic analysis of the input beyond the comprehension level, which helps learners notice the gap between their interlanguage and the target language. Consequently, more incorporation of input and modification of output (i.e., accuracy) could be found in learners' L2 production. Additionally, increasing cognitive task complexity along certain dimensions also induces greater functional demands on learners, which leads to greater complexity in L2 production (Robinson, 2001a).

Figure 1

Tasks, Resources, Learning and Performance (Robinson, 2001a)

Task demands		**Cognitive resources**		**Learning mechanisms**		**Performance effects**
More cognitively demanding tasks	⇨	More attention to input/output and noticing/rehearsal in memory	⇨	More rule and instance learning/stage shifts/proceduralization/ cue strengthening	⇨	More incorporation of input More modification of output

Robinson (2007a) further distinguishes between resource-directing and resource-dispersing dimensions in his Triadic Componential Framework related to task complexity manipulation, as shown in Figure 2. Resource-directing variables manipulate cognitive or conceptual demands, which accordingly influence learners' construction of meaning-form mapping. Therefore, higher complexity levels along these variables could engage more cognitive resources and thus lead to more attention to linguistic forms. For example, more memory resources are drawn to the content and more attention is allocated to the linguistic aspects when learners try to narrate events that occurred in the past (There-and-Then) than when they describe events concurrently (Here-and-Now). As a result, the more complex task (There-and-Then) triggers more accurate and complex L2 production. In dialogic tasks, higher cognitive demands induce more communication breakdowns leading to more meaning negotiations, noticing of linguistic aspect in the input, and modification of output.

In contrast, resource-dispersing variables manipulate performative or procedural demands, which affect learners' accessibility to their existing interlanguage system. When task complexity increases along these dimensions, attention diverges from the language, and thus accuracy and complexity of L2 production are expected to decrease. For example, when learners are given less time to think and plan, their attention is more likely to be directed to the content and strategies of performing this task rather than the linguistic aspects of the task. Increasing task complexity along resource-dispersing variables may not facilitate new linguistic knowledge construction, but it may improve learners' ability to access and use their existing L2 knowledge (Robinson, 2005).

Figure 2

The Triadic Componential Framework for Task Classification

Task Complexity (Cognitive factors)	*Task Condition* (Interactive Factors)	*Task Difficulty* (Learner factors)
(Classification criteria: cognitive demands)	(Classification criteria: interactional demands)	(Classification criteria: ability requirements)
(Classification procedure: information-theoretic analyses)	(Classification procedure: behavior-descriptive analyses)	(Classification procedure: ability assessment analyses)
(a) Resource-directing variables making cognitive/ conceptual demands	*(a) Participation variables* making interactional demands	*(a) Ability variables* and task-relevant resource differentials

Figure 2 Continued

Task Complexity (Cognitive factors)	*Task Condition* (Interactive Factors)	*Task Difficulty* (Learner factors)
+/− here and now	+/− open solution	h/l working memory
+/− few elements	+/− one-way flow	h/l reasoning
−/+ spatial reasoning	+/− convergent solution	h/l task-switching
−/+ causal reasoning	+/− few participants	h/l aptitude
−/+ intentional reasoning	+/− few contributions needed	h/l field independence
−/+ perspective-taking	+/− negotiation not needed	h/l mind/intention-reading
(b) Resource-dispersing variables making performative/ procedural demands	*(b) Participant variables* making interactant demands	*(b) Affective variables* and task-relevant state-trait differentials
+/− planning time	+/− same proficiency	h/l openness to experience
+/− single task	+/− same gender	h/l control of emotion
+/− task structure	+/− familiar	h/l task motivation
+/− few steps	+/− shared content knowledge	h/l processing anxiety
+/− independency of steps	+/− equal status and role	h/l willingness to communicate
+/− prior knowledge	+/− shared cultural knowledge	h/l self-efficacy

Adapted from "Criteria for grading and sequencing pedagogic tasks" by P. Robinson (2007a), In Maria del Pilar Garcia Mayo (ed.), Investigating tasks in formal language learning (pp. 7–27).

Based on the distinction between the two dimensions (i.e., resource-directing and resource-dispersing), Robinson (2010) developed two operational principles for task sequencing: (a) tasks should be sequenced from simple to complex based on their cognitive demands, and (b) task complexity should be increased along resource-dispersing dimensions first to foster learners' automatic access to their interlanguage, and then task complexity could be increased along resource-directing dimensions to trigger new form-meaning connections.

2.1.4 Skehan's Trade-Off Hypothesis

In contrast to the multiple resources theory, the predominant premises of Skehan's (1998, 2009) Trade-off Hypothesis are that human's attention and working memory capacity are limited (e.g., Cowan, 2001), and meaning has priority over form during information processing (VanPatten, 1990). Based on these two assumptions, Skehan (1998) proposes that increasing task complexity requires more attentional resources to the meaning and thus less attention can be committed to the linguistic aspects. Additionally, Skehan (2009) claims that there are also trade-off effects within attention to linguistic aspects (i.e., complexity, accuracy, and fluency) during task performance. For instance, if one devotes more attention to linguistic accuracy, linguistic complexity and fluency will be negatively impacted. Likewise, if one attempts to use more advanced and complex language, accuracy and fluency can be negatively affected. Skehan also argues that the major attentional resource competition is between complexity and accuracy. Therefore, manipulating task complexity can only induce greater linguistic complexity or accuracy but not both simultaneously.

Skehan (2009) reviews previous studies which found a simultaneous increase in complexity and accuracy in task performance (e.g., Foster & Skehan, 1999; Tavakoli & Skehan, 2005; Tavakoli & Foster, 2008), and argues that "the Cognition Hypothesis is not automatically needed to account for cases where complexity and accuracy come together" (p. 528). In contrast, the concurrent better performance in complexity and accuracy could be a result of "the joint operation of separate task and task condition factors" (p.510). For instance, Tavakoli and Foster (2008) examined how task complexity influences English learners' L2 performance in picture-story tasks (narrative). Four complexity levels were operationalized along story structure and background knowledge variables (i.e., +structure, +background; +structure, -background; -structure, +background; -structure, -background). The results showed that a task with a tight structure (i.e., a clear storyline) and background knowledge could induce both accurate and complex L2 production. This finding appears to support Robinson's Cognition Hypothesis that less cognitively demanding tasks along resource-dispersing dimensions could facilitate learners' access to their current interlanguage. Hence, more attention could be devoted to the linguistic form, which leads to increased accuracy and complexity. Nevertheless, Skehan (2009) argues that the limited attentional capacity theory could also explain the results. A clear task structure does not require much attention to the content, and thus leaves a great amount of attention to the form, which benefits accuracy. On the other hand, integrating information from the background and the pictures requires learners to use more functional and complex language in storytelling.

2.1.5 Research on Task Complexity

The distinct predictions of the Cognition Hypothesis and the Trade-Off Hypothesis that whether learners could pay simultaneous attention to both accuracy and complexity during task performance have promoted extensive empirical investigations on task complexity and task performance. Therefore, whether increasing task complexity could facilitate interaction and noticing of input during task performance leading to subsequent L2 development has also received great attention. Current task complexity research generally falls into three main categories: (1) task complexity and L2 performance, (2) task complexity and interaction, and (3) task complexity and L2 development. In what follows, I will present an extensive review of the selected studies across the three dimensions.

2.1.5.1 Task Complexity and L2 Performance

In studies examining the effects of task complexity on learners' L2 performance, a complexity, accuracy, and fluency (CAF) model has been widely used as the primary measure of learners' L2 production. Complexity is "the extent to which the language produced in performing a task is elaborate and varied" (Ellis, 2003, p. 340). Accuracy refers to the capacity to produce native-like speech without linguistic errors (Skehan, 2009). Fluency is "the production of language in real time without undue pausing or hesitation" (Ellis & Barkhuizen, 2005, p. 139). Skehan (2009) further argues that there has been a significant omission of considering lexical sophistication as a component of L2 production. He recommends independent measures of lexical complexity and syntactic complexity as two distinct aspects of complexity in the CAF model. In the following section, I will critically review empirical studies that focus on the relationship between cognitive task complexity and learners' L2 performance in the past three decades. Since task complexity has been operationalized in a variety of ways, I have selected a few representative studies that involved the most commonly examined task complexity variables in the existing literature (Sasayama, 2015). Along with Skehan's proposal, Table 2 summarizes the reviewed studies in terms of the four dimensions – syntactic complexity, accuracy, lexical complexity and fluency (CALF).

Table 2

A Summary of the Reviewed Studies on Task Complexity and L2 Performance

Study	Mode	Variables	Dimension	Task type	Syntactic	Accuracy	Lexical	Fluency
Robinson (1995)	oral	here-and-now	resource directing	picture-story (narrative)	+	+	N/A	-
Yuan & Ellis (2003)	writing	planning time	resource dispersing	picture-story (narrative)	-	mixed	mixed	mixed
Gilabert (2007a)	oral	planning time	resource dispersing	picture-story (narrative)	x	x	-	-
	oral	here-and-now	resource directing	picture-story (narrative)	-	+	-	+
Gilabert (2007b)	oral	here-and-now	resource directing	picture-story (narrative)	N/A	mixed	N/A	N/A
	oral	few elements	resource directing	Instruction-giving	N/A	+	N/A	N/A
	oral	reasoning	resource directing	decision-making	N/A	mixed	N/A	N/A
Ishikawa (2007)	writing	here-and-now	resource directing	picture-story (narrative)	+	+	mixed	+
Tavakoli (2009)	oral	task structure	resource dispersing	picture-story (narrative)	N/A	-	N/A	-
	oral	storyline	resource dispersing	picture-story (narrative)	-	N/A	mixed	N/A
Sasayama & Izumi (2012)	oral	planning time	resource dispersing	picture-story (narrative)	-	x	N/A	+
	oral	few elements	resource directing	picture-story (narrative)	+	-	N/A	-

Note. "+" indicates a positive effect of increasing task complexity; "-" indicates a negative effect of increasing task complexity; "x" indicates no significant effect of task complexity; "mixed" indicates mixed effect of increasing task complexity; "N/A" indicates that this dimension was not investigated.

Robinson (1995) investigated how task complexity impacted learners' oral narrative performance with a repeated-measure design. The participants were 12 college students with intermediate English proficiency level. Three different wordless narrative strips were used. Each strip had two versions (i.e., simple and complex) based on its cognitive load, which was manipulated along [+/- here-and-now]. The learners were either asked to narrate the story while looking at the strips (i.e., simple), or narrate the story after they reviewed the strips (i.e., complex). Learners' L2 production was measured for syntactic complexity – the number of S-nodes per T-unit and multipropositional utterances (i.e., the number of propositions in each utterance), accuracy – target-like use of articles, and fluency – the number of words per pausal unit. The results showed that the complex version of the tasks generated more accurate and complex utterances than the simple version. However, more fluent speech was found in the simple version. The findings confirmed the prediction of the Cognition Hypothesis that increasing task complexity along resource-directing variables could lead to greater accuracy and complexity but lower fluency in learners' L2 production.

Yuan and Ellis (2003) conducted a study on the effects of planning time on L2 oral production. Forty-two undergraduate learners of English were divided into three groups: no planning (i.e., 0.5-minute planning time), pre-task planning (i.e., 10 minutes planning time), and on-line planning (0.5-minute planning time). The participants were asked to use at least four sentences to describe each of the six pictures. Both the no planning group and the pre-task planning group were required to finish the task within five minutes, while the on-line planning group was given unlimited time for task completion. The learners' L2 production were measured for syntactic complexity – the number of clause per T-unit and the number of different grammatical verb forms, accuracy – the percentage of error-free clauses and the percentage of target-like verbs, lexical complexity – average type-token ratio in each segment (i.e., every 40 words), and fluency – the number of syllables and the number of meaningful syllables (i.e., excluding repeated/repaired syllables) per minute. The results showed that the no planning group obtained greater fluency and lexical complexity than the other two groups. However, the two planning groups achieved higher syntactic complexity. According to the comparison between the two planning groups, the pre-task planning group demonstrated greater fluency and vocabulary variation while the on-line planning group produced more accurate utterances. Overall, the findings indicate that providing planning time could facilitate syntactic complexity. The type of planning, on the other hand, influences fluency and accuracy. Based on this study, learners tend to pay more attention to accuracy rather than fluency if they are planning on-line. In contrast, fluency would be the priority if pre-task planning is provided.

Gilabert (2007a) examined the impact of task complexity on adult learners' English narrative oral production. Forty-eight Spanish L1 learners of English whose English proficiency levels were low-intermediate were involved in the study. A repeated-measure design was adopted. Four levels of task complexity were operationalized along [+/- planning time] and [+/- here-and-now] with four picture stories. Ten minutes planning time were provided for the two planned tasks (i.e., simple) and 50 seconds for the other two unplanned tasks (i.e., complex). Learners' L2 performance was measured through syntactic complexity – the number of S-Nodes per T-unit (i.e., the number of clauses in each T-unit), accuracy – the percentage of self-repairs, lexical complexity – the Guiraud's Index of lexical richness (i.e., the type/token ratio), and fluency – pruned speech rate (i.e., the number of syllables per minute). The results showed that, on the one hand, increasing task complexity along [+/- planning time] (i.e., decreasing planning time) negatively affected learners' lexical complexity and fluency. On the other hand, no significant effects were found for syntactic complexity and accuracy. This finding is in accordance with both Robinson's and Skehan's predictions that increasing task complexity along planning time (i.e., resource-dispersing variable) diverges learners' attention from the linguistic form and thus only complexity or accuracy might be improved but not both simultaneously. On the other hand, Gilabert also found that increasing task complexity along [+/- here-and-now] positively affected accuracy and fluency, but negatively influenced syntactic and lexical complexity. This finding tends to support Skehan's Trade-off Hypothesis that more complex tasks require more attention to the content rather than to the form. Therefore, the gains in accuracy were achieved at the expense of syntactic and lexical complexity.

In another study, Gilabert (2007b) investigated the effects of task complexity on adult learners' L2 oral language production in terms of self-repair, across varied task types (i.e., narrative task, instruction-giving task, and decision-making task). The participants were 42 Spanish L1 learners of English with a low-intermediate English proficiency level. The participants were asked to perform the three different task types with two task complexity levels (i.e., simple and complex). The narrative picture-story task was manipulated along [+/- here-and-now] and the study design was repeated-measure. The instruction-giving task was manipulated along [+/- few elements]. The simple version had fewer but more distinguishable landmarks on the map while the complex version had many similar landmarks for the learners to refer to. Finally, the decision-making task was manipulated along [+/- reasoning]. The task was called "Fire Chief", which required learners to rescue people from a building on fire. The participants had to determine the sequence of the people they rescue, and to justify their decisions. The simple version involved similar types of people, more resources, and mostly unrelated factors (e.g., the fire was static). The complex version, in contrast, included specific types of people (e.g., an injured child), fewer resources, and factors associated with each other. Gilabert only focused on the impact of task complexity on accuracy, measured by self-repair. The participants' self-repair was analyzed based on the number of errors per AS-unit, the errors/words ratio, repaired errors per AS-unit, the error-repair/words ratio, the number of repairs per AS-unit, the repairs/words ratio, the percentage of self-repair, the repaired/unrepaired error ratio, and the corrected repaired/unrepaired error ratio. The results demonstrated different impacts of task complexity on self-repair in different task types. First, in the narrative task, although the participants made fewer errors and had a higher repaired to unrepaired rate in the complex task, no significant differences were found in the number of errors per AS-unit and the self-repair rate between the two task complexity levels. Second, the results for the map task were clearer with both a larger number of repairs and a higher repair rate in the complex version. Finally, in the decision-making task, the cognitive demands of the task seemed not to significantly affect the number of errors and self-repairs. Taken together, the findings suggested that increasing task complexity along resource-directing variables may promote monitoring (i.e., attention) during the encoding process (e.g., the narrative task and the map task), which supports the Cognition Hypothesis. The results also indicated that task type might be a moderator variable that mitigates the effects of task complexity on L2 performance.

Ishikawa (2007) examined the effect of task complexity along [+/- here-and-now] on learners' L2 writing. Fifty-four Japanese 3rd year high school students who were learning English as a L2 were included. The participants were divided into two groups: the simple group and the complex group. Both groups were provided a strip cartoon with a prompt and were allowed to view it for five minutes before the writing session. Both groups were given 30 minutes for writing. There were two differences between the two task conditions. First, the prompt given to the simple group was written in the present tense while the prompt for the complex group was written in the past tense. Second, the simple group was allowed to look at the strip cartoon during the writing session. In contrast, the complex group had to return the strip cartoon to the teacher after viewing it for five minutes. Learners' L2 narrations were evaluated for syntactic complexity – the number of S-nodes and clauses per T-unit, the percentage of dependent clauses, and the number of S-nodes per clause; accuracy – the number of error-free T-units, the percentage of error-free clauses, and the percentage of target-like use of English articles; lexical complexity – the lexical to function words ratio, the percentage of lexical words and two kinds of type-token ratio; and fluency – the number of words per T-unit, the number of words per clause, and length of the writing. The results showed that increasing task complexity along [+/- here-and-now] tended to promote syntactic complexity, accuracy, and fluency. As for lexical complexity, the cognitive complexity of the task seemed to negatively affect lexical density with no noticeable effect on lexical variation. The overall findings were more compatible with the Cognition Hypothesis that learners' attention could be directed to both accuracy and complexity.

Tavakoli (2009) conducted a study examining the effects of task complexity along [+/- task structure] and [+/- storyline] on learners' L2 oral production. The participants were 60 Iranian adult learners of English. All the participants were at an intermediate English proficiency level. The study adopted a 2 x 3 factorial design. The three levels of task structure were: a clear event (simple), an apparent timeline of several events (+complex), and arbitrary sequence of several events (++complex). Additionally, at each structure level, there were two tasks that were different in the number of storylines (i.e., whether both background and foreground were provided). Learners' oral L2 production was measured in terms of syntactic complexity – the number of clauses per AS unit, accuracy – the percentage of error-free clauses, fluency – repetition, speech rate, and the number of pauses, and lexical complexity – D (i.e., VocD analysis). The results showed that more structured tasks induced more accurate and more fluent utterances. Tasks with both background and foreground (i.e., simpler tasks) facilitated syntactic complexity while mixed results were obtained for lexical complexity.

Sasayama and Izumi (2012) used picture narrative tasks to examine the effects of pre-task planning and the number of elements involved in the task on learners' L2 oral production. The participants were 23 Japanese high school students who were learning English as a foreign language. All participants performed a simple and a complex picture-based tasks varying along the number of elements involved in the task. Ten of the participants were given pre-task planning time and the others were not. Learners' L2 production were measured for accuracy, complexity and fluency. The results indicated that when task complexity was raised by adding more elements to the task, accuracy and fluency were negatively affected. However, syntactic complexity was positively affected. On the other hand, when task complexity was increased by removing pre-task planning, fluency was enhanced while syntactic complexity was negatively affected. No significant effect was found for accuracy.

To conclude, the above-reviewed studies, except one study (Sasayama & Izumi, 2012), suggest that increasing task complexity along resource-directing dimensions appears to facilitate accuracy in L2 production (Gilabert, 2007a, 2007b; Ishikawa, 2007; Robinson, 1995). Nevertheless, its effects on complexity and fluency remain unclear. On the other hand, increasing task complexity along resource-dispersing dimensions tends to negatively affect both syntactic and lexical complexity of learners' L2 production (Gilabert, 2007a; Tavaloli, 2009; Yuan & Ellis, 2003). However, its effects on accuracy and fluency are still ambiguous.

2.1.5.2 Task Complexity and Interaction

According to Long's (1996) Interaction Hypothesis, the key for language acquisition is negotiation of meaning, through which learners' attention could be directed from the content to the linguistic features. More attention to form could then facilitate "noticing", an internal process that is believed to be necessary for L2 acquisition (Schmidt, 2010). Therefore, understanding how the manipulation of task features can best promote meaning negotiation has become of great importance to both SLA researchers and educators. With regard to the relationship between cognitive task complexity and interaction, Robinson (2003, 2007b) proposes in his Cognition Hypothesis that increasing the cognitive demanding demands of a task could trigger more interaction and attention to form. This prediction has been tested by a great number of studies. Many of these studies not only examined the effects of task complexity on interaction but also its effects on the subsequent L2 development (e.g., Kim, 2012). In this section, I will only address studies that specifically focus on interaction, and the next section will discuss the empirical evidence for the effects of task complexity on L2 development. I have selected studies that were published in the past two decades and examined the effects of task complexity on the production of specific discourse moves such as negotiation of meaning and LRE. These studies were also representative for the most common variables (e.g., +/- few elements, +/- reasoning) for task complexity operationalization in the existing literature (Sasayama, 2015). Table 3 summarizes the design and the findings of each reviewed study in this section.

Table 3

A Summary of the Reviewed Studies on Task Complexity and Interaction

Study	Variables	Interactional type	Results
Hardy & Moore (2004)	Task familiarity	NNS-NNS	No significant effects of task complexity.
	Task structure	NNS-NNS	More meaning negotiation was observed in the complex task.
Nuevo (2006)	Reasoning	NNS-NNS	The influence of task complexity varied depending on the type of the interactional measures.
Robinson (2007b)	Intentional reasoning	NNS-NNS	More meaning negotiation and uptake were observed in the more complex tasks.
Révész (2011)	Reasoning & few elements	NNS-NNS	More LREs and interactional features were found in the complex task.
Solon et al. (2017)	Few elements	NNS-NNS	More pronunciation related LREs were found in the simple task but the difference was not significant.

Hardy and Moore (2004) examined how task structure and content familiarity influenced learner-learner interaction during task performance. Fifty-six (28 dyads) college students learning German were included in the study. Their German language proficiency was at the intermediate level. A Greco-Latin square design was utilized. The task was to observe a video clip in German and then evaluate the character in terms of personality, behavior, speech, and appearance. Two video clips, one familiar and one unfamiliar, were used. The familiar video was from the students' previous class materials so they were already familiar with the characters and the setting. In contrast, the unfamiliar video was never involved in previous classes and was completely new to the students. As for task structure, two complexity levels were operationalized based on whether the observation was already provided to the students. The simple task (+task structure) already provided the observations and asked the students to show the degree of agreement on these statements (i.e., rating). The complex task (-task structure), however, required the students to write down their own observations. After they completed their own observation and evaluation, they needed to discuss with their partner to reach a consensus on the rating. The students' conversations were coded and analyzed with regard to meaning negotiation strategies, such as clarification requests, confirmation checks, and comprehension checks. The results showed that increasing task complexity by reducing structural task support significantly promoted meaning negotiations. No significant effect was found for task familiarity on learner-learner interaction.

Nuevo (2006) also explored the relationship between task complexity and interaction-driven L2 learning opportunities. In the study, task complexity was manipulated along [+/- reasoning demands]. One hundred and thirteen learners performed a picture narration task and a decision-making task. L2 learning opportunities were analyzed based on nine different interactional moves including recasts, clarification requests, confirmation checks, comprehension checks, hypothesis testing, self-repairs, metalinguistic talks, noticing of linguistic deficiency, and other repetitions. The results showed that while learners generated more uptake of recasts, more comprehension checks, and more other repetitions in the simple task, they tended to test hypotheses (i.e., interlanguage) to a greater extent while performing the complex task. No significant effects of task complexity were detected on other interactional measures. The mixed findings suggested that the influence of task complexity varied depending on the type of the interactional measures.

Similarly, Robinson (2007b) also investigated the impact of task complexity on learners' interaction. The participants were 42 (21 dyads) Japanese undergraduate students who were learning English as a second language. Three task complexity levels were manipulated along [+/- intentional reasoning]. Each dyad was required to perform all three versions of the task. The task was to sequence the provided pictures to complete a story. There were causal relationships among the events in the pictures. One student (the speaker) first looked at the pictures and decided on the sequence, and then he/she narrated the story to his/her partner (the listener). The listener needed to sequence the pictures according to the speaker's description. Learners' interaction was assessed in terms of confirmation checks, the number of turns taken, clarification requests, confirmation checks, and uptake. The results showed that significantly more interactions with regard to all these indices were found in the more complex versions of the task.

To explore how task complexity influences interactional learning opportunities, Révész (2011) studied not only the quantity but also the quality of learners' interaction in different complexity versions of a task. Forty-three English adult learners were involved in the study. A decision-making task with two versions (i.e., simple and complex) was designed for the study. The task was to provide financial support to certain projects. Given the limited amount of funding, learners worked in groups to make decisions on which projects they should support. Task complexity was manipulated along [+/- reasoning] and [+/-few elements]. Language related episodes (LREs), "where students talk about language they are producing, question their language use, or other-or self-correct their language production" (Swain & Lapkin, 2001, p. 104), were identified to capture language learning opportunities. The LREs were then classified into two categories depending on whether the LRE was triggered by communication problems or by linguistic issues. Additionally, the LREs were coded and analyzed in terms of confirmation check, clarification request, recast, and metalinguistic talk. The results revealed an overall positive effect of greater task complexity on interactional language learning opportunities with significantly more LREs and metalinguistic talks observed in the complex task. A greater amount of all the other interactional features were also found in the complex task, although the difference was not significant.

More recently, Solon et al., (2017) investigated the effect of task complexity on pronunciation related LREs. Thirty-four (17 dyads) adult learners of Spanish completed two versions (i.e., simple and complex) of an information-gap map task. Task complexity was manipulated along [+/- few elements] with the simple version having fewer stops and elements on the route. The results showed that the learners produced a greater number of LREs during the completion of the simple task, although the difference was not significant. This finding contradicted previous studies (e.g., Révész, 2011) and the prediction of the Cognition Hypothesis that greater task complexity could lead to more attention to linguistic forms. As authors pointed out, a possible explanation of the finding the authors pointed out was that there could be a fundamental difference between phonetic targets and other linguistic aspects, such as grammar. Unlike grammar, which could be referred to be specific rules to be applied in the speech, there are a range of possibilities for phonetic combination, which requires "precise physical modification of articulators to produce" (Solon et al., 2017, p. 370). Another possibility Solon et al. mentioned was that since pronunciation training was usually not the main focus of foreign language class, the students may not be used to discuss pronunciation during task performance, which could also lead to fewer incidences of pronunciation related LREs.

To summarize, the review of the literature tends to support the prediction of Robinson's Cognition Hypothesis that increasing task complexity could induce more interaction and attention to linguistic forms during task performance (e.g., Hardy & Moore, 2004; Révész, 2011; Robinson, 2007b). Specifically, greater task complexity appears to promote the use of different interactional strategies in meaning negotiation, such as comprehension checks and metalinguistic talk (e.g., Révész, 2011). Additionally, Solon et al.'s (2017) study offers insights into the role of task complexity in L2 learning at the phonetic and phonological level.

2.1.5.3 Task Complexity and L2 Development

With regard to L2 development, Robinson (2007a, 2007b) proposes that increasing task complexity could promote interaction and thus facilitate s subsequent L2 development. Although researchers have provided empirical evidence to support the positive impact of manipulating task complexity on interactive learning opportunities, limited research has been carried out measuring the subsequent language gains. In what follows, I will review the most recent existing task complexity studies in the past decade that utilize specific language assessments for the evaluation of learners' L2 development.

Kim and Tracy-Ventura (2011) carried out a study investigating how task complexity affects the development of English past tense. The participants were 88 Korean college learners of English. The students were divided into three groups: simple, +complex, and ++complex. The cognitive demand of the tasks was manipulated along [+/- reasoning]. A pre-test, post-test, and delayed post-test design was adapted. Each student completed four interactive tasks at their group complexity level during the intervention phase. All these four tasks were at the same complexity level. For example, the simple group completed four simple tasks during the intervention. Learners' language development was measured by three oral production tests. The results from both the posttest and the delayed posttest indicated a positive effect of increasing task complexity on learners' English past tense development.

Kim (2012) examined the effect of task complexity on English question development in an English as a foreign language (EFL) classroom context. The participants were 191 Korean undergraduate students who were learning English as their L2. The participants were randomly assigned to four different groups: three experimental groups (i.e., simple, +complex, ++complex) and one comparison group. The comparison group received traditional English instruction while the three experimental groups received the task-based intervention. During the intervention, each student completed four tasks, same as in Kim and Tracy-Ventura (2011). Task complexity was manipulated along [+/- reasoning]. A pre-test, post-test, and delayed post-test design was adapted. Language learning opportunities were identified based on the number of LREs during task completion. Students' ability to use English questions was measured by three research-based oral production tests. Students' oral production was analyzed based on the language developmental model by Pienemann and Johnston (1987). Six progressive stages of English question development were identified in the students' L2 output. The results revealed a significant effect of task complexity on the amount of LREs: ++complex group > +complex group > simple group. With regard to English question development, the results indicated that the more complex the task group was, the more students in that group reached a higher stage of question development. Together, the findings suggested that more complex interactive tasks offered more language learning opportunities, which leads to subsequent L2 development.

Later, Kim and Taguchi (2015) examined task complexity and pragmatics development (i.e., request) in an EFL classroom context. Seventy-three Korean high school students were included in the study. The students were divided into two groups: the simple group and the complex group. Similarly, a pre-, post-, and delayed post-test design was used. The treatment consisted of two collaborative tasks with two levels of complexity along [+/- reasoning demands]. Students' interaction was analyzed based on the occurrence of pragmatic-related episodes (PREs). Students' learning outcome was assessed through a discourse completion test (DCT), in which the learners were required to use the intended speech act in English by following the situation promptly. The results revealed a strong positive impact of greater task complexity on certain PRE targets (i.e., contextual features, head acts, and preparators) but not on other PRE targets (i.e., grounders, hedges, and amplifiers). As for L2 development, the results did not demonstrate any significant differences between the simple and the complex task group in the post-test. However, only the complex task group maintained their gains in the delayed post-test, suggesting that task complexity might have a long-term effect on L2 pragmatic development.

The above three studies tested Robinson's prediction regarding the subsequent L2 gains that arose from interactive task performance influenced by task complexity. Overall, the findings corroborate the Cognition Hypothesis that increasing the cognitive demand of interactive tasks could facilitate interaction as well as subsequent L2 development (Kim, 2012; Kim & Tracy-Ventura, 2011). In particular, Kim and Taguchi's (2015) study suggests that the effect on L2 development might not be detectable until after a long-term period. Given the limited number of relevant studies in this specific area, it is difficult to provide conclusive evidence on the effect of task complexity dimensions and L2 learning outcomes in different areas. More studies are needed to show the relationship between task complexity dimensions and different components of L2 development.

In sum, findings from the reviewed literature indicate a critical role of task complexity in L2 learning. Generally speaking, increasing cognitive demands of a task encourages learners to posit allocate more attention to the linguistic form, which promotes "noticing" during interaction (e.g., Révész, 2011; Robinson, 2007b) and facilitates accuracy in L2 production (e.g., Gilabert, 2007a; Ishikawa, 2007; Robinson, 1995). Additionally, the effect of task complexity on L2 development might be carried over time (Kim & Taguchi, 2015). However, accuracy is only one aspect of L2 performance. No consensus has been made on whether greater accuracy is achieved at expense of other linguistic aspects (i.e., syntactic complexity, lexical complexity, and fluency). Furthermore, all the interpretations of the previous studies findings are based on the assumption that the designed complex tasks in the studies were indeed more cognitively demanding indeed. However, it is possible that the designed tasks assumed to have greater complexity may not be perceived as actually more demanding by learners. Therefore, in order to advance our understanding about the impact of task complexity on task performance, it is imperative to first attest the validity of task complexity manipulation.

2.1.6 Independent Measures of Task Complexity

Although a considerable body of research has been undertaken to investigate the role of task complexity in L2 learning, the validity of task complexity manipulation level based on manipulation of task cognitive loads has received little consideration. As noted by Norris and Ortega (2003) and Norris (2010), independent evidence must be provided to make sure that the manipulation of cognitive demands really induces the desired processes. To validate the cognitive load of tasks, researchers have proposed a number of independent measures of task complexity: (a) self-perceived task difficulty (Baralt, 2010; Gilabert, 2007b; Robinson, 2001b), (b) time judgment (Baralt, 2013; Sasayama, 2015), (c) stimulated recall (Kim et al., 2015), (d) expert judgments (Révész et al., 2016; Révész et al., 2014), (e) dual-task methodology (Révész et al., 2016; Révész et al., 2014; Sasayama, 2015), and (f) eye-tracking technology (Révész et al., 2014; Smith, 2012). Self-perceived task difficulty, time judgment, and stimulated recall rely on learners' subjective interpretation of the task. Expert judgments, dual-task methodology, and eye-tracking technology are more objective measures of task complexity.

The use of a subjective task difficulty questionnaire for task complexity identification was originally proposed and tested by Robinson (2007b). In the study, Robinson examined whether the level the assumed level of task complexity was correlated with learners' self-perceived task difficulty. The participants were asked to rate the difficulty of the task they just performed on a nine-point Likert scale. The variables included in the survey were task difficulty, stress, ability, interest, and motivation. The findings confirmed Robinson's prediction that learners who completed the complex task rated the task more difficult and felt more stressed and less confident in performing the task. Task difficulty questionnaire was then used as an independent measure of task complexity by other researchers in the field of TBLT (e.g., Baralt, 2010; Gilabert, 2007b).

Time judgment is a frequently used subjective measure of cognitive load in psychology. In a meta-analysis study, Block et al. (2010) reviewed 117 empirical studies examining the relationship between cognitive load and time duration judgment. The findings revealed that when time judgment was undertaken retrospectively, longer estimated time indicated higher cognitive demands. In the field of TBLT, time judgment is used by researchers as an independent measure of task complexity. Baralt (2013) evaluated task complexity with time judgment and found that the perceived duration of the more complex task was longer than the actual performance time.

Stimulated recall, which is believed to be able to capture learners' cognitive processing during task performance (e.g., Gass & Mackey, 2000), has also been employed as a measure of task complexity. Based on the data from stimulated recall, Kim et al. (2015) found that learners engaged in two types of cognitive processes (i.e., personal evaluation and task option evaluation) during information gap task performance. While both groups engaged in the personal evaluation, only the complex group further evaluated options provided in the task. Kim et al. concluded that different levels of cognitive demands were operationalized successfully as intended.

The other three measures of cognitive load – expert judgment, dual-task methodology, and eye-tracking – were examined by Révész et al. (2014). The tasks in the study were designed for the participants to learn how to express causal relationships in English. The participants were asked to match a "cause" item with a "consequence" item. In the simple version of the task, the casual relationship between the cause and the consequence was straightforward. In contrast, the complex version contained events that might be caused by different factors. The participants were asked to select the most direct cause of the event.

One of the task complexity measures Révész et al. (2014) chose was expert judgment. Expert judgment is often used as a means of gauging the difficulty of test items in language assessment. It is also utilized to evaluate the cognitive loads imposed on the learners in task-based performance assessments (Brown et al., 2002). In Révész et al.'s (2014) study, two doctoral students in applied linguistics were asked to rate the complexity of the corresponding item for each event on a 5-point Likert scale (i.e., 0 = minimum, 5 = maximum). The results showed that the complex items received higher ratings than the simple items, suggesting the successful manipulation of task complexity.

The researchers also used the dual-task methodology to evaluate task complexity. The dual-task methodology requires learners to perform a parallel but simpler secondary task after conducting the primary task. The rationale of this methodology is that participants' reaction time and accuracy on the secondary task will reflect the cognitive load imposed on them in the primary task. That is, completing the more cognitively demanding (versus less cognitively demanding) primary task leads to less accurate and slower performance of the secondary task. In Révész et al.'s (2014) study, immediately following the primary (experimental) task, the participants were given the secondary task: responding to only one color (i.e., red or green) while the colors were changing from one to the other on the computer screen. The results from the dual-task methodology were in line with the researchers' prediction that participants who carried out the design-to-be simple version of the experimental task produced more accurate responses in the secondary task. As for reaction time, no statistically significant difference was found between the simple and complex tasks. These findings echo Révész et al.'s (2016) study that only accuracy appeared to measure the cognitive demands of the primary task, while reaction time may not be sufficiently sensitive to validate task complexity manipulation. However, contrary to the above findings, Sasayama (2016) found longer reaction time for the more complex tasks, whereas there were no statistically significant differences for accuracy in the performance of the secondary task.

Finally, Révész et al. (2014) also employed eye-tracking technology to attest task complexity manipulation. The premise is that by analyzing individuals' locus, sequence, and duration of eye fixations, researchers could penetrate their cognitive processes: more and longer gaze duration represents greater cognitive engagement (Hyönä, 2010; Just & Carpenter, 1976). Based on this assumption, it could be hypothesized that a longer average duration of eye fixation could be found in the complex task. This hypothesis was confirmed by Révész et al. that a greater number of eye-fixation counts and longer durations were revealed under the complex condition.

Despite the fact that researchers have proposed a variety of ways to measure task complexity independently and to validate it, the majority of current research studies still simply assume that task complexity has been ideally manipulated as expected. Only a few studies have validated this important construct with empirical evidence. In a research synthesis, Sasayama et al. (2015) found that among 129 task complexity studies, only 18% of the studies supported the cognitive load of the experimental tasks using independent measures of task complexity. Failing to provide independent evidence of diverse cognitive task complexity could lead to inaccurate interpretation of research findings. To deal with this methodological issue and provide more robust results, multiple independent measures of task complexity should be employed to verify diverse cognitive loads imposed on L2 learners in research studies.

2.2 Incorporating Pedagogical Tasks in Subject Matter Content

With the growing empirical support for the facilitative role of TBLT in L2 learning, researchers and educators have started to extend the scope of research by investigating the use of pedagogical tasks in subject matter contents, in which the academic discourse is mainly used. In recent years, connecting TBLT to content and language integrated learning (CLIL) (Lopes, 2020; Lyster, 2015; Ortega, 2015;) has received growing attention from language researchers and educators. In what follows, the CLIL approach will be explained first before discussing the links between CLIL and TBLT.

2.2.1 Content and Language Integrated Learning

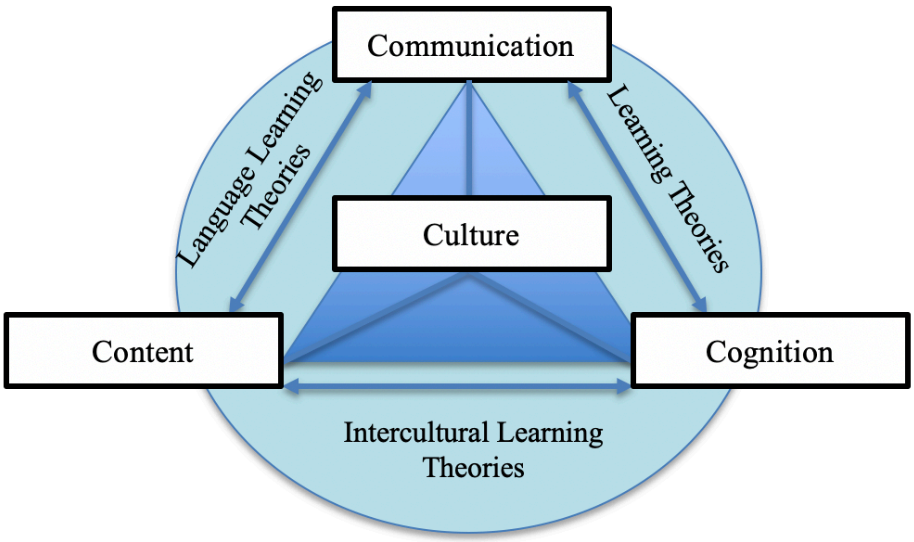
Content and language integrated learning (CLIL) was developed in Europe in the 1990s with the purpose of promoting multilingualism (Marsh, 2008). CLIL is defined as "a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language" (Coyle et al., 2010, p. 1). The core characteristic of CLIL lies in the concept of integration (of form and function in language learning). In other words, learners acquire subject knowledge through the language (i.e., language function) and develop language skills by learning the content (i.e., language form).

Coyle (2007) proposes a 4Cs framework for CLIL curriculum development, as shown in Figure 3. The framework highlights the connection between content, communication, cognition, and culture. Content refers to the subject knowledge within the curriculum. Communication indicates the role of the target language as a means of learning the content in an interactive environment. Cognition makes references to the internal information processing during (language and content) knowledge construction. Finally, cultural elements are needed to raise multicultural awareness and global citizenship.

In addition, Coyle (2007) further distinguishes between three different roles of the target language in CLIL: language *of* learning, language *for* learning, and language *through* learning. Language *of* learning constitutes the functional role of the language required for content learning. For example, the purpose of learning academic vocabulary is to scaffold subject content learning. Language *for* learning emphasizes discourse strategies in the target language used to promote classroom interactions, such as asking and answering questions and cooperative group work. Finally, the idea of language *through* learning is in congruent with the principle of TBLT that learners develop the target language by engaging in meaning-based learning activities.

Figure 3

The 4Cs Framework of CLIL (Coyle, 2007)



2.2.2 The Interface of CLIL and TBLT

In light of the above, both CLIL and TBLT view the target language as a means of communication and believe that language learning is achieved through interaction and collaboration. Most importantly, both approaches are based on the premise that language and meaning are inseparable (Ortega, 2015), although "meaning" is conceptualized differently by each approach. In CLIL, "meaning" represents subject-learning, while in TBLT, "meaning" refers to experiential and goal-oriented learning.

Besides the shared interests CLIL and TBLT have, each approach has its distinct features. First of all, the two approaches usually target different age groups. While TBLT has a primary focus on college-level learners, CLIL is mostly applied in secondary education. Second, each approach has its own educational context. TBLT programs are mostly implemented in second language contexts, where learners are exposed to the target language outside of the classroom in a variety of ways. In contrast, CLIL programs are thriving in foreign language contexts, where teachers aim to create an input-rich environment that is lacking outside of the classroom. Finally, the goals of educational effectiveness are different. The ultimate goal of TBLT is to transfer learning from pedagogical tasks to authentic tasks in the real world. In CLIL, on the other hand, the goal is to achieve balanced gains in both language learning and content learning. These divergences, however, do not necessarily signal or imply incompatibility between the two approaches. In fact, language researchers and educators have proposed that incorporating pedagogical tasks of TBLT in CLIL programs can promote negotiation of meaning, learner's attention to linguistic forms, and in-depth understanding of content knowledge (Lyster, 2015; Lopes, 2020; Meyer, 2010; Shehadeh, 2018).

2.2.3 Research on the Integration of TBLT and CLIL

Despite the commonalities TBLT and CLIL share, the links between the two approaches have been sparsely studied. The special issue of *System (*2015) journal constitutes the first attempt at investigating research interfaces between CLIL and TBLT (Ortega, 2015). The studies included in this special issue offer empirical evidence regarding the opportunities and challenges of blending the two approaches. For example, Lyster (2015) provides an exemplary illustration of using form-focused tasks to draw learners' attention to language in CLIL contexts in Canada. Lyster conducted two studies, each of which employed a different task-based approach that crossed borders between content areas or target languages. The first study integrated a focus on French grammatical gender across 5th-grade (10-11 years old) French immersion students. The experimental group completed noticing activities in the language arts class, awareness activities in the social studies class, and practice activities in the science class. In contrast, the comparison group took the regular CLIL program without focused tasks. The second study involved literacy tasks with a focus on French derivational morphology. The participants were 2nd-grade (7-8 years old) French immersion students who completed the focused tasks in one language (e.g., French) and then in the other (e.g., English). The comparison group, on the other hand, attended the regular CLIL classes (i.e., French literacy). The results of both studies show that the experimental groups significantly outperformed the comparison groups on the language assessments. Based on the findings, Lyster claims that "tasks would be pivotal in such a cross-curricular approach and would be designed to provide purposeful opportunities for strengthening connections between language and content learning" (p. 12). In the meanwhile, he points out two challenges educators may face when they attempt to integrate TBLT to CLIL: (a) language and content integrated task design, and (b) collaboration between teachers from different areas.

Llinares and Dalton-Puffer (2015) examined learners' use of evaluative language across five different take types (i.e., whole-class discussions, group-work discussions, individual interviews, oral presentations, and role-plays) in secondary social science classrooms (i.e., history, economics, and geography). The data were classroom discourses during task performance from three European CLIL contexts: Austria, Spain, and Finland. Learners' use of appraisal across tasks was coded and analyzed following Martin and White's (2005) appraisal model, in which three appraisal types were identified: attitude (i.e., feelings and emotional reactions, such as appreciate and judge), engagement (i.e., the sources of attitudes, including contract, expand, and justify), and graduation (i.e., grading phenomena that is either amplifying or weakening, such as *a lot* and *sort of*).

The results showed that there were clear differences in the frequency and distribution of different appraisal types across task types, with the role play task eliciting the richest evaluative language. The researchers argued that different task types provided different affordances for the use of evaluative language; and role play appeared to be more cognitively complex than the other four task types, since the use of appraisal is associated with the lexico-grammatical resources used to convey interpersonal meanings. Based on the findings, Llinares and Dalton-Puffer encouraged teachers to use a wide range of tasks in their CLIL classrooms, especially role play.

García Mayo and Lázaro Ibarrola (2015) examined the pedagogical blends of TBLT and CLIL by incorporating spot-the-difference tasks in CLIL and English as a foreign language (EFL) classroom. The researchers compared the amount of meaning negotiation instances elicited in children interaction during task performance between the two programs. The study involved two age groups: 8-9 and 10-11 years old. The results showed that learners in the CLIL program negotiated more than their EFL peers; and older learners in both contexts negotiated less frequently than younger learners. Based on the findings, the researchers contended that "the input characteristics the learners in CLIL programs are exposed to are likely to be the factor determining their better command of conversational adjustments" (p. 50). With reference to the results obtained from different age groups, the researchers explained that the older learners might have better linguistic abilities than the younger learners, which allowed them to communicate with fewer difficulties and thus reduced the amount of negotiation moves.

Recently, Li and Chen (2019) examined the effectiveness of incorporating pedagogical tasks in facilitating military English learning. The participants were 120 undergraduate students who were taking military English classes in addition to traditional English classes. The students were divided into two groups: CBLT group and TBLT group. Students in the CBLT group received regular content-based instruction with English as the primary language. In contrast, students in the TBLT group carried out specially designed reading tasks, with the purpose of directing students' attention to the language. The reading materials used by the two groups were the same. Learners' gains were evaluated via a reading test composed of four reading passages related to military affairs. The test was administered before and immediately after the class. The results showed that the TBLT group significantly outperformed their TBLT counterparts. In a follow-up interview, the CBLT students reported that their previous military knowledge helped them comprehend the content while performing the task, even if they did not understand the language. The TBLT students, however, stated that the specially designed tasks directed their attention to the language, and they then used the language to construct new content knowledge.

In sum, all the above four studies suggest that the integration of TBLT and CLIL is attainable not just at the theoretical level but also in pedagogical perspectives. Lyster's (2015) study indicates the potential promising of using pedagogical tasks in CLIL to strengthen connections between language and content learning. Llinares and Dalton-Puffer's (2015) study provides insights into the role of task type in facilitating language learning in content-based contexts. García Mayo and Lázaro Ibarrola's (2015) study suggests that pedagogical tasks in TBLT induce negotiation of meaning for learners from both CLIL and EFL settings. Finally, Li and Chen's (2019) study attests to the effectiveness of specially designed SLA tasks in promoting language learning in subject matter content. These studies set the stage for further investigation of the connection between TBLT and content-based instruction.

Although the combination of the two approaches has been approached from several different perspectives in recent years, there is a clear lack of empirical research on how task design features influence L2 learning in content-based settings, which is crucial in order to create the optimal environment that maximizes L2 learning. In light of this, the present study aims to examine the effects of task design features on interaction and L2 performance in one specific content area (i.e., math). Among a variety of task features, the study specifically focuses on cognitive task complexity, which is closely associated with learner cognitive operations during task performance. The experimental tasks are mathematical word problems with two cognitive complexity levels operationalized along the Triadic Componential Framework (Robinson, 2007a). To obtain a comprehensive understanding of the research context, characteristics of mathematical language as well as word problems will be discussed in the following section.

2.2.4 Mathematical Register

Linguistically, the word "register" refers to the subset of a language (e.g., English), usually used for specific purposes. For example, everyday informal register is the language used for daily communication. Likely, mathematical register is a specialized language used to achieve communication in the field of mathematics. According to Halliday (1975), mathematical register is composed of (a) natural language words in mathematical context (e.g., sum, point), (b) mathematical expressions, such as right triangle, and (c) terms from combinations of words (e.g., output, denominator). Building upon Halliday's proposal, Dale & Cuevas (1987) further analyzed mathematical register in terms of vocabulary, syntax, semantics, and discourse features.

One of the difficulties in learning mathematics vocabulary is that the same mathematics operation can be signaled in a variety of ways (Monaghan, 1999). For example, students need to know diverse expressions for "addition" in mathematics, such as add, sum, plus, together, combine, and increased by. In the meanwhile, it is also crucial to provide specific mathematical context for students to learn the vocabulary. Teaching or learning these isolated items without context could lead to a lot of confusion. For example, the word "by" refers to multiplication in "a number multiplied by 3". However, in the expression "a number exceeds 3 by 4", "by" signals addition. Therefore, mathematical vocabulary should be introduced and taught with mathematical tasks, rather than in the form of word list. Another potential issue in teaching mathematics vocabulary, proposed by Dale & Cuevas (1987), is related to the symbolic language of mathematics. Although mathematical symbols are believed to be most commonly known by people around the world, it should be noted that some symbols are used differently in different countries for operations. For instance, a comma is used to separate hundreds from thousands in the United States (e.g., 300,000,000), while it is used to separate ten thousand from hundred thousand (3,0000,0000) in China.

The highly frequent use of comparative structures is another unique characteristic of mathematical language (Knight & Hargis, 1977). Expressions that signal relationships, such as "greater than" and "as…as", are potentially confusing to ELLs. "The lack of one-to-one correspondence between mathematical symbols and the words they represent" makes it even more difficult for ELLs to translate expressions to symbols correctly (Dale & Cuevas, 1987, p. 15). For example, the sentence "the number a is five times greater than the number b" could be very difficult for some students to successfully translate into the symbolic solution equation.

Another potential difficulty for ELLs, especially for young ELLs, is the common use of logical connectors (e.g., if, because, as a result) in the mathematical language (Jarrett, 1999). For example, if a is 0, then $ab = 0$. According to Piaget's (1970) Stages of Development, young learners before the stage of formal operational are unlikely to solve problems in a hypothetical situation.

There might also be semantic stress posed on ELLs in mathematical problems, mainly due to the difficulty in identifying the referents of the keywords (Dale & Cuevas, 1987). For example, in the sentence “the number a multiplied by itself five times equals itself”, students must understand that “the number”, “a”, and “itself” refer to the same variable.

In addition, written mathematics texts are complex and not always accurate and precise (Barwell, 2005). To solve mathematical problems, students are supposed to both master the relevant mathematics knowledge and have sufficient experience with English mathematical texts. Only in this way, could students conceptually understand the problem and then relate it to everyday background knowledge. However, due to the fact that most ELLs do not have a lot of opportunities to experience mathematical discourse in their regular ESL/EFL class, it is difficult for them to successfully comprehend mathematical texts.

2.2.5 Mathematical Word Problems

Word problems are one of the most commonly used mathematical tasks in schools around the world. However, word problems have also been viewed as the most difficult problem type by most students (Daroczy et al., 2015). According to Gerofsky (1996), a word problem usually has three components: a set-up component, an information component, and a questioning component. The set-up component provides the background of the “story”, such as time, location, and characters. This component is usually irrelevant to the solution. So sometimes the set-up component is embedded in the other two components. The information component is the key to solving the problem, which contains the essential information to the solution. Students must identify the relationship among the variables in order to solve the problem. The last component is the goal of the problem, which usually takes the form of a question. The following word problem reflects all the three components with the set-up component embedded in the information component.

Example 1

"Sally loaned $7.00 to Betty. But Sally borrowed $15.00 from Estella and $32.00 from Joan. Moreover, Joan owes $3.00 to Estella and $7.00 to Betty. One day the girls got together at Betty's house to straighten out their accounts. Which girl left with $18.00 more than she came with?" (Whimbey et al., 2013, p. 244)

The language use in word problems is generally concise and for calculation purposes only. Unlike other mathematical task types, there is usually no visual aid for word problems. Therefore, if a student has difficulty understanding some words or phrases, it is unlikely for him or her to perform the correct calculations. Daroczy et al. (2015) argued that "unfamiliar (low-frequency) words, polysemous words, idiomatic or culturally specific lexical references" (p. 4) are the main sources of comprehension difficulties for ELLs. To solve the problem in Example 1, the student not only has to know the meanings of "loaned", "borrowed", and "owed", but also the phrase "straighten out", which is an unfamiliar expression for a lot of ELLs. The syntactic structure of the question even further complicates the problem. The student may interpret the question as "which girl had $18.00" or "which girl had more money than she came with". The "she" in the question could also be very confusing. Since there are many girls' names in the problem, the student could be confused about which name "she" refers to.

To summarize, the linguistic challenges addressed above call for researchers' and educators' attention to the need of developing learners' mathematical language proficiency. Although CLIL programs have been developed to facilitate students' academic language learning with subject matter content, researchers have found that "the realities of CLIL implementation clearly emphasize its content-side" (Llinares & Dalton-Puffer, 2015, p. 70). A possible solution to this is to incorporate carefully designed pedagogical tasks of TBLT in CLIL contexts (Llinares & Dalton-Puffer, 2015; Lopes, 2020; Ortega, 2015). As such, understanding how task design features influence L2 learning in academic content has become of great importance for both language researchers and educators. Since task complexity has been viewed as a feature that contributes to the intrinsic cognitive demands of tasks (Robinson & Gilabert, 2012), and academic language development is highly related to content and cognitive demand (Cummins, 2001), the study aims to illuminate the effects of different cognitive task complexity on how learners engage and allocate their attentional resources while performing subject-specific tasks (i.e., mathematical tasks). In addition, in response to Ortega's (2015) call for technology-mediated task-based CLIL research, the study seeks to investigate academic language learning not only in face-to-face but also in synchronous computer-mediated communication (SCMC) environments.

2.3 Using Tasks in Synchronous Computer-Mediated Communication

Synchronous computer-mediated communication (SCMC) refers to network-based real-time conversation. While SCMC includes text-based, video-based, and voice-based communication, most of the previous research on TBLT and technology focuses on synchronous text-based interaction (e.g., online chat) probably due to its considerable resemblance to FTF interaction. Text-based SCMC is believed to be "the more reliable and economically feasible way to connect groups of learners" (Blake, 2009, p. 227). In this study, "SCMC" specifically refers to text-based SCMC.

In the early 1990s, researchers started to explore SCMC by examining classroom discourse in distance learning programs (e.g., Chun, 1994; Kelm, 1992; Kern, 1995). A body of research has revealed the benefits of SCMC in promoting learner participation (Kelm, 1992;

Kern, 1995; Warschauer, 1996), reducing task anxiety (Chun, 1998; Sauro, 2009), and facilitating the development of learners' interactive competencies (Chun, 1994; Darhower, 2002). Recently, the emphasis of the research has been shifted to interaction and focus on form in SCMC under an interactionist perspective (Long, 1996). This strand of studies is mainly task-based and has provided compelling evidence showing that SCMC is effective in generating opportunities for negotiation of meaning (e.g., Blake, 2000; Chen & Eslami, 2013; Eslami & Kung, 2016; Kung & Eslami, 2015, 2018; Pellettieri, 2000; Smith, 2003, 2004; Toyoda & Harrison, 2002; Yilmaz, 2011; Yilmaz & Granena, 2010).

Whereas both SCMC and FTF require real-time interaction (i.e., immediate response), SCMC might have some advantages over FTF communication (Chun, 1998). In comparing the efficacy of SCMC and FTF communication in facilitating L2 learning, researchers have proposed that SCMC supports a written record of the information and thus allows for more processing time, which in turn promotes comprehension and allows for more attention to the output (Chapelle, 2001; Smith, 2004). A number of empirical studies have been conducted to test this proposal (e.g., Chapelle, 2001; Lai & Zhao, 2006; Moradi & Farvardin, 2020; Payne & Whitney, 2002; Rouhshad et al., 2016; Yuksel & Inan, 2014).

For example, Lai and Zhao (2006) examined learners' noticing in the SCMC and the FTF modes with a spot-the-difference task. Twelve learners of English formed six mixed-proficiency dyads. Stimulated recall was used to identify instances of noticing immediately after task performance. The number of meaning negotiation occurrences was also calculated. The results showed that although there was significantly more meaning negotiation in the FTF interaction, a significantly higher percentage of noticing occurred (i.e., noticing of interactional feedback and noticing of self-errors) in the SCMC mode. Lai and Zhao found that online chat allowed learners to spend more time reviewing their L2 production, which then promoted learners' noticing of their own linguistic errors in SCMC. Yuksel and Inan (2014) replicated this study with 64 English learners and obtained similar results. Yuksel and Inan concluded that while FTF promoted negotiation of meaning, SCMC led to more noticing.

However, in contrast with the findings of the above two studies, Gurzynski-Weiss and Baralt (2014) also measured learners' noticing in interactional feedback with stimulated recall in FTF and SCMC modes but did not find any significant difference between the two groups. The study involved 24 intermediate learners of Spanish and the experimental task was an information-gap task, which required the learners to fill a living room or a kitchen with furniture. Although more time was found to be spent on the task in the SCMC mode, the increased time did not lead to more instances of meaning negotiation. The greater average performance time, however, was due to the dual-mode (i.e., reading, typing, and processing at the same time) feature of SCMC. Baralt and Gurzynski-Weiss asserted that it may just take longer for learners to read, decode, encode, and then type the messages in SCMC compared to FTF. Additionally, the results also revealed that learners modified their output significantly more often after receiving feedback in the FTF interaction than in the SCMC mode. This may be due to the fact that in SCMC, the feedback providers did not provide sufficient opportunities for their interlocutors to use the feedback for modified output as in the FTF interaction.

In a similar vein, Rouhshad et al. (2016) compared the frequency as well as the quality of meaning negotiations in terms of successful uptake and modified output in FTF and SCMC environments. The participants were 24 adult English learners in Australia. The results from the two decision-making tasks were aligned with Gurzynski-Weiss and Baralt's (2014) findings that task completion in SCMC required more time investment, but it did not elicit as many negotiations as in FTF interactions. Furthermore, the researchers also found fewer opportunities for successful uptake and output modifications in SCMC than in FTF mode.

More recently, Moradi and Farvardin (2020) considered whether the mode differentially impact oral interaction with mix-proficiency dyads. The study aimed to examine the frequency of meaning negotiation and modified output generated in FTF and SCMC interactions. Forty-five Iranian adult learners of English were divided into two group based their English proficiency: elementary group and upper intermediate group. Thirty-two of them were then randomly selected to form 16 mixed-proficiency dyads. Each dyad completed a jigsaw task and a spot-the-difference task. Their oral interactions were recorded and analyzed for meaning negotiation and modified output. The results showed that while the interaction mode did not have a significant effect on the frequency of meaning negotiation, it influenced the production of modified output with more modified output in SCMC.

To sum up, the findings from the existing literature suggest that (a) interaction mode (FTF vs. SCMC) impacts the frequency of meaning negotiation with more opportunities for negotiations in FTF (Gurzynski-Weiss & Baralt, 2014; Lai & Zhao, 2006; Moradi & Farvardin, 2020; Rouhshad et al., 2016; Yuksel & Inan, 2014), and (b) corrective feedback in FTF is more likely to trigger output modifications than in SCMC (Gurzynski-Weiss & Baralt, 2014; Rouhshad et al., 2016), with Moradi and Farvardin's (2020) study as an exception. However, the absence of uptake and modified output does not mean noticing does not happen in SCMC. Both Lai and Zhao (2006), and Yuksel and Inan (2014) found a higher percentage of noticing in negotiation instances using a direct measure of noticing - stimulated recall. Moreover, noticing of self-errors has been found to be greatly promoted in SCMC (Lai & Zhao, 2006) but it cannot be captured by uptake instances. For a more nuanced understanding of research evaluating interaction in SCMC and FTF environments, researchers should consider both types of noticing: noticing of feedback and noticing of one's own linguistic errors.

2.3.1 Task Complexity in SCMC

Although task-based interaction in SCMC mode has received great attention in recent decades, much of the research focuses on whether SCMC or FTF promotes L2 development based on their interactive features during task performance. There has been little research examining the influence of specific task design features in L2 learning. To date, only a limited number of studies have investigated the effect of task complexity in SCMC compared to FTF interactions.

Baralt (2013) investigated the interactive effects of task complexity and task modality on interaction (i.e., the efficacy of recasts) and the development of the Spanish past subjunctive. A two-level task complexity was manipulated along with [+/- intentional reasoning]. Eighty-four adult learners were divided into four groups: FTF + Complex, FTF – Complex, SCMC + Complex, SCMC – Complex. The learning outcomes were measured with a multiple-choice receptive test and two productive tasks (one in FTF and one in SCMC). A pre-, post-, and delayed post-test design was used.

The results showed that more cognitively complex tasks yielded more gains in FTF while simple tasks led to higher achievement in SCMC. A follow-up analysis on the qualitative analysis of interactions in each condition was carried out to examine why the effects of cognitive complexity were different in FTF and SCMC modes. Based on the interaction data, Baralt found that the SCMC – C group produced fewer and shorter turns. This was because the lower cognitive level in SCMC led to a slower communication pace, which allowed learners to draft messages based on the responses they received from their interlocutors. After receiving corrective feedback in relation to linguistic mistakes in their previously sent messages, learners oftentimes double checked their messages for any linguistic errors before sending the messages out. Therefore, it appeared to be that more time and more attentional resources were available for the learners to check the linguistic features in the SCMC – C condition.

In contrast, more turns with longer duration were observed in the SCMC + C condition. Most of the time, the learners were typing their messages when they received the feedback, which was often neglected. The fact that both the researcher and the learners were typing and sending their messages simultaneously made turn taking problematic and inefficient. Cognitive overload was found as the main cause of the learners' inaccurate L2 production.

However, the interaction in FTF – C condition worked quite differently from that in SCMC. In the FTF – C condition, little scaffolding was needed during task performance. The learners simply received the feedback, repeated it, and then moved on to the next point. Not a lot of attention was required in this setting. On the contrary, the FTF + C condition resulted in a much greater amount of scaffolding. Learners' attention was successfully directed to their erroneous utterances by the researcher's immediate feedback, which allowed them to compare and modify the output. In other words, higher cognitive loads in FTF forced the learners to both take in and apply the feedback.

One issue addressed in the study was the "transferability of L2 development from one mode to the other" (Baralt, 2013, p. 720). In this study, all groups demonstrated some gains in both the FTF and the CMC immediate post-test. However, the SCMC + C group was unable to maintain most of the gains in the FTF production task during the post-test, indicating the possible transferability issue of L2 development from written text to oral production. This finding was in contrast to Payne and Whitney's (2002) finding indicating that learners who spent half of their class time in the chatroom outperformed their FTF peers in the oral production assessment. The discrepancy in the findings of the two studies may be due to the difference in the intensity of the treatment sessions in the two studies. In Payne and Whitney's (2002) study, the learners met for a total of 21 online sessions across a 15-week semester, while in Baralt's (2013) study, the learners only met with the researcher four times in total, involving only two treatment sessions within two days.

Adams et al. (2015) also examined the role of task complexity in SCMC but on L2 production. In this study, task complexity was manipulated along [+/- task structure] as well as [+/- language support], which was defined as "…language-focused pre-task planning by pushing learners to explicitly consider language forms that would be useful as they carried out the task" (Adams et al., 2015, p. 69). Although language support was not specified in the Triadic Componential Framework, the researchers claimed that it resembled the pre-task planning in the framework. Therefore, both of the variables were considered as resource-dispersing demands. The participants were ninety-six undergraduate learners of English in Malaysia, who were divided into four groups: low task structure without language support (-TS, -LS), low task structure with language support (-TS, +LS), high task structure without language support (+TS, -LS), high task structure with language support (+TS, +LS). The students' online chat exchanges were collected and analyzed in terms of the accuracy and the complexity of L2 writing. L2 writing accuracy was assessed through mean errors per AS-unit, target-like use of auxiliary verbs, and modal verbs. Complexity was measured based on structural (i.e., clauses per AS-unit and words/turn) and lexical complexity (i.e., lexical frequency and the Guiraud index).

The results showed that the learners tended to produce less accurate L2 production when performing the more complex tasks, while no statistically significant impact on writing complexity was revealed. The findings partially supported the Cognition Hypothesis that increasing task complexity along resource-dispersing variables will lead to lower accuracy and lower complexity in L2 production. In order to better interpret the results, a follow-up interview was administered regarding learners' perception of the task performance and language learning.

During the interview, the participants reported that the task structure guided their writing and allowed them to focus on the language form in their output. One student mentioned that "… I think that made us more attentive to the language our teammates used. I mean it allowed us to be more conscious of others' language expressions" (Adams et al., 2015, p. 75). Similarly, providing pre-task language support also reduced the cognitive load of the task and fostered consciousness-raising in accuracy. With regard to complexity, the students reported that it was very difficult for them to maintain both accuracy and complexity, so the high cognitive demand pushed them to prioritize one over the other. Due to the rapid online information exchanges, the students were striving to catch up with others' turns, which resulted in simpler sentence structures and low lexical variety.

Taken together, the findings from the above two studies indicate that there are interactive effects of task complexity and modality on L2 learning in SCMC. Both studies found that learners could pay greater attention to linguistic forms while performing simple tasks than complex tasks in SCMC. This was probably attributed to the disassociation of message production and transmission in SCMC, which is unlikely to occur in FTF. Additionally, using more complex tasks in SCMC may cause cognitive overload, which may lead to confusion and to ignoring of corrective feedback. In this regard, the prediction of Robinson's Cognition Hypothesis that increasing cognitive task complexity leads to more attention to form may not be applicable in the SCMC condition.

2.4 Conclusion

This literature review offers insights into the analysis of learners' task performance depending on a variety of task complexity levels. It also highlights the importance of implementing independent measures of task complexity in research to verify the cognitive loads experienced by the learners. Even though a large number of research studies have examined the effect of task complexity on L2 learning, only a few studies have examined the validity of their task complexity manipulation by other measures. The lack of validation of task complexity could possibly lead to inaccurate results as well as invalid interpretation of research findings. Therefore, it is crucial for future research to consider using multiple measures of task complexity to obtain a more accurate understanding of how cognitive demands impact learners' task performance.

The review also addresses the research gap related to pedagogical task design in content-based language learning programs such as CLIL. As several researchers and educators have recently pointed out, well-designed tasks hold great promise for optimal linguistic learning in subject matter contents (Lopes, 2020; Lyster, 2015; Ortega, 2015). However, current research on content-focused programs mainly focuses on the overall effectiveness of the program based on students' academic achievements or the perceptions of the teachers and the students. Little attention has been paid to how the manipulation of task design features can promote L2 learning in a subject matter content.

Furthermore, in response to current trends in the increasing use of technology in language teaching and learning, the review of the current literature suggests an interactive effect of task complexity and task performance modality (i.e., FTF vs. SCMC) on L2 learning. However, how task complexity and modality interact with each other is still an under-researched area. More research is needed to extend the study of task design and its effect on L2 learning in SCMC settings.

To conclude, the review of the current literature indicates research gaps in the following three areas: (a) task complexity research with independent measures of cognitive load, (b) task design research in subject-matter contents, and (c) task design research in SCMC environments. To fill these gaps, the present study aims to examine the effects of cognitive task complexity and task modality on interaction and mathematical language performance. Self-ratings of task difficulty and retrospective time duration judgment are utilized as independent measures of task complexity.

CHAPTER III

EXPLORING THE LINKS BETWEEN TBLT AND CLIL:

THE EFFECTS OF TASK COMPLEXITY AND MODALITY ON INTERACTION

3.1 Introduction

How to design pedagogical tasks that maximize language acquisition has been a central topic in the field of second language acquisition (SLA). From a cognitive perspective, Robinson's (2001, 2011) Cognition Hypothesis predicts that increasing cognitive task complexity promotes interaction-driven L2 learning opportunities. Although this assumption has been tested by a number of researchers (e.g., Baralt, 2014; Kim, 2009; Kim et al., 2015; Révész, 2011), the majority of these studies were conducted in the face-to-face (FTF) mode. Little is known about how task complexity affects interaction in online environments.

Due to the increasing number of long-distance language learning programs, online communication has been widely used as a medium of interaction. Among various types of computer-mediated communication (CMC) (e.g., text-based, audio-based, video-based), text-based synchronous CMC (SCMC) has received the greatest attention from SLA researchers. It has been proposed that compared to FTF interaction, SCMC provides a written record of the text and thus allows for more processing time, which in turn promotes comprehension and allows for more attention to the output (Chapelle, 2001). This proposal has led to a fair number of investigations on language use in task-based SCMC. Accumulated empirical evidence has shown that SCMC amplifies learners' attention to linguistic features, and enhances "noticing" (Chapelle, 2001; Kitade, 2000; Pellettieri, 2000; Lai & Zhao, 2006; Long, 2007; Payne & Whitney, 2002; Yuksel & Inan, 2014). Although SCMC research has advanced significantly in the recent decades, little attention has been paid to the effects of cognitive task demand on interaction in SCMC. Since cognitive task complexity plays a critical role in L2 learning (Robinson, 2011), research is needed to investigate whether task complexity differently influences interaction in SCMC and FTF modes.

So far, most task-based research, either in FTF or SCMC environments, focuses on basic interpersonal language for daily communication. Nevertheless, the impact of task design features on academic language development has not yet been sufficiently addressed. Recently, the proposal of integrating TBLT to content and language integrated learning (CLIL) (Ortega, 2015) has brought researchers and educators' attention to pedagogical task design in content areas. Based on the premise that both CLIL and TBLT view the main aim of learning a language to be able to communicate and believe that language learning is achieved through interaction and collaboration, many researchers (e.g., Lopes, 2019; Meyer, 2010; Ortega, 2015; Shehadeh, 2018) have suggested the use of well-designed subject-specific tasks to facilitate language learning in content-based instruction such as CLIL.

To date, the examination of pedagogical tasks in content areas has just started. A limited number of relevant studies have provided empirical evidence showing that the combination of TBLT and CLIL not only reinforces the connection between language and content (Lyster, 2015) but also promotes negotiation of meaning (García Mayo & Lázaro Ibarrola, 2015), which is believed to be critical in L2 learning (Long, 1996). To continue this line of research, the present study aims to investigate the effects of cognitive task complexity on interaction in content-based learning (i.e., mathematics). To explore how task modality (i.e., the interaction environment) affects task outcome, the study focuses on conversational interaction in both FTF and SCMC environments.

3.2 Background Literature

3.2.1 The Cognition Hypothesis, Task Complexity, and Interaction

The Cognition Hypothesis proposed by Robinson (2001b) provides the theoretical foundation for task design and task sequencing from a cognitive perspective. Robinson (2001b) defines cognitive task complexity as the "attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner" (p. 29), and asserts that cognitive task complexity should be the main factor to consider in task sequencing. Later, in his Triadic Componential Framework, Robinson (2007a, 2011) further distinguishes between resource-directing and resource-dispersing dimensions for task complexity manipulation. Resource-directing variables such as [+/- few elements] manipulate meaning-form mappings. Robinson (2011) hypothesizes that increasing task complexity along resource-directing variables can "result(s) in greater attention to, and uptake of, forms made salient during the provision of reactive focus on forms techniques such as recasts" (p.18). In contrast, resource-dispersing variables such as [+/- planning time] operate on learners' accessibility to their interlanguage system, but do not account for new linguistic knowledge construction. Increasing task complexity along this dimension disperses learners' attention from the linguistic aspects of the target language and thus leads to less accurate and complex L2 production.

Stimulated by the Cognition Hypothesis, a substantial body of research has investigated the role of task complexity in L2 learning. In particular, researchers have examined the effects of task complexity on interactional moves such as recasts (i.e., target-like reformulation of an erroneous utterance by the interlocutor), clarification requests (i.e., the listener's requests to clarify the interlocutor's certain utterances), confirmation checks (i.e., the listener's repetition of the interlocutor's certain utterances for confirmation), and comprehension checks (i.e., the speaker checks on the interlocutor's understanding) (e.g., Hardy & Moore, 2004; Kim et al., 2015; Nuevo, 2006; Révész, 2011; Robinson, 2007b).

For instance, Hardy and Moore (2004) examined how task structure and content familiarity affected learner-learner interaction. Fifty-six college students who were learning German as an L2 performed computer-supported tasks (i.e., watching a video clip in German and evaluating the characters in the video). Learners' interaction was assessed based on the frequency of conversational negotiation episodes (e.g., clarification requests, agreements, repairs) and the interrelation of these episodes (i.e., the average number of episodes within one topic). The results showed that increasing task complexity by reducing the degree of structural support significantly promoted meaning negotiation. However, no significant effect was found for task familiarity on interaction.

Nuevo (2006) also explored the relationship between task complexity and interaction-driven L2 learning opportunities. In this study, task complexity was manipulated along [+/- reasoning demands]. One hundred and thirteen learners performed a picture narration task and a decision-making task. L2 learning opportunities were analyzed based on nine different interactional moves including recasts, clarification requests, confirmation checks, comprehension checks, hypothesis formulation, self-repairs, metalinguistic talks, noticing a linguistic deficiency (i.e., when learners state that they do not know a specific expression in the L2), and other repetitions (i.e., imitation). The results showed that learners generated more uptake of recasts, more comprehension checks, and more other repetitions in the simple task. Nevertheless, they verbalized their hypotheses about the L2 more often while performing the complex task. In other words, learners were more likely to check whether their utterances were correct with their interlocutors in the complex task. No significant effects of task complexity were detected on other interactional measures. The mixed findings suggested that the influence of task complexity varied depending on the type of the interactional measures.

Robinson (2007b) investigated the impact of task complexity along [+/- intentional reasoning] on confirmation checks, the number of turns taken, clarification requests, and uptake. Forty-two Japanese undergraduate students who were learning English as an L2 completed picture story tasks with three different cognitive complexity levels. The results were in accordance with the Cognition Hypothesis as increasing task complexity along the resource-directing variable led to significantly more turns, confirmation checks, clarification requests, and partial uptake (i.e., learners used the target form but omitted, substituted, or added other elements).

Révész (2011) examined the role of task complexity in the occurrence of language related episodes (LREs) as well as the exchanges within the LREs (i.e., confirmation checks, clarification requests, recasts, and metalinguistic talk). Decision-making tasks with different cognitive loads along [+/- reasoning] and [+/- few elements] were administered to 43 English adult learners. The results did not show significant differences for confirmation checks, clarification requests, and recasts, but it revealed a significantly greater amount of LREs and metalinguistic talk in the more cognitively demanding task.

Kim and Taguchi (2015) investigated the effect of task complexity on pragmatics development (i.e., request-making expressions). A pre-, post-, and delayed post-test design was used in this study. Task complexity was manipulated along [+/- reasoning]. Seventy-three Korean learners of English were assigned to one of the following groups: the control group, the simple group, and the complex group. The control group took the pre-, post-, and delayed post-tests only. Learners in the other two groups performed one version of the treatment task (collaborative) based on which group they were assigned to, and took the pre-, post-, and delayed post-tests. Their oral interaction during task performance was recorded and coded for pragmatics related episodes (PREs). The results revealed a strong positive impact of greater task complexity on certain PRE targets (i.e., contextual features, head acts, and preparators) but not on other PRE targets (i.e., grounders, hedges, and amplifiers). In terms of learning outcomes, the results did not demonstrate any significant differences between the simple and the complex task group in the post-test. However, only the complex task group maintained their gains in the delayed post-test, suggesting that task complexity might have a long-term effect on L2 pragmatic development.

In a more recent study, Solon et al. (2017) explored the relationship between task complexity and pronunciation related LREs with 34 (17 dyads) learners of Spanish. In this study, two task complexity levels were operationalized along [+/- few elements]. Each dyad performed both the simple and the complex versions of an information gap task. The results showed that there was no significant effect of task complexity on pronunciation related LREs.

Taken together, previous task complexity studies fail to provide conclusive support for the Cognition Hypothesis, which claims that learners pay more attention to linguistic forms while performing more cognitively demanding tasks. Although Robinson (2007b) found that there certainly was an increase in the interactional moves in the more complex task, other studies (e.g., Hardy & Moore, 2004; Nuevo, 2006; Révész, 2011; Solon et al., 2017) found varied effects of task complexity on different interactional moves. The inconsistency in the findings suggests a need for further research on the relationship between task complexity and interaction. One of the goals of the present study, therefore, is to elucidate how task complexity operates on interaction-driven L2 learning opportunities.

3.2.2 Task-based Interaction in Synchronous Computer-Mediated Communication

With the advancement of technology, there has been a trend towards the incorporation of technology in classroom teaching in recent decades. Given its great resemblance to FTF communication but in a written mode, synchronous computer-mediated communication (SCMC) in the form of written conversation has received great attention from language researchers (e.g., Aoki, 1995; Chen & Eslami, 2013; Chun, 1998; Eslami & Kung, 2016; Gurzynski-Weiss & Baralt, 2014; Kern, 1995; Lai & Zhao, 2006; Warschauer, 1996). Research that compares classroom discourses in FTF and SCMC settings shows that learners are more willing to participate in discussions and to interact with others in SCMC than in FTF communication (e.g., Chun, 1998; Kern, 1995; Warschauer, 1996). Furthermore, SCMC reduces learners' anxiety and facilitates interaction by building a sense of learning community (Aoki, 1995; Warschauer, 1996).

Another advantage of SCMC over FTF proposed by researchers is that the written presentation of the discourse in SCMC amplifies learners' attention to linguistic forms and enhances "noticing the gap" (Chapelle, 2001; Kitade, 2000; Pellettieri, 2000; Long, 2007; Payne & Whitney, 2002). A few studies have empirically tested this claim within task-based research contexts. (Baralt, 2013; Eslami & Kung, 2016, 2018; Gurzynski-Weiss & Baralt, 2014; Lai & Zhao, 2006; Moradi & Farvardin, 2020; Rouhshad et al., 2016; Yuksel & Inan, 2014). For example, Lai and Zhao (2006) examined L2 learning opportunities in SCMC and FTF interaction. In this study, twelve learners of English formed six mixed-proficiency dyads to perform a spot-the-difference picture task. Learners' oral conversation and online chats were collected and coded for recasts, negotiation of meaning (i.e., indicators of non-understanding), and self-initiated repair. A stimulated recall session was carried out subsequently to identify instances of noticing. The results showed that although there were significantly more instances of meaning negotiation in the FTF interaction, SCMC yielded a significantly higher percentage of noticing, especially learners' noticing of their own linguistic mistakes.

Gurzynski-Weiss and Baralt (2014) measured learners' noticing of interactional feedback as well as opportunities for modified output following the feedback in FTF and SCMC interactions. The researchers used stimulated recall as the measure of noticing in this study. Twenty-four intermediate learners of Spanish were divided into two groups based on the interaction mode and performed an information-gap task. The results revealed that learners in both modes were able to notice and accurately perceive the feedback in a similar manner. However, learners modified their output significantly more often after receiving feedback in FTF interaction than in SCMC. The researchers also found that oftentimes, the feedback providers did not offer sufficient opportunities for their interlocutors to modify their output in SCMC. Based on the findings, Gurzynski-Weiss and Baralt argued that the greater amount of processing time available in SCMC did not guarantee more attention to linguistic forms. The extra time may be just used for reading, decoding, encoding, and typing messages due to the dual-mode feature (i.e., reading and writing) of SCMC.

Rouhshad et al. (2016) examined the nature of negotiations generated in FTF and SCMC interaction. The researchers further distinguished between negotiation for meaning and negotiation for form in the study. Twenty-four adult learners of English with intermediate proficiency levels formed 12 dyads. All dyads performed two similar decision-making tasks, one in FTF and one in SCMC setting. Learners' oral conversation and online chats were collected and coded for each type of negotiation as well as interactional moves within negotiation episodes - successful uptake (i.e., successful uptake of the feedback) and modified output. The results showed that FTF interaction induced a significantly higher rate of negotiations for meaning than SCMC, although learners spent more time on the task in SCMC. No significant difference was detected for negotiation for form across modes. In addition, higher percentages of successful uptake and modified output were obtained in FTF interaction.

More recently, Moradi and Farvardin (2020) considered whether the mode differentially impacts oral interaction with mix-proficiency dyads. The study aimed to examine the frequency of meaning negotiation and modified output generated in FTF and SCMC interactions. Forty-five Iranian adult learners of English were divided into two groups based on their English proficiency: elementary group and upper-intermediate group. Thirty-two of them were then randomly selected to form 16 mixed-proficiency dyads. Each dyad completed a jigsaw task and a spot-the-difference task. Their oral interactions were recorded and analyzed for meaning negotiation and modified output. The results showed that while the interaction mode did not have a significant effect on the frequency of meaning negotiation, it influenced the production of modified output, with more modified output in SCMC.

In sum, previous research on interaction in FTF and SCMC settings suggests that FTF communication appears to be more effective in generating negotiation of meaning. It seems to be the case that the extra processing time available in SCMC allows learners to review previous messages in case of incomprehension, which eliminates the need for negotiation (Rouhshad et al., 2016; Smith, 2009). However, one cannot simply claim that SCMC is less effective in facilitating L2 learning. It is worth noting that except for Lai and Zhao's (2006) study, the other two studies only focus on noticing triggered by the interlocutor's feedback. Learners' noticing of their own linguistic errors was not taken into consideration. Based on Lai and Zhao's (2006) study, one can speculate that while FTF interaction promotes negotiation of meaning, SCMC enhances self-initiated noticing. However, there is a need for studies to empirically test this hypothesis. The present study thus aims to investigate negotiations and (self- and other-triggered) noticing in FTF and SCMC modes.

3.2.2.1 Task Complexity and the Interaction Mode

Despite a handful of studies that investigate learners' focus on form in SCMC, there have been very few attempts to explore the effects of cognitive task complexity on learners' attention allocation in this context. To date, there have been only two empirical studies that examine how task complexity impacts interaction and L2 development in SCMC. Baralt (2013) examined how task complexity and task modality moderated the efficacy of recasts with 84 adult learners of Spanish. Participants were divided into five groups depending on the interaction mode and task version (i.e., control, FTF-C, FTF+C, SCMC-C, SCMC+C). Each participant carried out a story-retell task with the researcher. Occurrences of recasts received during task performance were collected and analyzed. Language learning outcomes were measured by pre- and post-assessments. The results showed that the more cognitively demanding task carried out in the FTF mode led to the most learning. In the SCMC mode, performing the cognitively simple task led to more learning. Baralt observed that communication in the tasks with higher cognitive demands in the online environment was challenging and difficult for learners. As the SCMC environment allowed learners to type and send messages simultaneously without waiting time, both interlocutors were trying to explain their complicated ideas with long sentences. "Split negotiation routines" (i.e., the primary clause and its subordinate clause are split in nonadjacent turns) occurred frequently. Learners were trying hard to collect all the pieces of information in order to comprehend their interlocutors, paying little attention to the linguistic features in the target language. In a more recent study that examined the effects of task complexity across online and FTF modes, Adams, Alwi, and Newton (2015) also reported that while conducting the more complex task, learners were striving to catch up with their interlocutors' turns without paying a lot of attention to the linguistic form.

Despite progress in understanding how task complexity may operate on learners' attention to linguistic forms in SCMC, empirical studies are still limited. Comparative research on the effects of task complexity on interaction across online and FTF modes will help develop a better understanding of the potentials of the interaction in online vs. FTF mode in L2 development. Furthermore, it will also shed light on whether the predictions of the Cognition Hypothesis are confirmed for L2 learning in online environments. The present study, therefore, attempts to contribute to this line of research by examining how task modality moderates the effects of task complexity on negotiations and noticing.

3.2.3 Exploring Tasks in Content-based Learning

So far, task-based research, either in FTF or SCMC environments, has mainly focused on basic interpersonal language skills for daily communication. Recently, the proposal of integrating TBLT to content and language integrated learning (CLIL) (Ortega, 2015) has brought researchers and educators' attention to the use of pedagogical tasks (developed in TBLT) in content areas. CLIL is "a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language" (Coyle et al., 2010, p. 1). Despite the dual foci of CLIL, researchers and educators have seen a clear emphasis on the content over the language in CLIL implementation phase (Llinares & Dalton-Puffer, 2015). In many CLIL programs, lessons are taught by content specialists who may or may not possess qualifications for language teaching (Dalton-Puffer, 2011). Moreover, assessment in CLIL classes is mainly driven by the content (Linares & Dalton-Puffer, 2015). Such practices may lead to reduced devotion to L2 learning by students participating in CLIL programs. Blasco (2014), for example, found that there was a gradual decrease of accuracy from learners aged 9-10 throughout a two-year CLIL study period, although fluency and syntactic complexity were improved.

In response to the issues raised in the implementation of CLIL, researchers and educators have proposed the combination of TBLT and CLIL to promote language learning parallel to content learning (Lopes, 2019; Ortega, 2015; Shehadeh, 2018). Although there are several differences in terms of target age group, educational purpose, and learning context between the two approaches, the premise that meaning and form are inseparable is highly significant in the principles of both TBLT and CLIL (Ortega, 2015). Furthermore, in both approaches, language is viewed as a means of communication, which can be developed through interaction and collaboration. In light of these shared features, researchers argue that the use of pedagogical tasks (developed in TBLT) in content-based learning could facilitate attention to linguistic forms (Lopes, 2019; Lyster, 2015). This argument, however, needs to be tested by empirical research. To date, the examination of pedagogical tasks in content areas and their effectiveness in promoting both language and content learning has just started. Limited empirical evidence has shown that the combination of TBLT and CLIL not only reinforces the connection between language and content (Lyster, 2015) but also promotes negotiation of meaning (García Mayo &Lázaro Ibarrola, 2015). These studies set the stage for further investigation of the connection between the two approaches. To continue this line of research, the present study aims to examine the effects of task design features on interaction-driven L2 learning opportunities in content-based learning. Among a variety of task features, the study specifically focuses on cognitive task complexity, which is closely associated with learner cognitive operations during task performance (e.g., Baralt, 2013).

The experimental tasks are mathematical word problems with two cognitive complexity levels operationalized based on the Triadic Componential Framework (Robinson, 2007a). The reasons why mathematical word problems are chosen as the experimental tasks are that (a) learners' language proficiency appears to be a significant factor that impacts their mathematical achievement (NCELA, 2017), and (b) word problems are perceived as the most difficult and as cognitively demanding problem type by most students (Daroczy et al., 2015).

Following previous research (e.g., Sato & Takatsuka, 2016; Swain & Lapkin, 2001), LRE and self-initiated repair are used as the measures of interaction in the present study. LREs are short dialogues where learners "talk about the language they are producing, question their language use, or correct themselves or others" (Swain & Lapkin, 1998, p. 326). An LRE contains three parts: trigger, response, and uptake (optional). It indicates possible communication breakdown and opportunities for meaning negotiation. While an LRE may contain self-corrections, it is mainly driven by interactional moves. Previous research has shown that the interaction mode has different effects on different types of modified output (i.e., self-corrections with or without prompts from interlocutors) (Lai & Zhao, 2006). In order to distinguish self-corrections without having triggers by interlocutors from self-corrections in LREs, the present study involves self-initiated repairs (i.e., immediate self-correction that do not require responses from interlocutors) as a measure of focus on form in addition to LREs. The following example illustrates the two types of self-corrections.

Example 2

A: Did you go the party yesterday?

B: Yes, I **go… went** (*self-initiated repair*) to the party. What about you?

A: No, I didn't. I had an assignment to turn in last night.

B: Oh, I **turn in** my assignment first.

A: **What do you mean?**

B: I mean I **turned in** (*self-correction prompted by the interlocutor*) my assignment before going to the party.

A: Oh, that's great.

The following research questions directed the study:

1. How does cognitive task complexity affect interaction-driven L2 learning opportunities, as measured by LREs and self-initiated repairs, during the mathematical problem-solving process in the FTF mode?

2. How does cognitive task complexity affect interaction-driven L2 learning opportunities, as measured by LREs and self-initiated repairs, during the mathematical problem-solving process in text-based online chats?

3. How does task modality (i.e., FTF vs. SCMC) affect interaction-driven L2 learning opportunities, as measured by LREs and self-initiated repairs, across cognitive task complexity?

3.3 Method

3.3.1 Participants

A total of 82 college students (55 males, 27 females) consisting of 41 English native speakers (NSs) and 41 English non-native speakers (NNSs) were included in this study. The participants were undergraduate or graduate students studying at a public university in the United States. Their ages ranged from 18 to 33 (M= 25.57, SD= 3.55). The participants were majoring in various fields. The NSs were from social science areas such as education, psychology, international studies, communication, and sociology. The NNSs were from science or engineering areas such as mechanical engineering, computer engineering, biochemistry, and chemistry. Among the NNSs, 37 participants reported being native speakers of Mandarin. The remaining four NNSs were native speakers of Korean (n=1), Japanese (n=1), Turkish (n=1), and Kazakh (n=1). Based on their responses in the background questionnaire, the NNSs' English proficiency was at a high-intermediate level (average iBT TOEFL = 96.67, SD= 4.55, range 86 - 102). All the NNSs used technologies in English extensively for a variety of purposes, such as social networking and text messaging.

3.3.2 Design

Each NNS was paired with an NS to form 41 NS-NNS dyads. The rationale for forming NS-NNS dyads is that NS-NNS interactions have been found to be more effective in eliciting meaning negotiation opportunities, corrective feedback, and modified output than NNS-NNS interactions in both FTF and SCMC modes (Iwashita, 2001; Mackey et al., 2003). The dyads were then divided into two groups: the FTF group (n=21) and the SCMC group (n=20). Each dyad worked on two versions of a mathematical task, a simple version and a complex version, under FTF or SCMC condition. To eliminate the potential order effects, half of the dyads performed the simple task first and the other half performed the complex task first. The independent variables in the study were task complexity (within-subject) and task modality (between-subject). The dependent variables were a series of interactional measures such as frequency counts of self-initiated repairs and LREs.

3.3.3 Materials

3.3.3.1 Background Questionnaire

The background questionnaire contained three sections. The first section asked for participants' demographic information (e.g., age and gender). The second section included open-ended questions asking NNSs' English learning experiences (e.g., how many years they have been learning English?). The last section was related to the NNSs' experience of using English in an online environment (see Appendix A).

3.3.3.2 Experimental Tasks

The experimental tasks were two mathematical word problems designed by the researcher based on the literature (e.g., Whimbey et al., 2013) (see Appendix B). Task complexity was manipulated along [+/- few steps] following Robinson's (2007a) Triadic Componential Framework. Robinson argues that pedagogical tasks that require more steps to complete are more cognitively demanding than tasks that require fewer steps. This assertion is in compliance with the Cognitive Load Theory in math education, in which Sweller (2010) claims that the intrinsic cognitive load (i.e., the information that must be understood to solve the problem) of a math problem increases when the level of element interactivity is raised. Element interactivity refers to the extent to which individual elements in the problem are connected to each other. In this regard, a problem in need of a more complex problem-solving procedure indicates a higher level of element interactivity and thus is more cognitively demanding.

Both word problems asked participants to calculate the time needed for two characters to complete a certain task (i.e., planting flowers or painting a room). The main differences lied in whether the work efficiency was already provided in the problem and the number of incidents that occurred while completing the task in the story. In the simple task, the total work and work efficiency were provided, and a fewer number of incidents were involved in the story. This task thus was algorithmic and mainly focused on describing a simple procedure that was presented in the task. In contrast, the complex task required participants to work out the efficiency first based on the given information, and then calculate the final answer by considering a more complicated storyline. The complex task required more steps to complete than the simple task. The design of the experimental tasks was supported by Smith and Stein (1998)'s taxonomy of mathematical tasks based on the cognitive demand of the task. In the taxonomy, mathematical tasks that are algorithmic and involve procedures without connections to the concepts that underlie the procedure, such as the simple task in the study, are considered to have lower cognitive demands. Mathematical tasks that require a deeper understanding of abstract concepts and involve processes analyzing what needs to be done and how, such as the complex task in the study, are considered to have higher cognitive demands.

3.3.3.3 Task Difficulty and Time Judgment Questionnaire

Although the effect of task complexity on L2 learning has been extensively examined, most studies have not validated their task complexity manipulation with empirical evidence. It has been suggested that task complexity operationalization should be empirically evaluated with independent measures of cognitive load (e.g., Norris & Ortega, 2003). Given that many of the measures used in recent task complexity studies are adopted from the field of psychology (e.g., time duration judgment) and have not been largely investigated in the field of SLA, it is difficult to determine whether the measure is appropriate for task complexity validation (Sasayama, 2016). One of the contributions of the present study, therefore, is to gauge the cognitive loads of the experimental tasks with two different measures of task complexity – self-ratings of task difficulty and retrospective duration judgment.

In this study, a task difficulty and time judgment questionnaire was used to validate the cognitive demands of the experimental tasks (see Appendix C). The questionnaire asked for the participants' perception of task difficulty as well as the estimated time they spent on the task. Previous research has shown that learners perceive more complex tasks as more difficult, and the estimated task completion time was longer than the actual performance time (Baralt, 2013; Robinson, 2001b). The questionnaire was originally designed by Baralt (2013) for Spanish learners, and it was adapted in this study for English learners. The questionnaire consisted of two parts. The first part asked for the estimated time spent on the task. The second part was composed of 15 items regarding the degree of task difficulty. The participants were asked to rate each statement on a 6-point Likert scale with 1 as strongly disagree and 6 as strongly agree. The internal consistency of the second part of the questionnaire was 0.82 for the simple task and 0.87 for the complex task, indicating a good consistency of the items in the questionnaire.

To validate task complexity, a pilot study (N=20) was conducted before the present study. The participants in the pilot study were undergraduate and graduate NSs and NNSs from a public university in the US. The participants performed the two tasks (in FTF) and reported on the level of task difficulty for each task. The results showed that participants perceived the complex task (M=4.7, SD=1.06) as more difficult than the simple task (M=1.6, SD=0.70), supporting the validity of the level of task complexity. Although the task complexity level was validated in the pilot study, to ensure the validity of task complexity level in both FTF and SCMC modes, we also asked the participants involved in each task to evaluate the task difficulty level and estimated perceived time spent on the task to further validate the complexity of the tasks used in this study in each mode.

3.3.4 Procedure

The participants in the study were recruited by sending emails to students in the university email lists and classroom visits. In the recruiting email, a brief introduction of the study, criteria for potential participants, and contact information of the researcher were provided. The potential participants contacted the researcher directly through email. In addition to that, the researcher also talked about the study with professors who were teaching the related courses. Upon professors' permission, the researcher visited the classroom, explained the study, and recruited potential participants.

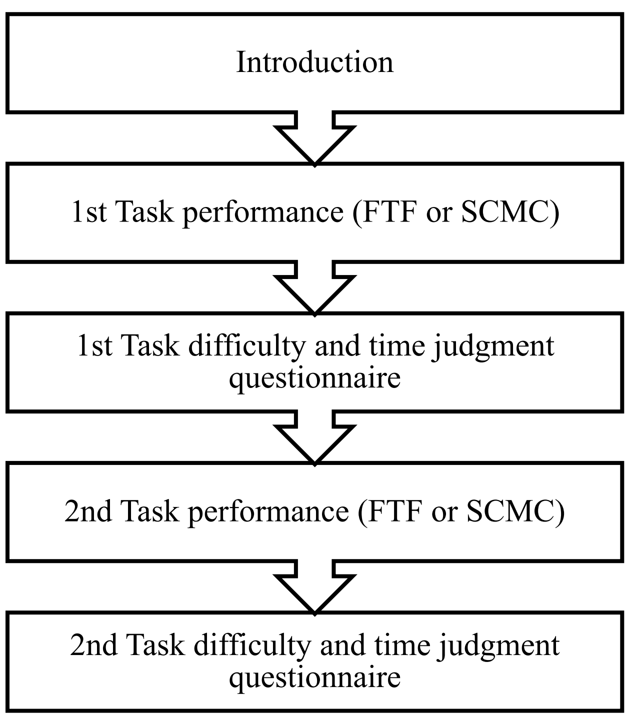
Through the above processes, a total of 90 participants, consisting of 47 English NSs and 43 NNSs, signed up for the study. The researcher then sent out the background questionnaire via Qualtrics for them to fill out. Four people did not complete the questionnaire so they were removed from the study. Each NS was then randomly paired with an NNS to form an NS-NNS dyad for task completion. In the task completion session, another four participants were absent from the meeting. They, therefore, were removed from the study. Finally, 82 participants (41 dyads) were included in the study. All participants were recruited on a voluntary basis. Each participant was paid a $10 or $15 Starbucks gift card, depending on the length of the session, as compensation for their participation.

All phases of data collection were completed within one meeting session. The session lasted 1 to 1.5 hours. The procedure of the data collection session is illustrated in Figure 4. The session started with the researcher briefly introducing the study and presenting the consent form to the participants to sign. The participants then introduced themselves to their partners. After that, the researcher shared the experimental task materials and instructions with the participants. The participants then started the first task, through FTF or online interaction. The

FTF dyads discussed the word problem by staying in the same room facing each other. Their interactions were recorded with digital audio-recorders. The researcher was present in the room during the process to take observational notes and coordinate the session. On the other hand, participants in the SCMC group stayed in separate rooms and communicated via the chat box in Google Hangouts. Their chat histories were recorded. The researcher stayed with the NNSs to take observational notes. Immediately following completion of the first task, the participants filled out the Task Difficulty and Time Judgment Questionnaire. After that, they continued to work on the second task and completed the questionnaire again based on their perception of the second task.

Figure 4

The Procedure of Data Collection in the Study



3.3.5 Data Coding

The audio recordings from the FTF interactions were first transcribed and then coded for the occurrence of LREs. Chat histories from the SCMC group were also coded for the occurrence of LREs. Once LREs were identified, they were coded for a variety of characteristics. The coding scheme was adapted from Loewen's (2005) analysis of focus-on-form episodes (FFEs), which was defined as "a brief, spontaneous focus on a linguistic item within the context of a meaning-focused task" (p. 363). Table 4 presents the coding scheme for LREs used in the present study. Furthermore, self-initiated repair was defined as occurrences when learners correct their own linguistic errors without having triggers by their interlocutors. Self-initiated repairs indicate learners monitoring their output and thus play an important role in prompting modified output (e.g., Sato & Takatsuka, 2016; Shehadeh, 2001).

Table 4

Characteristics of LREs

Characteristics	Definition	Categories
Type	When an LRE is instigated	• Reactive: Error correction • Preemptive: NNS-initiated query
Linguistic focus	Linguistic target	• Grammar • Vocabulary • Pronunciation / Spelling
Source	The reason to instigate an LRE	• Code: Inaccurate use of linguistic item with no apparent miscommunication • Message: Problem understanding meaning

Table 4 Continued

Characteristics	Definition	Categories
Complexity	Length	• Simple: Only one response move • Complex: More than one response move
Directness	Explicitness of feedback	• Indirect: Implicit (e.g., recast or repetition) • Direct: Explicit (e.g., metalingual explanation)
Emphasis	Combination of complexity and directness	• Light: Indirect and simple • Heavy: Direct, complex, or both
Response	Type of feedback provided by the NS	• Provision: NS gives information about a language form • Elicitation: NS attempts to draw out from NNS a language form or information about a language form • Provision & Elicitation
Uptake	NNS response to feedback	• Uptake: NNS produces response • No uptake: NNS produces no response
Successful uptake	Quality of student response	• Successful uptake: NNS incorporates linguistic information into production or shows solid evidence of understanding • Unsuccessful uptake: NNS does not incorporate linguistic information into production

The example in Table 5 illustrates the coding system used in the study. This episode occurred during one of the calculation processes. The participants were discussing how to calculate the work efficiency in the problem. The NNS's inaccurate language use caused nonunderstanding of the NS interlocutor. Therefore, this LRE is *reactive* on the part of the NNS. The linguistic focus of this LRE is related to language use (*vocabulary)*. Meanwhile, the problematic linguistic item impeded comprehension. Thus, the source is *meaning*. Moreover, it is a *complex* LRE because it required more than one response move. The NS used implicit corrective feedback - clarification request and recast, which makes the feedback *indirect*. However, given that the LRE is complex, the emphasis was considered as *heavy*. The NS first used clarification request (*Change to what?*) to elicit linguistic information from the NNS, and then used recast (*change the place of them, put the time on the top*) to provide the correction, without directly indicating that the NNS' utterance was inaccurate. Therefore, the response was coded as *elicitation & provision*. Finally, the NNS not only responded to the NS' feedback, but also incorporated the linguistic information in his production, suggesting an occurrence of successful uptake.

Table 5

Example of LRE Coding

<table>
<tr><td rowspan="10">NNS: I mean we need to change this two to two places. Yeah. Change it. Yeah. Change this, these two numbers.
NS: Change to what?
NNS: Change these two numbers
NS: Change to what though?
NNS: Change what though?
NS: You said to change these numbers.
NNS: Yeah. change these two numbers, 420 and 7 because…
NS: To what?
NNS: Because we needed to calculate how many time the Alex will spend to paint one square.
NS: Yes, I know. But you said to change this. But to what? What should we change this to?
NNS: Change these two place.
NS: Oh, just change the place of them?
NNS: Yes!!
NS: So, you want to put the time on the top?
NNS: Yes, the time on the top.
NS: OK.</td><td>Characteristics</td><td>Category</td></tr>
<tr><td>Type</td><td>Reactive</td></tr>
<tr><td>Linguistic focus</td><td>Vocabulary</td></tr>
<tr><td>Source</td><td>Meaning</td></tr>
<tr><td>Complexity</td><td>Complex</td></tr>
<tr><td>Directness</td><td>Indirect</td></tr>
<tr><td>Emphasis</td><td>Heavy</td></tr>
<tr><td>Response</td><td>Elicitation & Provision</td></tr>
<tr><td>Uptake</td><td>Uptake</td></tr>
<tr><td>Successful uptake</td><td>Successful uptake</td></tr>
</table>

The researcher coded all the LREs (n= 237) that were generated from the NS-NNS conversations. In order to investigate the reliability of the coding, about 20% of the samples (n = 48) were coded by another coder. The kappa coefficients for the LRE characteristics ranged from k = .68 to .93, and the inter-coder agreement scores were above 85%, indicating substantial agreement between the coders.

3.3.6 Statistical Analysis

First, raw frequencies and percentages of the conversational moves (i.e., self-initiated repair, LRE, and LRE characteristics) were calculated to provide an overview of self-initiated repair and LRE production. Then, to account for the differences in time on task across task versions, rates of self-initiated repair and LRE by dyad per minute were calculated. Outliers were defined as a score that is three standard deviations away from the mean. All the outliers were identified and eliminated from the sample for inferential statistical analysis. There was one outlier for task difficulty ratings, one for duration judgment, and one for interaction measures. The data were normally distributed and met the conditions for use of parametric tests. Paired *t*-tests, repeated-measures ANOVAs, and factorial ANOVAs were conducted, with the significance level set at $\alpha = .05$. Partial η^2 was calculated to measure effect size. Partial η^2 values of .01, .06, and .14 were considered small, medium, and large, respectively (Baralt, 2013). All the inferential statistics were performed in STATA.

3.4 Results

Prior to answering the research questions, the manipulation of task complexity was examined by analyzing the scores from Task Difficulty and Retrospective Time Judgment Questionnaire. Average self-perceived task difficulty ratings for each task and average subjective-to-objective (i.e., estimated-to-actual) time duration ratios for each task were first calculated. Task difficulty ratings and subjective-to-objective ratios were then compared between the simple and complex tasks. The results showed that learners rated the complex task (M = 39.23, SD = 12.77) as more difficult than the simple task (M = 29.65, SD = 9.57). Results from the paired *t*-test revealed that the differences were significant, $t\ (39) = 5.65$, $p < .001$. In relation to time judgment, the average ratio of the complex task was larger than that of the complex task (1.03 vs. 0.94), although the difference was not significant, $t\ (39) = 1.42$, $p = 0.16$. According to Block et al. (2010), increasing cognitive demands would lead to a higher subjective-to-objective ratio in retrospective duration judgment. Taken together, the results confirmed the validity of task complexity manipulation.

3.4.1 RQ1. How does Cognitive Task Complexity Affect Interaction-Driven L2 Learning Opportunities during the Mathematical Problem-Solving Process in the FTF Mode?

Results for the measures of L2 learning opportunities during the simple and complex tasks in the FTF mode are listed in Table 6. In order to account for the differences in time upon completion of the two tasks, rates of self-initiated repair and LRE production for each dyad per minute on each task were calculated. The mean rates for self-initiated repair and LRE between the simple and complex tasks were compared using repeated measures ANOVAs. Table 7 presents the descriptive and inferential statistics for the rates of self-initiated repair and LRE (in terms of linguistic focus) across task versions.

The results showed that learners produced more self-initiated repair instances in the complex task (n=21) than the simple task (n=9). Nevertheless, the rates of self-initiated repair were the same across task versions. As for LREs, there were a greater amount of LREs observed in the complex task (n=86) than the simple task (n=47). However, once again, the rates of LREs across task demands were similar. Although the differences in the rate of LREs across task demands were not statistically significant, the results revealed a large effect of task complexity on LRE production (partial $\eta^2 = 0.13$).

Similar distributions of LRE characteristics were found between the two versions of the task. For both tasks, 94% of the LREs were triggered by NNSs' erroneous utterances, and 6% of the LREs were a result of NNS-initiated query. Grammar-focused LRE was the most common LRE type produced on both tasks, followed by vocabulary-focused LRE. Pronunciation-focused LRE was the least frequent linguistic foci in LREs. As for the effect of cognitive demands on the linguistic focus of the LREs, a significantly higher rate of grammar-focused LRE was obtained in the complex task, with a large effect size, $F(1, 20) = 4.45$, $p = 0.04$, partial $\eta^2 = 0.18$. No significant effects were found for the other two types of LREs (i.e., vocabulary-focused and pronunciation-focused LREs).

The majority of the LREs were initiated for the purpose of correcting inaccurate use of linguistic forms on both tasks. The complex task (18%) generated a higher percentage of negotiation episodes (LREs) caused by NNSs' erroneous utterances than the simple task (9%). For both tasks, about 90% of the LREs were short and only required one response move. The feedback provided by the NSs was generally implicit and had a light emphasis on linguistic forms. The main purpose of corrective feedback was to provide information about linguistic forms. Furthermore, increasing cognitive demands did not lead to differences in the rate of uptake. However, it appeared to affect the production of successful uptake. The complex task yielded a higher rate of successful uptake (42%) than the simplex task (29%).

Table 6

Learning Opportunities in the FTF Mode across Task Versions

		FTF_S		FTF_C	
Measures		Amount	Percentage	Amount	Percentage
Self-initiated repair		9	N/A	21	N/A
LRE		47	N/A	86	N/A
Type	Reactive	44	94%	82	94%
	Preemptive	3	6%	5	6%
Linguistic	Grammar	30	64%	53	62%
	Vocabulary	11	23%	21	24%
	Pronunciation	6	13%	13	15%
Source	Code	43	91%	72	84%
	Message	4	9%	15	18%
Complexity	Simple	42	89%	80	93%
	Complex	5	11%	7	9%
Directiveness	Direct	4	9%	13	15%
	Indirect	43	91%	74	86%
Emphasis	Light	41	87%	74	86%
	Heavy	6	13%	13	15%
Response	Provision	43	91%	78	91%
	Elicitation	5	11%	11	13%
Uptake	Uptake	42	89%	78	91%
	No uptake	5	11%	9	10%
	Successful	12	29%	33	42%
	Unsuccessful	30	71%	45	58%

Table 7

The Number of Self-initiated Repairs and LREs (in terms of linguistic focus) per Minute during Completion of the Two Tasks in the FTF Mode

	Self-initiated Repair M (SD)	Grammar-focused LRE M (SD)	Vocabulary-focused LRE M (SD)	Pronunciation-focused LRE M (SD)	LRE (Total) M (SD)
FTF_S	0.05 (0.07)	0.17 (0.15)	0.06 (0.08)	0.03 (0.08)	0.27 (0.21)
FTF_C	0.05 (0.09)	0.13 (0.14)	0.05 (0.08)	0.03 (0.06)	0.21 (0.20)
Sig.	$p = 0.51$	$p = 0.04$	$p = 0.18$	$p = 0.74$	$p = 0.09$
Effect	$\eta^2 = 0.02$	$\eta^2 = 0.18$	$\eta^2 = 0.09$	$\eta^2 = 0.01$	$\eta^2 = 0.13$

3.4.2 RQ1. How does Cognitive Task Complexity Affect Interaction-Driven L2 Learning Opportunities during the Mathematical Problem-Solving Process in the SCMC Mode?

The overall number and percentage of self-initiated repair and LRE details during the text-based online chats across cognitive conditions are shown in Table 8. The descriptive and inferential statistics for the rates of self-initiated repair and LRE are presented in Table 9. The results revealed more incidences of self-initiated repair in the complex task (n=7) than the simple task (n=3). Nevertheless, the rates were the same across cognitive conditions. A similar trend was found for the production of LREs. While the total number of LREs in the complex task (n=64) was greater than that in the simple task (n=40), the differences in rates were not statistically significant. The effect size for LRE, however, appeared to be large (partial $\eta^2 = 0.13$).

LREs were similarly distributed across types and linguistic focus on the simple and the complex tasks. For both tasks, the majority of the LREs started with NNSs' linguistic errors. Only a small portion of the LREs was initiated by NNSs' query. For both tasks, the predominant linguistic focus of the LREs was grammar, followed by vocabulary. None of the LREs was related to spelling. Results from the repeated measures ANOVAs showed that changes in cognitive demands did not lead to significant differences in the linguistic focus of the LREs. However, a large effect size was found for grammar-focused LRE, $F(1, 18) = 3.70$, $p = 0.07$, partial $\eta^2 = 0.17$.

In both tasks, the main cause or source of the LREs was associated with linguistic accuracy rather than miscommunication. The results showed that the percentage of language-related communication breakdown and negotiations increased along with cognitive loads. There were a higher number of LREs negotiating forms in the complex task (23%) compared with the simple task (10%). The LREs were short in length across task versions. A higher percentage of more explicit feedback, with a heavier emphasis on linguistic forms, was observed in the complex task. As for the function of feedback, a higher percentage of the feedback was used to elicit learners' linguistic knowledge in the complex task (20%) than the simple one (2%). There was a lower rate of uptake in the complex task compared with the simple task. However, increased task complexity led to a much higher rate of successful uptake, 45% in the complex task compared to 13% in the simple task.

Table 8

Learning Opportunities in the SCMC Mode across Task Versions

		SCMC_S		SCMC_C	
Measures		Amount	Percentage	Amount	Percentage
Self-initiated repair		3	N/A	7	N/A
LRE		40	N/A	64	N/A
Type	Reactive	38	95%	62	98%
	Preemptive	2	5%	1	2%
Linguistic	Grammar	34	85%	51	80%
	Vocabulary	6	15%	13	20%
	Spelling	0	0%	0	0%
Source	Code	36	90%	49	77%
	Message	4	10%	15	23%
Complexity	Simple	40	100%	61	95%
	Complex	0	0%	3	5%
Directiveness	Direct	7	18%	20	31%
	Indirect	33	82%	44	69%
Emphasis	Light	33	82%	46	72%
	Heavy	7	18%	18	28%
Response	Provision	39	98%	51	80%
	Elicitation	1	2%	13	20%
Uptake	Uptake	31	78%	44	69%
	No uptake	9	22%	20	31%
	Successful	4	13%	20	45%
	Unsuccessful	27	87%	24	55%

Table 9

The Average Number of Self-initiated Repairs and LREs (in terms of linguistic focus) per Minute during Completion of the Two Tasks in the SCMC Mode

	Self-initiated repair M (SD)	Grammar-focused LRE M (SD)	Vocabulary-focused LRE M (SD)	Spelling-focused LRE M (SD)	LRE (Total) M (SD)
SCMC_S	0.01 (0.01)	0.08 (0.03)	0.02 (0.01)	0.00 (0.00)	0.10(0.03)
SCMC_C	0.01 (0.02)	0.06 (0.04)	0.01(0.01)	0.00 (0.00)	0.07 (0.05)
Sig.	$p = 0.82$	$p = 0.07$	$p = 0.36$	N/A	$p = 0.13$
Effect	$\eta^2 = 0.002$	$\eta^2 = 0.17$	$\eta^2 = 0.05$	N/A	$\eta^2 = 0.12$

3.4.3 RQ 3. How does Task Modality (i.e., FTF vs. SCMC) Affect Interaction-Driven L2 Learning Opportunities across Cognitive Task Complexity?

Table 10 summarizes the descriptive statistics for L2 learning opportunities during task completion in different modes. As can be seen, more incidences of self-initiated repair and LRE were generated in FTF interaction than in SCMC, regardless of the cognitive condition. Results from mixed-design ANOVAs confirmed the above observation that the mode of interaction (FTF vs. SCMC) had significant effects on self-initiated repairs, $F(1, 38) = 12.18$, $p = 0.001$, partial $\eta^2 = 0.24$, and LREs, $F(1, 38) = 13.71$, $p < 0.001$, partial $\eta^2 = 0.27$.

The descriptive statistics showed that uptake was produced at a higher rate in FTF mode than in SCMC mode across task complexity levels. This observation was aligned with the ANOVA results, which indicated a significant main effect of mode on uptake, with a large effect size, $F(1, 38) = 6.11$, $p = 0.002$, partial $\eta^2 = 0.16$.

Whereas the mode of interaction appeared to be the main factor that affected self-initiated repair, LRE, and overall uptake, the results for successful uptake were mixed. The simple task completed during the FTF interaction generated a higher rate of successful uptake (29%) than in the SCMC interaction (13%). On the other hand, the complex task in the FTF mode yielded a lower rate of successful uptake (42%) than that in the SCMC mode (45%). To statistically examine the relationship between task complexity, mode, and successful uptake, a mixed-design ANOVA was carried out. The results showed that main effect was found only for task complexity, with a large effect size, $F(1, 38) = 7.18$, $p = 0.01$, partial $\eta^2 = 0.22$. No significant effects were found for the mode of interaction. These results suggest that a higher rate of successful uptake is associated with increased cognitive demands in the task, rather than with the interaction mode.

Table 10

Descriptive Statistics for the Rate of Self-Initiated Repair, LRE, Uptake, and Successful Uptake across Modes and Cognitive Conditions

	Self-initiated repair		LRE (Total)		Uptake (%)		Successful Uptake (%)	
	M	SD	M	SD	M	SD	M	SD
FTF_S	0.05	0.07	0.27	0.21	89	16	29	30
FTF_C	0.05	0.09	0.21	0.20	91	41	42	36
SCMC_S	0.01	0.01	0.10	0.03	78	25	13	27
SCMC_C	0.01	0.02	0.07	0.05	69	33	45	38

The interaction mode had varied effects on the other LRE characteristics. No significant effect was detected for the type, source, and emphasis of the LREs. The linguistic focus of LREs were significantly impacted by the mode with a higher percentage of grammar-focused LREs generated in the SCMC interaction, $F(1, 38) = 7.22$, $p = 0.01$, partial $\eta^2 = 0.16$ and a higher percentage of vocabulary-focused LREs, $F(1, 38) = 7.18$, $p = 0.01$, partial $\eta^2 = 0.22$, and pronunciation/spelling-focused LREs, $F(1, 38) = 4.78$, $p = 0.04$, partial $\eta^2 = 0.11$, in the FTF interaction. It was also found that the feedback induced in SCMC was more explicit than that in the FTF interaction, $F(1, 38) = 5.78$, $p = 0.02$, partial $\eta^2 = 0.16$. Moreover, the results also revealed significant interaction effects between task complexity and the mode on the length of the LREs, $F(1, 38) = 5.62$, $p = 0.02$, partial $\eta^2 = 0.16$, and the purpose of the feedback in the LRE (i.e., provision or elicitation), $F(1, 38) = 6.71$, $p = 0.01$, partial $\eta^2 = 0.19$. Increases in task complexity led to a higher percentage of simple LREs (i.e., LREs that only involved one response move) in the FTF interaction, but the opposite was found for the SCMC mode. In addition, increasing task complexity resulted in a higher percentage of feedback that aimed to elicit linguistic information from the NNS in the SCMC mode, but no effect was found in the FTF mode.

In summary, the results showed that increasing task complexity led to higher rates of meaning-focused LREs (in comparison to code-focused LREs, see Table 4 for more details) and successful uptake in both FTF and SCMC interactions. In addition, the complex task in SCMC induced higher percentages of explicit feedback, modified language elicitation moves, and LREs related to grammar. The results also suggested that changes in cognitive demands did not lead to differences in the rates of self-initiated repair and LRE, regardless of the mode. However, cognitive task complexity had significant effects on successful uptake: increases in task complexity led to higher rates of successful uptake in both modes. With regard to the effect of the interaction mode, there were higher rates of self-initiated repair, LRE, and uptake produced in the FTF mode than in the SCMC mode across task complexity levels. However, the interaction mode did not have any significant effect on successful uptake.

3.5 Discussion

Drawing on Robinson's (2011) Cognition Hypothesis, the present study aimed to examine how task complexity affects L2 learning opportunities in content areas across two different communication modes (FTF & SCMC). In particular, the study explored the effects of cognitive demands on the production of LRE and self-initiated repair during the mathematical problem-solving process in FTF and SCMC interactions. Two levels of task complexity (i.e., low and high) were operationalized along the [+/- few steps] variable in the resource-dispersing dimension based on the Triadic Componential Framework (Robinson, 2007a). In order to ensure that the two experimental tasks indeed posed different cognitive demands on the participants, self-ratings of task difficulty and retrospective time judgment were employed to validate task complexity manipulation. The results from both measures confirmed that the complex task was perceived as more cognitively demanding than the simple task. To examine NS-NNS interaction, LREs were identified, counted, and coded based on Loewen's (2005) analysis of focus-on-form episodes (FFEs). Learners' self-initiated repair instances were also collected as indicators of incidental focus on form. Mean rates of self-initiated repair, LREs (i.e., per dyad per minute), and interactive moves within LREs (e.g., successful uptake) were calculated and compared between the simple and complex tasks using repeated measures and mixed-design ANOVAs.

3.5.1 The Effects of Task Complexity on FTF Interaction

The first research question was concerned with the effects of task complexity on FTF interaction in the mathematical content area. The results showed that there were no significant differences in mean rates of self-initiated repair and LRE across cognitive conditions. Nevertheless, the complex task yielded a higher rate of successful uptake (42%) than the simple task (29%). Overall, the results suggested that learners paid greater attention to linguistic forms while they were solving the more complex word problem in the FTF mode. The results partially support the Cognition Hypothesis, which claims that increasing cognitive task complexity promotes negotiation of meaning, noticing, and uptake of more salient input in feedback (Robinson, 2011).

The possible explanation as to why changes in task complexity did not lead to differences in LREs could be the nature of the experimental tasks in this study – mathematical word problems. Unlike information-gap and jigsaw tasks, which require a great deal of information exchange between task performers and thus promote meaning negotiation, interaction in problem-solving tasks is not mandatory (Pica et al., 1993). In other words, it is the participants' choice to interact or remain silent. In the case of the mathematical word problems in the present study, it is likely that once one of the interlocutors knew how to solve the problem, the other person just listened to him/her. As long as the participants could understand each other, they did not feel obligated to negotiate. This may have mitigated the impact of task complexity on interaction. As Hsu (2020) has claimed, "the lack of the need to exchange information may entail the lack of the need to negotiate, thus again leading to the lack of complexity distinctiveness between the two tasks" (p. 11).

The results also showed that task complexity did not have significant effects on self-initiated repairs, suggesting that changes in cognitive task complexity did not affect the NNSs' noticing of their own linguistic errors in the FTF mode. This is possibly attributable to the academic nature of the experimental tasks. Previous research has indicated that learners' previous subject knowledge can help them with task performance, even if they lack the related language knowledge (Li & Chen, 2019; Usó-Juan, 2006). It is likely to be the case that the NNSs in the FTF group were mostly concerned with the academic content and paid little attention to their L2 output while performing both tasks. Although the complex task was more cognitively demanding and more challenging to perform, the NNSs' previous mathematical knowledge may have compensated for their linguistic deficiency. This implies that learners' previous subject knowledge could be a factor that moderates the effect of task complexity on learners' noticing of their own linguistic errors in the FTF mode.

Finally, the results showed that the more cognitively demanding task induced a higher rate of successful uptake in FTF interaction. In the Cognition Hypothesis, Robinson (2003) proposes that increasing task complexity results in more communication breakdowns and thus promotes learners' noticing of more salient input in the interlocutor's feedback. However, as task complexity did not affect the rate of LREs in the present study, the reason why attention to feedback was enhanced along with increased task complexity remains unclear. It is possible that the NNSs realized that the complex task was more complicated and could easily cause misunderstanding, so they paid greater attention to their interlocutors' responses, which in turn facilitated noticing of feedback.

3.5.2 The Effects of Task Complexity on SCMC Interaction

The second research question tested the effects of task complexity on text-based online interaction in the mathematical content area. Once again, there were no significant differences in mean rates of self-initiated repair and LRE between the simple and complex tasks. However, increased task complexity resulted in a much higher rate of successful uptake in the complex task (45%) than in the simple task (13%). These findings imply that increasing task complexity fosters attention to linguistic forms in content areas during text-based online interaction, which again partially supports the Cognition Hypothesis. The findings, however, are not aligned with Baralt's (2013) study, in which the researcher found that more noticing of feedback occurred in the cognitively simple task in the SCMC mode. Baralt argues that increased task complexity in SCMC interaction may cause cognitive overload. She found that split negotiation routines occurred frequently during the performance of the complex task in the SCMC mode, especially when learners and their interlocutor (i.e., the researcher) were sending or typing messages simultaneously. Consequently, corrective feedback was often delayed and missed during the conversation. However, in the present study, a different pattern emerged. That is, turn-taking was mainly sequential in the complex task in SCMC mode.

Although long turns and split sentences also occurred frequently in the present study especially when learners were trying to explain the reasons behind some calculations in the complex task, the researcher observed that participants were usually waiting for their interlocutors to finish explaining their thoughts before typing or sending their responses. When an erroneous utterance was produced, the NSs waited for the learners to finish explaining, and then provided corrective feedback. Learners usually read the responses first, and then compared them with the messages they just sent. This is demonstrated in the following example (from the complex task performance):

Example 3

NNS: No, it should be much less than that.

NNS: The total area of room should be 12*35.

NNS: For now, we have already painted 372.

NNS: So, the work left should be 12*35 – 372, right?

NS: Yes, you are right. I made a mistake in my math. It should be 48 inches leftover.

NNS: It's OK. 2 digits multiple can be difficult sometimes.

NNS: 48 is the correct number.

NS: I would say that "2 digits being multiplied can be difficult".

NS: Is there a way to find the exact time?

NNS: Thanks! I didn't come up with that phrase.

NNS: Sure, we can find the exact time.

In the above example, the NNS used four consecutive turns to explain his thoughts, during which the NS did not interrupt him but rather waited until the NNS finished the explanation. After the NS provided feedback with the correct use of the language, the NNS responded to her by notifying his comprehension of the feedback. As can be seen, there were not that many overlaps, and turn-taking was mainly sequential. This might partially attributable to the "typing indicator" function provided in Google Hangouts Chat. If one person is typing, the other person will see “xxx (the person’s name) is typing…” on his/her end. This may reduce the chance when both interlocutors were sending or typing messages simultaneously.

Another explanation for the differences in the interaction pattern (i.e., turn-taking and overlaps) between Baralt’s (2013) study and the present study are probably due to different characteristics of the experimental tasks. The more complex task (i.e., a picture-story task) in Baralt’s study required learners to reflect on the intentional reasons that explained the main characters’ actions in the story. In this sense, the task is an “open” task, meaning that it allows for several possible outcomes in meeting the task goal (Pica et al., 1993). As such, it might be difficult to anticipate what an interlocutor wants to say. On the other hand, the complex task in the present study was a mathematical word problem, which is considered as a “closed” task because it permits a single outcome – the time required for the characters to finish painting the room. Although there might be different ways to solve the problem, the use of specific mathematical formulas was required by the problem (e.g., the total work divided by the time token equals work efficiency). Participants’ previous subject knowledge may help them anticipate and comprehend the content, which made more attentional resources available for noticing of feedback.

This explanation is aligned with the researcher’s observation that the NNSs spent a lot of time typing messages while performing the complex task. Oftentimes, they scrolled up to view previously provided feedback, and then rephrased their message. This indicates that learners tended to be more cautious about what they sent, whether the language was accurate and precise, in the complex task than in the simple task. It might be the case that the simple task did not require as much analysis and explanation as the complex task. Thus, learners may not need to pay as much attention to both the content and the language while performing the simple task.

This explanation is in accordance with the finding that there was a lower percentage of successful uptake in the simple task. In contrast, interlocutors on both sides might be aware that the complex task was more cognitively demanding and could easily cause confusion. Therefore, they were more careful about their messages and linear organizations of turns during the problem-solving process.

3.5.3 The Role of the Interaction Mode

The third research question investigated the potential interactive effects between task complexity and modality. Overall, the results did not show significant interaction effects between the two variables on the number of self-initiated repairs and LREs. The findings suggest that task modality and task complexity play critical roles in noticing of one's own linguistic errors and noticing of feedback, respectively. Significantly more self-initiated repairs and LREs were yielded in the FTF mode than in the SCMC mode. Increasing task complexity led to higher rates of successful uptake irrelevant of the interaction mode.

The findings, however, need to be interpreted with caution. Although lower rates of self-initiated repair were found in online chat histories, it should not be simply interpreted that learners rarely noticed their linguistic errors in the SCMC mode. It is possible that self-initiated repairs were not fully captured in the chat history. As previously mentioned, the researcher noticed that many NNSs spent a lot of time typing messages. Oftentimes, the NNSs typed a few words, paused, deleted the words, and then rephrased them before sending out the message. It is very likely that the NNSs may have been monitoring and self-correcting their language while typing, but this could not be captured by the chat history. This is supported by Moradi and Farvardin (2020), who have claimed that "the participants were more cautious of their grammatical mistakes in the SCMC mode than the FTF mode and they tried to revise their messages repeatedly before sending them" (p. 10). This can also explain why there were fewer LREs generated in the SCMC interaction (N = 104) than in the FTF interaction (N = 133), even though the duration of task performance in the SCMC mode was longer than that in the FTF mode. Since the NNSs took time to modify their L2 output before the messages were sent, there may have been fewer linguistic errors for negotiation episodes to occur. As Smith (2004) has noted, the extra processing time provided by SCMC facilitates comprehension and promotes accurate L2 production. Close observations of learners' reflection process and practices during the SCMC interaction are needed in future research to obtain a better understanding about how task complexity influences incidental focus on form in online environments.

Although the study provides initial empirical evidence related to the effects of task complexity and task modality on NS-NNS interaction in subject matter L2 learning context, it is not without limitations. First, the present investigation measured learners' incidental focus on form using LRE and self-initiated repair. Although self-initiated repair and successful uptake in LREs have been argued to be strong indicators of noticing and interlanguage development (Loewen, 2005), it does not clearly depict learners' internal processes. Given that learners may disperse their attention over many aspects of task performance (Kormos & Trebits, 2011; Leow, 2012), there is a need for implementing more direct measures of incidental noticing, such as stimulated recall and eye-tracking technology, in future research (Kim et al., 2015). Second, the study only focused on the interactional moves during task performance rather than the actual L2 learning outcomes. Future investigations adopting language assessments to examine whether task complexity would lead to the actual academic L2 development in different modes are warranted. Third, the research setting in the present study may limit the generalization of the findings.

Given that the research is not classroom-based, caution should be taken when applying the findings to classroom settings (Robinson, 2011).

3.6 Conclusion

In response to calls for the integration of TBLT and CLIL (Lopes, 2019; Ortega, 2015), the present study examined the effects of task complexity and task modality on NS-NNS interaction in content areas (while solving a math problem). The study extends the scope of task complexity research and provides empirical evidence with regard to whether the predictions of the Cognition Hypothesis can be applied to academic language learning in content areas. Additionally, the study contributes to the literature showing that task modality plays an imperative role in facilitating interaction-driven L2 learning opportunities. Overall, the results indicate that increasing task complexity led to a higher rate of successful uptake in both FTF and online modes during mathematical problem-solving processes. Moreover, FTF interaction appeared to better facilitate self-initiated focus on form than text-based SCMC. The findings provide some pedagogical implications regarding task design in content-based instruction. First, the study suggests that cognitive task complexity can be operationalized to enhance learners' attention to linguistic forms in subject matter classrooms. Second, the interaction environment (i.e., FTF vs. SCMC) should be carefully selected as it plays an important role in L2 learning in content areas. The study also offers some methodological implications for future research. First, multiple measures of cognitive loads should be incorporated for triangulation purposes and to validate task complexity manipulation. Second, when investigating cognitive operations in text-based online environments, it is recommended to record the screen as well as learners' whole interaction processes. Different types of concurrent data would allow researchers to picture learners' internal processes more accurately.

CHAPTER IV

THE EFFECTS OF TASK COMPLEXITY AND MODALITY ON MATHEMATICAL LANGUAGE PERFORMANCE

4.1 Introduction

Tasks, as the basis of Task-based language teaching (TBLT), have been extensively studied for their facilitative role in second language (L2) development. Researchers have mainly focused on the criteria with which tasks could be classified and sequenced. Two main frameworks – Robinson's (2011) Cognition Hypothesis and Skehan's (1998) Trade-Off Hypothesis – provide different accounts as to how tasks should be sequenced in order to obtain desired changes in linguistic complexity, accuracy, and fluency (CAF).

While the Cognition Hypothesis (Robinson, 2011) adopts Wickens' (1989) multiple and non-competitive attentional resource model, the Trade-Off Hypothesis (Skehan, 1998, 2009) argues for a single and limited attentional resource model (Cowan, 2001). The different theoretical foundations of the two hypotheses have led to different predictions about the effect of cognitive demands on L2 performance. Robinson (2001b) argues for a facilitative role of increasing task complexity in improving both the complexity and the accuracy of learners' L2 production. In contrast, Skehan (1998) claims that manipulating task complexity can lead to greater linguistic complexity or accuracy but not both simultaneously.

The contrasting predictions of the Cognition Hypothesis and the Trade-Off Hypothesis have promoted extensive empirical investigations regarding the role of task complexity in L2 learning in recent decades (e.g., Gilabert, 2007a, 2007b; Ishikawa, 2007; Robinson, 1995; Sasayama & Izumi, 2012; Tavakoli, 2009; Yuan & Ellis, 2003). Nevertheless, among these task complexity studies, only a few have used a computer-mediated communication (CMC) setting (e.g., Adams et al., 2015; Baralt, 2013; Baralt & Leow, 2016; Hsu, 2020). With the advancement of technology, increasingly more researchers have begun to explore L2 learning in online settings using CMC. Findings from previous research indicate the promising potential for L2 development in SCMC and call for investigation regarding the association between task design features and L2 learning opportunities in SCMC (e.g., Coniam & Wong, 2004; Eslami & Kung, 2016; Kung & Eslami, 2018, 2019; Fuente, 2003; Yanguas, 2012; Ziegler, 2016).

Although there have been a considerable number of studies that have examined how task complexity influences L2 learning in either FTF or SCMC setting, the effects of cognitive task complexity in content-based instruction (CBI) have beenbarely explored. CBI or content-based language teaching (CBLT) refers to "an approach to language instruction that integrates the presentation of topics or tasks from subject matter classes (e.g., math, social studies) within the context of teaching a second or foreign language" (Crandall & Tucker, 1990, p. 187). Academic language, the type of language that is closely associated with subject matter learning. is commonly used in CBI. Cummins (1981) has made clear distinctions between basic interpersonal conversational skills (BICS) and cognitive academic language proficiency (CALP). While the former refers to language skills used in daily communication, the latter includes language skills used to comprehend and construct academic knowledge. It is suggested by Cummins (1984) that compared to BICS, CALP requires additional context-embedded language learning opportunities. In this sense, CBLT facilitates the development of CALP with a rich and meaningful context that integrates subject knowledge and language learning.

Despite the fact that the CBLT model provides an ideal learning environment for CALP, an issue has been pointed out concerning the actual implementation of CBLT that there has been a clear emphasis on content over the language in content-based language learning programs

(Dalton-Puffer, 2011; Llinares & Dalton-Puffer, 2015; Lyster, 2017; Sato & Loewen, 2019). As the use of tasks has shown to be highly effective in L2 learning, language researchers and educators (e.g., Lyster, 2015; Ortega, 2015) have suggested the incorporation of pedagogical tasks in content areas. As such, investigating task design in content areas has become of great importance in the field of CBLT. Although there have been a few studies that examine the use of pedagogical tasks in CBLT (e.g., García Mayo & Lázaro Ibarrola, 2015; Lyster, 2015), none of these studies have looked into the role of task design in L2 learning in subject matter content. To address this and the above-mentioned research gaps, the present study seeks to examine the influence of cognitive task complexity on learners' L2 performance (i.e., complexity, accuracy, and fluency) in FTF and SCMC conditions in content-based language learning.

4.2 Background Literature

4.2.1 Task Complexity and L2 Performance

Robinson (2001b) defines task complexity as "the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner" (p. 29). In other words, task complexity is mainly dependent on the inherent design features of the task and influences learners' attentional resource allocation. Robinson's (2001a, 2001b, 2003, 2011) Cognition Hypothesis has led to a series of predictions regarding how task complexity affects attention deployment during task performance.

Built upon the multiple attentional resource model (Wickens, 1989), Robinson believes that learners' attention is not in competition while processing content and linguistic forms during task performance. The increase in cognitive task complexity can promote learners to pay more attention to the input as well as their output (i.e., self-monitoring). Consequently, greater accuracy could be found in learners' L2 production. In addition, increasing cognitive task complexity along certain dimensions (e.g., +/- few elements) induces greater functional demands on learners, which promotes the process of conceptualization and proceduralization, leading to greater complexity in L2 production (Robinson, 2001a).

Robinson (2007a) further distinguishes between resource-directing (e.g., here-and-now) and resource-dispersing dimensions (e.g., planning time) in his Triadic Componential Framework related to task complexity. Within this framework, Robinson argues that while resource-directing variables manipulate form-meaning mappings, resource-dispersing variables influence learners' access to the interlanguage system. Thus, when task complexity is increased along resource-directing variables, more cognitive resources are engaged for the acquisition of the target form, which leads to more accurate and complex L2 production. In contrast, when task complexity is raised along resource-dispersing variables, more cognitive resources are engaged to comprehend the content rather than analyzing aspects of linguistic forms, which reduces complexity, accuracy, and fluency.

Informed by the Cognition Hypothesis, a large body of research has been conducted to examine the impact of task complexity on learners' L2 performance (e.g., Frear & Bitchener, 2015; Gilabert, 2007a; 2007b; Ghanbarzadeh & Gholami, 2014; Kim & Payant, 2017; Revesz, 2011). A complexity, accuracy, and fluency (CAF) model is widely used as the primary measure of learners' L2 performance in most studies. Complexity refers to the extent to which the language is intricate and varied during task performance (Ellis, 2003). Accuracy entails learners' ability to produce error-free utterances in L2. Fluency is defined as "the production of language in real time without undue pausing or hesitation" (Ellis & Barkhuizen, 2005, p. 139). Additionally, Skehan (2009) suggests the use of independent measures of lexical complexity and syntactic complexity as two distinct aspects of complexity in the CAF model. In the following section, I will critically review some pertinent empirical studies that focus on the relationship between cognitive task complexity and learners' L2 performance. Since task complexity has been operationalized in a variety of ways, I have selected a few representative studies that involved the most common task complexity variables in the existing literature (Sasayama, 2015). Table 11 summarizes the design features as well as the research findings of the reviewed studies.

Table 11

A Summary of the Reviewed Studies on Task Complexity and L2 Production

Studies	Modality	Task complexity variable	Dimension	Findings (Increasing task complexity leads to…)
Robinson (1995)	Oral	+/- here-and-now	Resource-directing	+ Complexity + Accuracy - Fluency
Yuan & Ellis (2003)	Writing	+/- planning time	Resource-dispersing	- Syntactic complexity Mixed results for lexical complexity, accuracy, and fluency
Gilabert (2007a)	Oral	+/- here-and-now	Resource-directing	- Complexity + Accuracy + Fluency
		+/- planning time	Resource-dispersing	x Syntactic complexity - Lexical complexity x Accuracy - Fluency
*Gilabert (2007b)	Oral	+/- here-and-now	Resource-directing	Mixed results for accuracy
		+/- few elements	Resource-directing	+ Accuracy
		+/- reasoning	Resource-directing	Mixed results for accuracy

Table 11 Continued

Studies	Modality	Task complexity variable	Dimension	Findings (Increasing task complexity leads to…)
Sasayama & Izumi (2012)	Oral	+/- planning time	Resource-dispersing	- Syntactic complexity x Accuracy + Fluency
		+/- few elements	Resource-directing	+ Syntactic complexity - Accuracy - Fluency

Note. "+" indicates a positive effect of increasing task complexity; "-" indicates a negative effect of increasing task complexity; "x" indicates no significant effect of task complexity; "mixed" indicates mixed effect of increasing task complexity. *Gilabert (2007b) only measured accuracy.

Among the studies on task complexity and L2 performance, one of the most frequently investigated resource-directing variables is [+/- here-and-now]. For example, Robinson (1995) examined the impact of task complexity on English adult learners' oral narrative performance. Three wordless narrative strips were used to elicit learners' L2. Two levels of task complexity (i.e., simple and complex) were operationalized along [+/- here-and-now]. Each participant performed both the simple and the complex versions of the task. The results showed that learners demonstrated more complex and accurate, but less fluent L2 output in the complex task than the simple task.

Likewise, Gilabert (2007a) inquired into the effect of cognitive task complexity on English adult learners' L2 production and manipulated task complexity along [+/- here-and-now]. Additionally, the researcher also operationalized [+/- planning time] as another task complexity variable. Forty-eight Spanish L1 learners of English whose English proficiency level were low-intermediate performed both the simple and the complex tasks. The results partially aligned with Robinson's (1995) finding that increasing task complexity along [+/- here-and-now] induced more accurate L2 production. However, both syntactic complexity and lexical complexity were negatively affected by the increased cognitive demands. Fluency was enhanced in the complex task. With regard to the [+/- planning time] variable, the complex task (i.e., less planning time) induced lower lexical complexity and fluency. No significant effects were found for syntactic complexity and accuracy.

Other resource-directing variables such as [+/- reasoning] and [+/- few elements] have also been addressed in the literature. For instance, Gilabert (2007b) examined the effects of task complexity on adult learners' L2 oral production with a specific emphasis on accuracy, which was measured via self-repair. Task complexity variables involved in the study were [+/- here-and-now], [+/- few elements], and [+/- reasoning]. The treatment tasks were a narrative task, an instruction-giving task, and a decision-making task. Forty-two Spanish L1 learners of English with a low-intermediate English proficiency level performed three sets (i.e., simple and complex) of the tasks. The results showed that the impact of task complexity varied across different tasks. First, in the narrative task, although the participants made fewer errors and had a higher ratio of repaired to unrepaired errors performing the complex task, no significant differences were found for the number of errors per AS-unit (i.e., analysis of speech unit) and the percentage of self-repairs. Second, a clearer pattern was revealed for the map task as a larger number of repairs and a higher repair rate in the complex task was observed. Finally, for the decision-making task, the cognitive demands of the task did not significantly affect the number of errors and self-repairs. The findings indicate that task type might be a moderating variable in the relationship between task complexity and accuracy.

With regard to resource-dispersing dimension, pre-task planning has received a great amount of attention in task complexity research. For example, Yuan and Ellis (2003) conducted a study on the effects of pre-task planning on learners' L2 oral production with picture narrative tasks. The participants were 42 undergraduate learners of English. Three types of planning were operated: no planning, pre-task planning, and online planning. The results showed that the no planning group obtained greater fluency and lexical complexity than the other two groups. However, the two planning groups achieved higher syntactic complexity. Between the two planning groups, the pre-task planning group demonstrated greater fluency and lexical complexity, while the online planning group produced utterances that were more accurate. Overall, the findings indicate that providing planning time facilitated syntactic complexity. The type of planning, on the other hand, influenced fluency and accuracy.

Similarly, Sasayama and Izumi (2012) used picture narrative tasks to examine the effects of pre-task planning and the number of elements involved in the task on learners' L2 oral production. The participants were 23 Japanese high school students who were learning English as a foreign language. All participants performed a simple and a complex picture-based tasks varying along the number of elements involved in the task. Ten of the participants were given pre-task planning time and the others were not. Learners' L2 production was measured for accuracy, complexity, and fluency. The results indicated that when task complexity was raised by adding more elements to the task, accuracy, and fluency were negatively affected but syntactic complexity was positively affected. On the other hand, when removing pre-task planning, fluency was enhanced, while syntactic complexity was negatively affected. No significant effect was found for accuracy.

In sum, the findings from previous studies only provide partial support for the Cognition Hypothesis. The line of research manipulating task complexity along resource-directing variables indicates an overall positive impact of task complexity on accuracy. Nevertheless, its impact on syntactic and lexical complexity, and fluency is still inconclusive. Moreover, increasing task complexity along resource-dispersing dimensions tends to negatively affect syntactic complexity while its impact on accuracy, fluency, and lexical complexity remains unclear. The inconsistency in the findings indicates a need for more research examining the relationship between task complexity and L2 performance. Therefore, the present study aims to investigate the effect of task complexity along [+/- few steps] on adult learners' oral L2 performance. Additionally, the study also attempts to explore whether task complexity influences L2 performance differently in FTF and SCMC modes.

4.2.2 Task Complexity in Synchronous Computer-mediated Communication

All the studies discussed above were conducted in an FTF research setting. As communication in multiuser virtual environments has become prevalent in recent decades, SLA researchers have shown great interest in task-based language learning in online environments. Among various types of computer-mediated communication (CMC) (e.g., text-based, audio-based, video-based), text-based synchronous CMC (SCMC) has received the greatest attention from language researchers and educators. Synchronous computer-mediated communication (SCMC) is considered "a hybrid between written and oral conversation" (Baralt, 2013, p. 696). Compared to FTF interaction, SCMC supports a written record of the information and allows for more processing time for comprehension and attention to form (Chapelle, 2001; Smith, 2004). Additionally, there has been empirical evidence showing that the language gains through online chat could be transferred to oral proficiency (Payne & Whitney, 2002).

A growing body of research has revealed the advantages of SCMC in promoting learner participation (Kelm, 1992; Kern, 1995; Warschauer, 1996), reducing task anxiety (Sauro, 2009), and facilitating the development of interactive competencies (Chun, 1994; Darhower, 2002; Eslami & Kung, 2016). With regard to language learning, SCMC may induce more noticing (Chapelle, 2001; Lai & Zhao, 2006; Long, 2007; Payne & Whitney, 2002; Yuksel & Inan, 2014), more accurate L2 production (Payne & Whitney, 2002; Pelletieri, 2000), and greater L2 development (Payne & Whitney, 2002; Sykes, 2005). However, there is a relative lack of research that investigates the relationship between task complexity and L2 learning in SCMC. One of the few task complexity studies in SCMC was conducted by Baralt (2013). He investigated the effects of task complexity and task modality on interaction (i.e., the efficacy of recasts) and the development of the Spanish past subjunctive 84 adult learners of Spanish. Two levels of task complexity were manipulated along the [+/- intentional reasoning]. The participants were divided into five groups: control, FTF + Complex, FTF – Complex, SCMC + Complex, SCMC – Complex. The learning outcomes were measured with a multiple-choice receptive test and two productive tasks (one in FTF and one in SCMC). The results revealed that the simple task in SCMC led to most gains in Spanish past subjunctive, followed by the complex task in FTF, and the complex task in SCMC led to the least amount of progress.

These findings suggest an interactive effect of task complexity and task modality on L2 learning. In the study, learners' performance in FTF interaction aligned with the prediction of the Cognition Hypothesis that more complex tasks promote attention to form and thus result in greater gains in L2. In contrast, increased cognitive demands did not promote L2 learning in SCMC. To better interpret the results, in the same study, Baralt conducted a follow-up analysis of learners' online discourse. He found that when the task was more cognitively demanding in

SCMC, learners spent a lot of time and effort in providing and explaining their arguments. Oftentimes, it took many turns to complete one explanation. Since both interlocutors were typing concurrently in SCMC, erroneous utterances and feedback were largely overlooked. This could explain why the SCMC + Complex group made the least gains.

Another study was conducted by Adams, Alwi, and Newton (2015), who also examined the role of task complexity in SCMC but on L2 performance. Task complexity was manipulated along [+/- task structure] as well as [+/- language support] dimensions, which was defined as "…language-focused pre-task planning by pushing learners to explicitly consider language forms that would be useful as they carried out the task" (Adams et al., p. 69). Both variables were identified as in the resource-dispersing dimension. Although the variable "language support" was not specified in the Triadic Componential Framework, the researchers claimed that its effect resembled that of the pre-task planning dimension in the framework. The participants were divided into four groups: low task structure without language support (-TS, -LS), low task structure with language support (-TS, +LS), high task structure without language support (+TS, -LS), high task structure with language support (+TS, +LS). Participants' online chat exchanges were collected and analyzed in terms of accuracy, syntactic and lexical complexity. The results showed that learners' L2 production was less accurate when they were performing the more complex tasks. Nevertheless, no statistically significant impact of task complexity was detected on syntactic and lexical complexity. The findings partially supported the Cognition Hypothesis that increasing task complexity along resource-dispersing variables leads to lower accuracy and complexity in L2 production.

To the best of my knowledge, the above two studies are the only existing empirical studies that examine how task complexity affects L2 performance in a text-based SCMC mode.

Other studies may also examine task complexity in CMC but not in text-based SCMC environments. For example, York (2019) investigated the effects of task complexity on L2 performance but in an audio-based SCMC setting. Given the limited empirical evidence, it is still unclear whether text-based SCMC plays a role in moderating the effect of task complexity on L2 performance. In order to extend our understanding toward the relationship between task complexity, mode, and L2 performance, the present study aims to investigate whether task complexity differentially influences learners' L2 performance in FTF interaction and text-based SCMC.

4.2.3 Pedagogical Task Design in Content-based Language Teaching

Although a substantial number of studies have examined how task complexity influences L2 learning in either FTF or SCMC setting, research concerning the effects of cognitive task complexity in content areas has remained scarce. Crandall and Tucker (1990) define CBLT as "an approach to language instruction that integrates the presentation of topics or tasks from subject matter classes (e.g., math, social studies) within the context of teaching a second or foreign language" (p. 187).

The theoretical frameworks that underlie CBLT are Krashen's (1982) Input Hypothesis and Cummins' (1979, 1981, 2000) framework of language proficiency. In the Input Hypothesis, Krashen (1982) claims that meaningful and comprehensible input is the key to language acquisition. Cummins (1979, 1981, 2000) posits a dichotomy between basic interpersonal communication skills (BICS), which are language skills used in daily communication, and cognitive academic language proficiency (CALP), which includes language skills used to comprehend and construct academic knowledge. Cummins claims that while BICS may take one to two years to develop, CALP requires content-based instruction and takes five to seven years to achieve. Built upon these two frameworks, CBLT develops students' CALP by providing a meaningful and contextually rich academic language learning environment (i.e., subject matter content).

Leaver and Stryker (1989) summarize four core characteristics of CBLT: (a) subject matter (i.e., the curriculum should be developed based on the subject matter), (b) use of authentic texts (i.e., both the teaching and learning materials and learning activities should reflect authentic use of the target language in the real world), (c) learning of new information (i.e., there should be opportunities for learners to use the target language to learn and construct new subject-specific information), and (d) appropriate to the specific needs of students (i.e., the instruction should be cognitively and affectively appropriate to the students). In this sense, similar to TBLT, CBLT emphasizes authentic language use, views the target language as a tool for communication rather than a learning subject, and considers the influence of cognitive demands of learning activities on learners.

Based on the degree of focus on language and content, Lyster (2017) categorizes CBLT programs on a continuum with language-driven programs at one end and content-driven programs at the other end (see Figure 5). At the language-driven end are foreign language classrooms that aim to develop learners' L2 proficiency with themes or topics (e.g., weather, travel, and environmental issues). In this type of CBLT program, themes and topics are used to facilitate language learning rather than for content knowledge development. The middle of the continuum are programs that focus on both language and content, such as content and language integrated learning (CLIL) programs and English-medium instruction (EMI). Learners involved in these programs are supposed to take content classes taught in a foreign language. At the content-driven end are immersion programs, in which more than 50% of the curriculum is delivered in the target language.

Figure 5

Range of CBLT Settings (Lyster, 2017, p. 3)

Language-driven	⟵⟶	Content-driven
e.g., theme-based language course	e.g., content course(s) + language courses	e.g., immersion programs (50%+ in target language)

Despite the ideal environment CBLT provides for L2 development and meaningful contexts for both comprehension and production, concerns rise along with the actual implementation of CBLT programs. A major issue in CBLT that has been brought up by many researchers is the lack of attention to formal aspects of language (Dalton-Puffer, 2011; Llinares & Dalton-Puffer, 2015; Lyster, 2017; Sato & Loewen, 2019). As Lyster (2017) points out, “It has become clear in CBLT that many target-language features, especially (but not limited to) those involving morphosyntax, are not learned by ‘osmosis’ and instead require a more intentional instructional focus” (p. 6). Research on CBLT (e.g., Blasco, 2014) has shown that although learners develop high levels of communicative abilities, their accuracy tends to be less developed. To reinforce learners’ attention to linguistic features in content areas, language researchers and educators have proposed the use of pedagogical tasks, which are developed by following the principles of TBLT, in discipline related learning (Lopes, 2019; Lyster, 2015; Ortega, 2015).

With two studies, Lyster (2015) provides an exemplary illustration of using form-focused tasks to draw learners' attention to language in CLIL contexts in Canada. The first study integrated a focus on French grammatical gender across 5th-grade (10-11 years old) French immersion students. The experimental group completed noticing activities in the language arts class, awareness activities in the social studies class, and practice activities in the science class. In contrast, the comparison group took the regular CLIL program without focused tasks. The second study involved literacy tasks with a focus on French derivational morphology. The participants were 2nd-grade (7-8 years old) French immersion students who completed the focused tasks in one language (e.g., French) and then in the other (e.g., English). The comparison group, on the other hand, attended the regular CLIL classes (i.e., French literacy). The results of both studies showed that the experimental groups significantly outperformed the comparison groups on the language assessments. Based on the findings, Lyster claims that "Task would be pivotal in such a cross-curricular approach and would be designed to provide purposeful opportunities for strengthening connections between language and content learning" (p. 12).

In a recent study, Li and Chen (2019) compared the effectiveness of CBLT and TBLT in teaching reading to learners of military English. A pre-posttest design was adopted in the study. The participants were 120 undergraduate students who were taking military English classes in addition to traditional English classes. All participants received reading comprehension instruction but in different ways. The teacher directed students in the CBLT group to read and analyze the texts on military affairs via group discussion and student presentation, without a specific focus on the language. On the other hand, the TBLT group conducted specially designed reading tasks (with the same texts), with the purpose of directing students' attention to the language (e.g., drawing an information table, carrying out a class survey). Learners' gains were evaluated via a reading test composed of four reading passages related to military affairs. A follow-up interview was carried out for students' perception of their learning experience. The results echoed Lyster's (2017) argument and showed that without a conscious emphasis on language, students in the CBLT group tended to neglect the language and were mostly concerned with the content (i.e., military knowledge). Students in the CBLT group reported that their previous military knowledge helped them comprehend the content, even if they did not understand the language. In contrast, students in the TBLT group were oriented to the language and used the language to construct new content knowledge. Results from the reading tests showed that students in the TBLT group achieved significantly more gains than their peers in the CBLT group.

The findings of the above two studies provide evidence for the effectiveness of incorporating pedagogical tasks in content-based environments. An important issue needing more research is in what ways pedagogical tasks should be designed to facilitate language learning in content areas. In other words, how to manipulate task features to maximize language learning in subject matter contexts has become the key to the successful integration of TBLT and CBLT. Although there have been a considerable number of studies that examine the relationship between task design features and L2 learning, research on task design features in content areas is sparse. To address this gap, the present study aims to examine the effects of cognitive task complexity in content-based learning (i.e., math). In particular, the study examines the effects of task complexity on learners' L2 performance (i.e., accuracy, syntactic complexity, lexical complexity, and fluency) while they are solving mathematical word problems. To ensure that the experimental tasks are indeed different in cognitive demands, the study utilizes independent measures of task complexity to validate task design manipulation.

4.2.4 Independent Measures of Task Complexity

One methodological issue that has been largely overlooked by many researchers in task complexity studies is the validity of task complexity manipulation. In other words, although numerous task complexity studies have sought to test the predictions of the Cognition Hypothesis, very few of them have provided evidence showing that the designed complex tasks were indeed more cognitively demanding to learners. Norris and Ortega (2003) have addressed this issue and suggested the use of independent measures of cognitive demands to validate the presumed different levels of task complexity in research.

Accordingly, researchers have proposed and employed a few independent measures of task complexity: (a) self-perceived task difficulty (Baralt, 2010; Gilabert, 2007b; Ishikawa, 2011; Lee, 2019; Robinson, 2001b), (b) time judgment (Baralt, 2013; Sasayama, 2015), (c) stimulated recall (Kim et al., 2015), (d) expert judgments (Révész et al., 2016; Révész et al., 2014), (e) dual-task methodology (Révész et al., 2016; Révész et al., 2014; Sasayama, 2015), and (f) eye-tracking (Révész et al., 2014). Following Révész (2013) who suggests a combination of different measures of task complexity to best estimate the cognitive loads of tasks, the present study employs self-perceived task difficulty and time judgment to validate task complexity manipulation.

Self-perceived task difficulty is one of the most commonly used independent measures of task complexity. For instance, in Ishikawa's (2011) study, 24 learners of English conducted three monologic reasoning tasks with different cognitive demands and completed the task difficulty questionnaire by following each of the three tasks. The results showed that the designed-to-be more cognitively demanding task was actually perceived as more difficult to perform by the learners. In a more recent study, Lee (2019) also used a task difficulty questionnaire to examine

42 English native speakers' perceptions of task difficulty, required mental effort, and stress. The study involved three types of oral monologic tasks: map task, seating arrangement task, and car accident task. Each task type had three cognitive complexity levels from simplest to most complex. The results were aligned with Ishikawa's findings that the more complex tasks were perceived as more difficult by the participants, irrespective of task type. Taken together, the findings from the two studies suggest that self-perceived task difficulty is a valid assessment of task complexity.

Subjective time judgment is another common measure of cognitive load in psychology, and it has been recently used as a measure of task complexity in task-based research (e.g., Baralt, 2013). The judgment can be either prospective or retrospective depending on when the participants are required to do the judgment task. In the prospective paradigm, before task performance, participants are told that they will need to estimate the duration of their performance after the completion of the task. In the retrospective paradigm, on the other hand, participants are not aware of the judgment until they finish their performance.

To date, very few researchers have employed duration judgment in their studies in the field of TBLT. Baralt's (2013) study was the first attempt to use retrospective time judgment as a measure of task complexity. In the study, 84 learners of Spanish were randomly assigned to a control group or one of the four experimental groups: simple task in FTF mode, complex task in FTF mode, simple task in SCMC mode, and complex task in SCMC mode. After completing the task, learners were requested to estimate the duration of their task performance. The time differences between the estimation and the actual duration were calculated for analysis. The results showed that the estimated durations of complex tasks were significantly longer than the actual performance time, and the estimation of simple tasks was significantly shorter than the actual time. Based on the findings, Baralt stated that retrospective time estimation could be one way to validate task complexity manipulation.

Building on the existing literature, the present study employed both self-perceived task difficulty and retrospective time judgment to estimate the cognitive loads of the experimental tasks. Nevertheless, the subjective-to-objective ratios (i.e., the ratio of estimation to actual time), instead of time differences in Baralt's (2013) study, were calculated for data analysis. This data analysis approach is supported by literature in the field of psychology that an increase in cognitive load leads to a higher subjective-to-objective ratio in the retrospective condition (Block et al., 2010).

4.2.5 Research Questions

The current literature shows inconsistent findings with regard to the influence of task complexity on learners' L2 performance. It is also unknown whether the mode (i.e., FTF or SCMC) moderates the effects of task complexity on learners' task performance. Furthermore, task design in CBLT has yet to be examined from a cognitive perspective. To address these research gaps, the present study seeks to examine the impact of task complexity on learners' L2 performance in SCMC and FTF conditions in content-based language learning. The study utilizes a simple and a complex mathematical word problem to examine the impact of different cognitive loads on learners' language performance. Findings from this research will not only provide insights into pedagogical task design in the CBLT context but also offer additional empirical evidence regarding the utility of different task complexity evaluation scales (i.e., self-perceived difficulty and time judgment).

Prior to the study, the validity of task complexity was measured based on a pilot study with a group (N=20) who had a set of characteristics (i.e., English non-native speakers (NNSs), and English native speakers (NSs)at a public university in the US) similar to the participants in the current study. The results from the pilot study showed that the complex task (M=4.7, SD=1.06) was perceived as more difficult than the simple task (M=1.6, SD=0.70), supporting our manipulation of task complexity. Furthermore, we also investigated whether cognitive complexity is perceived differently in FTF and SCMC modes with the participants in the present study. The following research question directed the study:

1. How does task complexity affect learners' oral language production in terms of syntactic complexity, accuracy, lexical complexity, and fluency (CALF) as they described the steps followed in solving the problem after their engagement in solving mathematical word problems with their interlocutors in the FTF mode?

2. How does task complexity affect learners' oral language production in terms of syntactic complexity, accuracy, lexical complexity, and fluency (CALF) as they described the steps followed in solving the problem after their engagement in solving mathematical word problems with their interlocutors in text-based SCMC?

3. Does the impact of task complexity on learners' language production as they described the steps followed in solving the problem after their engagement in solving mathematical word problems with their interlocutors in terms of CALF differ depending on the interactive mode (i.e., FTF vs. SCMC)?

4. Do different intended levels of task complexity lead to different levels of perceived task performance, as measured by self-perceived difficulty and time judgment?

4.3 Method

4.3.1 Participants

The participants in the study were 82 undergraduate and graduate students (55 males, 27 females) at a public university in Texas, United States. Their ages ranged from 18 to 33 (M= 25.57, SD= 3.55). Half of the participants (n=41) were English native speakers (NSs), and the other half (n=41) were English non-native speakers (NNSs), who were international students enrolled in the university. All the NNSs, except four, were native speakers of Mandarin (n=37). The remaining four NNSs were native speakers of Korean (n=1), Japanese (n=1), Turkish (n=1), and Kazakh (n=1). The participants were majoring in various fields. Table 12 and Table 13 demonstrate the NSs' and NNSs' are of study at the time of the study. The NNSs' English proficiency was measured based on their most recent iBT TOEFL scores. The average iBT TOEFL score was 96.67 (SD= 4.55, range 86 - 102), indicating a high-intermediate English proficiency level.

Table 12

English Native Speakers' Field of Study

Major	# of NSs
Education	21
Psychology	9
International Studies	4
Communication	3
Finance	2
Sociology	1
Public Health	1

Table 13

English Non-Native Speakers' Field of Study

Major	# of NNSs
Mechanical Engineering	11
Computer Engineering	8
Economics	6
Biochemistry	4
Chemistry	3
Civil Engineering	3
Electronic Engineering	2
Petroleum Engineering	2
Physics	1
Public Health	1

4.3.2 Design

A repeated measure design was employed in the study. Two versions (i.e., simple and complex) of a mathematic word problem were designed and utilized as the experimental tasks based on Robinson's (2007a) Triadic Componential Framework. Cognitive task complexity was manipulated along the resource-dispersing variable [+/- few steps]. Each dyad (one NS and one NNS) completed both task versions under a FTF or SCMC condition. A counterbalanced design was utilized to control for order effects. Under each condition, half of the dyads performed the simple task first and the other half performed the complex task first (Table 14).

Table 14

Study Participant Groups

FTF		SCMC	
Group 1	Group 2	Group 3	Group 4
Simple task	Complex task	Simple task	Complex task
Complex task	Simple task	Complex task	Simple task

Task completion consisted of two steps. The first step involved an interactive problem-solving process, during which the NNS and the NS worked together to find the solution to the problem. After that, the NNS was asked to describe the problem-solving process independently to the researcher. The NSS's oral narrations were recorded and transcribed for analysis.

4.3.3 Materials

4.3.3.1 Background Questionnaire

This instrument consisted of three parts. The first part asked for the demographic information of the participants (e.g., age, gender). The second part contained open-ended questions related to NNSs' English education experiences. The third part concerned the participants' experience of using technology, such as the frequency of online chat (see Appendix A). The NNSs were asked to complete all three parts. The NSs only needed to complete the first and the third parts since the second part was not relevant to them.

4.3.3.2 Experimental Tasks

The experimental tasks were two mathematical word problems of different cognitive complexity levels, which were designed by the researcher with reference to the literature related to mathematical word problems (e.g., Whimbey et al., 2013). Both word problems involved a story of two people who intended to collaborate on a certain task (i.e., planting flowers or painting a room). A series of incidents occurred during the task of planting flowers or painting a room. Both word problems asked the participants to figure out the time needed to complete the task by considering all the incidents in the story. The word problems were reviewed by two doctoral students in math education to make sure that the mathematical knowledge required in the problems was appropriate to the participants' overall mathematical competence. This was also confirmed by task performance in the pilot study.

The two problems required the same mathematical knowledge but were different in cognitive demands. The main differences lied in whether the work efficiency was already provided in the problem and the number of incidents that occurred while completing the task in the story. The simple task provided participants with the work efficiency and involved a fewer number of incidents in the story. In contrast, the complex task required the participants to calculate the work efficiency first based on the provided information, and also had a more complicated storyline (see Appendix B).

4.3.3.3 Task Difficulty and Time Judgment Questionnaire

To gauge the cognitive demands of the experimental tasks, a questionnaire asking about task difficulty and the estimated duration of task performance (i.e., the narration) was administered to all participants (see Appendix D). The questionnaire contained two parts. The first part asked the participants to estimate the time they spent on the task. The second part required the participants to rate how challenging the narration was on a 6-point Likert scale, with 1 as the easiest and 6 as the most difficult. The participants filled out the questionnaire immediately after completing each task. To make sure participants understood the steps to follow, a detailed verbal instruction was provided to all the participants prior to completing the questionnaire and participants' questions were addressed by the researcher.

4.3.4 Procedure

Participants were recruited by emails as well as classroom visits. The researcher sent a recruiting email through the university email system to students in the email list. In the email, a brief introduction of the study, criteria for potential participants, and contact information of the researcher were provided. Potential participants who were interested in the study contacted the researcher for consideration. The researcher also explained the study to professors who were teaching the related courses. After getting permission from the professors, the researcher went to different classes to introduce the study and provided contact information for potential participants.

The participants who showed interest to be included in the study consisted of 47 English NSs and 43 NNSs. The researcher then sent out the background questionnaire to the interested participants via Qualtrics. Four participants were not included in the study since they did not fill out the background questionnaire.

In the next step, the researcher randomly paired each NNS with an NS to form an NS-NNS dyad. The researcher scheduled a study session with each dyad based on their availability. The study sessions were carried out in the group study rooms at different libraries in the university. Participants who were not able to complete all the sessions were removed from the study (n=4). Finally, 41 NS-NNS dyads involving 82 participants were included in the study. Each participant was paid with a $10 or a $15 Starbucks gift card, depending on the length of the session, as compensation for their participation. The steps followed for collecting the experimental data (Figure 6), which lasted about one hour to one hour and half, are explained in the following sections.

4.3.4.1 The FTF Groups

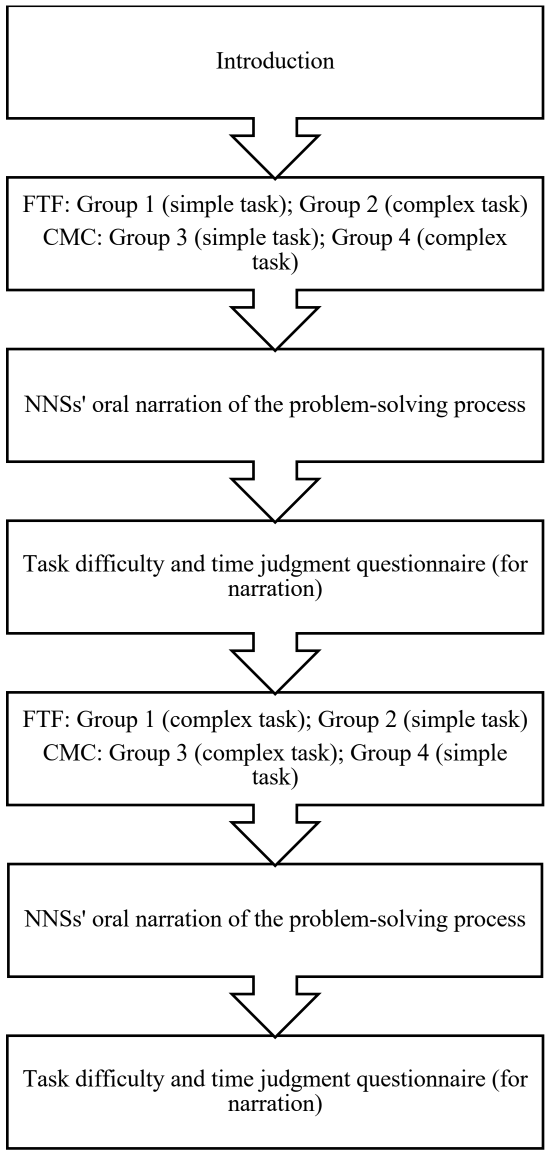
At the beginning of the session, the researcher briefly introduced the study and gave the consent form to participants to sign. The two participants in a dyad then introduced themselves to each other. After the introduction period, the participants were given one version of the task to complete. The NS and NNS discussed the word problem and worked out the solution together. Their interactions were audio-recorded. As soon as the participants informed the researcher that the task of solving the math problem was completed, the researcher gave the task difficulty and time judgment questionnaire to each of the participants to fill out independently. After the questionnaire was completed, the NS was asked to take a short break outside the room. The NNS stayed in the room and was asked to orally describe the problem-solving process. The oral narrations were audio recorded. Once the narration was complete, the researcher invited the NS back to the room to complete the other version of the task with the NNS. Once again, their discussions were recorded. The same process was applied for the second task.

4.3.4.2 The SCMC Groups

The session for the SCMC groups also started with the introduction. However, after that, the NS and the NNS were assigned to different rooms. Each participant was provided a copy of the word problems. The participants were asked to use Google Hangouts to interact during the math problem-solving task. Their messages in Google Hangouts were recorded. Similar to the FTF groups, the participants completed the task difficulty and time judgment questionnaire immediately after each task. The NNS was also asked to orally explain the problem-solving process for each task. The oral narrations of the steps followed to solve the math problem were recorded.

Figure 6

The Procedure of the Study



4.3.5 Data Analysis

All NNSs' oral narrations of the problem-solving process were transcribed. Linguistic performance was analyzed in terms of accuracy, syntactic complexity, lexical diversity, and fluency. The first step of data analysis was segmenting the L2 utterances into AS-units (i.e., analysis of speech units). According to Foster, Tonkyn, and Wigglesworth (2000), "an AS-unit is a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (p. 365). Unlike the T-unit, which is defined as "a main clause plus all subordinate clauses and non-clausal structures attached to or embedded in it" (Hunt, 1970, p. 4), the AS-unit considers the inclusion of independent sub-clausal units that are common in oral production (Foster et al., 2000). In the example that follows below, "It is interesting" is an independent clause because it contains a subject (i.e., it) and a verb (i.e., is), and it can stand by itself as a sentence. On the other hand, "that you always pick" and "what she chose" are subordinate or dependent clauses because they are dependent on the main clause.

Example 4

It is interesting that you always pick what she chose. (3 clauses, 1 AS unit)

4.3.5.1 Accuracy

As suggested by Robinson and Gilabert (2007), both general and specific measures of accuracy should be included for L2 performance analysis. Therefore, we used three different measures of accuracy: (a) the percentage of error-free clauses, (b) the percentage of error-free AS-units, and (c) the percentage of accurate use of grammatical verbs. Grammatical verb forms include tense (e.g., present, past), modality (e.g., should, have to), and subject-verb agreement. The specific calculation of each measure is illustrated below:

The percentage of error-free clauses	=	The total number of error-free clauses / The total number of clauses
The percentage of error-free AS-unit	=	The total number of error-free AS-units / The total number of AS-units

The percentage of accurate use of grammatical verb forms	=	The total number of correct grammatical verb forms / The total number of grammatical verb forms

4.3.5.2 Syntactic Complexity

The evaluation of syntactic complexity involved three different measures: (a) the number of clauses per AS-unit, (b) the number of dependent clauses per AS-unit, and (c) the number of words per AS-unit. The calculation of each measure is as follows:

The number of clauses per AS-unit	=	The total number of clauses / The total number of AS-units
The number of dependent clauses per AS-unit	=	The total number of dependent clauses / The total number of AS-units
The number of words per AS-unit	=	The total number of words / The total number of AS-units

4.3.5.3 Lexical Diversity

For the analysis of lexical diversity, three measures were utilized: (a) type-token ratio (TTR), (b) D score (e.g., Révész, 2011), and (c) Measure of Textual Lexical Diversity (MTLD) (e.g., Lee, 2019). Type-token ratio (TTR) is a common measure of lexical variation in SLA research (e.g., Xing, 2015). It is calculated by using the total number of distinct words (i.e., type) divided by the total number of words (i.e., tokens) in a text. However, the disadvantage of using TTR is that the length of the text can easily affect the TTR value – the increase in the length of a text leads to a systematic decrease in the value of TTR (Malvern & Richards, 2002). This means that TTR does not allow for valid comparisons between learners' L2 productions that vary considerably in length. In order to account for this issue, researchers (e.g., Malvern & Richard, 1997) have introduced D-score as an additional measure of lexical diversity. The calculation of D value involves "a random selection of tokens in plotting the curve of TTR against increasing token size for the text to be analyzed" (Kormos, 2011, p. 154). A higher D value entails a greater diversity of a text. Although D value is not a function of text length as TTR, it requires a minimum of 50 words for calculation. In this sense, texts that have less than 50 words cannot be analyzed using this method. In consideration of this issue, MTLD is introduced as a sequential analysis for lexical diversity (McCarthy, 2005). The MTLD value presents "the average number of words in a row for which a certain TTR is maintained" (Fergadiotis et al., 2013). It has been used in many research studies and is found to be a strong indicator of lexical diversity (e.g., Lee, 2019). Following McCarthy and Jarvis's (2010) suggestion that multiple indices should be employed to assess lexical diversity, all the above three measures (i.e., TTR, MTLD, and D scores) were utilized in the present study. TTR, MTLD, and D scores were calculated automatically via a web-based text analysis software called Text Inspector.

4.3.5.4 Fluency

Fluency reflects the efficiency in the L2 production process. Measures of fluency included (a) speech fluency and (b) repair fluency. Speech fluency was measured by calculating the number of syllables per minute (Cho, 2018). General repair fluency was measured by calculating the total number of repeated words or phrases divided by the total number of words (Skehan, 2003). Two types of repetition were then identified: verbatim and substitutive repetitions (Sasayama & Izumi, 2012). Verbatim repetitions occur when learners are looking for an appropriate word, and substitutive repetitions take place when learners attempt to correct their own errors (Bygate, 1996).

The researcher independently coded all the utterances. Twenty percent of the data was double-coded by a research assistant who was trained in coding the data. The inter-coder reliability was 82.6% for the number of AS-units, 94.2% for the number of clauses, 91.4% for the number of subordinate or dependent clauses, 83.7% for error-free clauses, and 97.7% for the correct use of grammatical verb forms.

4.3.6 Statistical Analyses

Descriptive statistics were calculated for each of the following measures: (a) average self-perceived task difficulty ratings for each experimental task, (b) average actual and estimated time durations for each experimental task, (c) linguistic performance measures (i.e., accuracy, syntactic complexity, lexical variation, and fluency) for each experimental task. Outliers were defined as a score that is three standard deviations away from the mean. All the outliers were identified and eliminated from the sample for inferential statistical analysis. There was one outlier for task difficulty ratings, one for duration judgment, and one for linguistic performance measures. The data were normally distributed and met the conditions for use of parametric tests. A series of paired *t*-tests and mix-design ANOVAs were conducted. All analyses were completed in STATA with an alpha level of 0.5. Partial η^2 was calculated to measure effect size. For partial η^2 values, .01, .06, and .14 were considered small, medium, and large, respectively (Baralt, 2013).

4.4 Results

4.4.1 Linguistic Performance of the FTF Group

The first research question concerned the impact of cognitive task complexity on NNSs' linguistic performance in the FTF mode. A series of paired *t*-tests were computed to examine whether there were statistically significant effects of task complexity on accuracy, syntactic complexity, lexical diversity, and fluency.

4.4.1.1 Accuracy

Three measures were utilized to assess the accuracy of NNSs' oral production: (a) the percentage of error-free clauses, (b) the percentage of error-free AS-units, and (c) the percentage of accurate use of grammatical verbs. Table 15 presents the descriptive and inferential statistics for each measure of accuracy across task conditions. The results showed that the simple task yielded higher percentages of error-free clauses, error-free AS-units, and accurate use of grammatical verbs. However, the differences were not statistically significant.

Table 15

Mean, Standard Deviation, and Paired t-tests Results for Accuracy Measures of the FTF Group

	FTF_S		FTF_C		*t*	*p*
	M	SD	M	SD		
Percentage of error-free clauses	81.39	12.20	81.25	6.50	0.07	0.95
Percentage of error-free AS-unit	70.11	17.46	66.78	11.79	0.91	0.37
Percentage of accurate use of grammatical verbs	90.32	10.20	89.75	4.50	0.36	0.73

4.4.1.2 Syntactic Complexity

The measures that were used to evaluate syntactic complexity were: (a) the number of clauses per AS-unit, (b) the number of dependent clauses per AS-unit, and (c) the number of words per AS-unit. Table 16 summarizes the descriptive and inferential statistics for each measure across task conditions. The results showed that NNSs produced longer AS-units, and more independent and dependent clauses per AS-units on the more cognitively demanding task. Once again, none of the differences were statistically significant.

Table 16

Mean, Standard Deviation, and Paired t-tests Results for Syntactic Complexity Measures of the FTF Group

	FTF_S		FTF_C		*t*	*p*
	M	SD	M	SD		
Number of clauses per AS-unit	1.80	0.49	1.96	0.53	-1.92	0.07
Number of dependent clauses per AS-unit	0.59	0.33	0.67	0.41	1.25	0.23
Number of words per AS-unit	14.57	3.94	15.33	3.05	-1.32	0.20

4.4.1.3 Lexical Variation

For the analysis of lexical diversity, TTR, D-value, and MTLD were calculated. Table 17 illustrates the descriptive and inferential statistics for each measure across task conditions. Results of paired *t*-test conducted on lexical variation in terms of TTR indicated significant effects for cognitive task complexity, t (20) = 7.95, $p < 0.001$, indicating the greater lexical richness of NNSs' narrative production on the simple task. On the other hand, the analysis of D-value and MTLD showed contrasting results in that lexical variation was slightly greater on the complex task than the simple task. The differences were not statistically significant. Since the length of the L2 narratives ranged from 98 to 994 words and TTR is easily affected by text length, the results from D-value and MTLD could be considered to be more reliable. Therefore, the results suggested that task complexity did not significantly impact lexical variation.

Table 17

Mean, Standard Deviation, and Paired t-tests Results for Lexical Variation Measures of the FTF Group

	FTF_S		FTF_C		*t*	*p*
	M	SD	M	SD		
TTR	0.42	0.08	0.31	0.08	7.95	0.00
D-value	40.41	9.56	40.67	7.72	-0.13	0.90
MTLD	33.04	8.88	33.32	8.79	-0.15	0.88

4.4.1.4 Fluency

Four measures were employed to assess fluency: (a) speech fluency (number of syllables per minute), (b) general repair fluency (percentage of repeated words/ phrases), (c) percentage of verbatim repetitions, and (d) percentage of substitutive repetitions. Table 18 summarizes the descriptive and inferential statistics for each measure across task conditions. The results showed that the more cognitively demanding task led to greater speech fluency but lower repair fluency (i.e., general, verbatim, and substitutive). In other words, NNSs produced more syllables per minute and fewer repetitions (both verbatim and substitutive) on the complex task. Among the four different measures related to fluency, the only significant effect of task complexity was on general repair fluency, t (20) = 2.51, p = 0.02.

Table 18

Mean, Standard Deviation, and Paired t-tests Results for Fluency Measures of the FTF Group

	FTF_S		FTF_C		t	p
	M	SD	M	SD		
Number of syllables per minute	140.56	20.54	141.35	22.30	-0.31	0.76
Percentage of repeated words/phrases	8.58	6.87	6.35	4.72	2.51	0.02
Percentage of verbatim repetitions	4.97	5.56	4.15	4.03	0.94	0.36
Percentage of substitutive repetitions	3.61	4.01	2.19	2.06	1.44	0.17

To summarize, for the FTF group, the manipulation of task complexity did not significantly affect the accuracy, syntactic complexity, and lexical variation of NNSs' L2 oral production. Nevertheless, the more cognitive demanding task led to significantly fewer repetitions than the simple task.

4.4.2 **Linguistic Performance of the SCMC Group**

The second research question investigated how task complexity affected NNSs' oral language production in the SCMC context. Paired t-tests were conducted to identify any statistically significant effects of cognitive demands on accuracy, syntactic complexity, lexical variation, and fluency. The measures of each linguistic dimension were the same as the ones used for the FTF group. The results for each linguistic dimension are reported in the following sections.

4.4.2.1 Accuracy

Table 19 presents the descriptive and inferential statistics for each measure of accuracy across task versions. As shown in Table 19, increased cognitive demands led to higher percentages of error-free clauses, error-free AS-unit, and accurate use of grammatical verbs. However, the differences were statistically significant only for the percentage of error-free clauses, t (18) = -2.32, p = 0.03.

Table 19

Mean, Standard Deviation, and Paired t-tests Results for Accuracy Measures of the SCMC Groups

	SCMC_S		SCMC_C		*t*	*p*
	M	SD	M	SD		
Percentage of error-free clauses	79.68	12.01	82.34	18.72	-2.32	0.03
Percentage of error-free AS-unit	70.60	13.55	73.44	21.85	-1.58	0.13
Percentage of accurate use of grammatical verbs	88.74	8.99	89.34	18.48	-2.05	0.06

4.4.2.2 Syntactic Complexity

Table 20 summarizes the descriptive and inferential statistics for each measure of syntactic complexity across task versions. The results showed that the complex task led to greater syntactic complexity for all three measures. However, the differences were not statistically significant.

Table 20

Mean, Standard Deviation, and Paired t-tests Results for Syntactic Complexity Measures of the SCMC Group

	SCMC_S		SCMC_C		*t*	*p*
	M	SD	M	SD		
Number of clauses per AS-unit	1.67	0.39	1.76	0.43	-1.42	0.17
Number of dependent clauses per AS-unit	0.45	0.24	0.53	0.29	-1.05	0.31
Number of words per AS-unit	13.89	2.91	14.42	3.20	-1.78	0.09

4.4.2.3 Lexical Variation

Table 21 illustrates the descriptive and inferential statistics for each measure of lexical variation across task versions. As we can see, greater lexical richness was identified in the simple task for all three measures. The effect of task complexity measured by TTR was found to be statistically significant, $t(18) = 6.40$, $p < 0.001$. Nevertheless, paired *t*-tests for D-value and MTLD did not reveal any statistically significant effects of cognitive task complexity.

Considering that the variation in the text length of L2 production was relatively large (minimum = 73, maximum = 650) and TTR is easily affected by text length (Malvern & Richards, 2002), the significance shown in the TTR measure results are not surprising. The D-value and MTLD measure results indicate that lexical variation was not significantly influenced by task complexity.

Table 21

Mean, Standard Deviation, and Paired t-tests Results for Lexical Variation Measures of the SCMC Group

	SCMC_S		SCMC_C		*t*	*p*
	M	SD	M	SD		
TTR	0.48	0.08	0.32	0.10	6.40	0.00
D-value	40.48	8.89	38.70	11.35	0.12	0.90
MTLD	37.30	8.03	35.04	8.01	0.40	0.69

4.4.2.4 Fluency

Table 22 demonstrates the descriptive and inferential statistics for each measure of fluency across task versions. As shown in Table 22, text fluency, measured by the number of syllables per minute, was greater on the simplex task than the complex task. In contrast, general repair fluency and verbatim repetition had a slight increase in the complex task compared with the simple task. Substitutive repetition was similar across cognitive conditions. None of the effects were statistically significant.

Table 22

Mean, Standard Deviation, and Paired t-tests Results for Fluency Measures of the SCMC Group

	SCMC_S		SCMC_C		*t*	*p*
	M	SD	M	SD		
Number of syllables per minute	134.08	28.02	126.17	32.28	0.80	0.43
Percentage of repeated words/phrases	4.59	6.57	4.99	3.87	-0.29	0.77
Percentage of verbatim repetitions	2.00	2.37	2.47	2.07	-1.00	0.33
Percentage of substitutive repetitions	2.59	4.73	2.55	3.43	0.10	0.92

In sum, for the SCMC group, increasing cognitive demands led to significantly more accurate use of the target language, as measured by error-free clause per AS-unit. The results show that task complexity level did not result in statistically significant changes in syntactic complexity, lexical variation, and fluency.

4.4.3 Linguistic Performance: FTF vs. SCMC

The third research question asked whether the effects of task complexity on NNSs' L2 oral production differ in FTF vs. SCMC mode. To answer this question, the linguistic performance of the FTF group was compared to that of the SCMC group across task versions. A series of mixed-design ANOVAs were performed to identify any statistically significant differences between the two groups. The within-subject variable was task complexity, and the between-subject variable was the interaction mode. The results are reported in the following sections for each linguistic dimension.

4.4.3.1 Accuracy

Table 23 summarizes the results from the multivariate analysis for the effects of task complexity and interaction mode on accuracy measures. The results did not reveal any statistically significant main effects for task complexity or task modality on any of the three measures of accuracy. Nevertheless, a medium effect size was found for the main effect of task complexity on accuracy as measured by the percentage of error-free clauses, $F(1, 40) = 3.46$, $p = 0.07$, partial $\eta^2 = 0.08$. The interaction effect between task complexity and task modality was not statistically significant. However, medium effect sizes were obtained for the interaction effect, indicating that the impact of task complexity on accuracy was moderately affected by the interaction mode.

Table 23

Mixed-design ANOVA Results for Task Complexity and Mode Effects on Accuracy Measures

Source		SS	df	MS	F	*p*	Partial η^2
Task complexity							
	Error-free clauses (%)	0.02	1	0.02	3.46	0.07	0.08
	Error-free AS-unit (%)	0.01	1	0.01	0.29	0.59	0.01
	Accurate use of Grammatical verbs (%)	0.01	1	0.01	1.46	0.23	0.04
Mode							
	Error-free clauses (%)	0.01	1	0.01	0.29	0.59	0.01
	Error-free AS-unit (%)	0.06	1	0.06	1.53	0.22	0.04
	Accurate use of Grammatical verbs (%)	0.01	1	0.01	0.35	0.56	0.01

Table 23 Continued

Source	SS	df	MS	F	p	Partial η^2
Task complexity * Mode						
Error-free clauses (%)	0.02	1	0.02	3.77	0.06	0.09
Error-free AS-unit (%)	0.04	1	0.04	3.16	0.08	0.08
Accurate use of grammatical verbs (%)	0.02	1	0.02	2.91	0.10	0.07

4.4.3.2 Syntactic Complexity

Table 24 presents results from the multivariate analysis for the effects of task complexity and interaction mode on syntactic complexity measures. The results showed that there was a significant main effect of task complexity when syntactic complexity was measured by the number of clauses per AS-unit, with a large effect size, $F(1, 40) = 5.35$, $p = 0.03$, Partial $\eta^2 = 0.12$; and by the number of words per AS-unit, with a large effect size, $F(1, 40) = 4.93$, $p = 0.03$, Partial $\eta^2 = 0.11$. No significant main effects were found for the interaction mode (FTF vs. SCMC) or for the interaction between task complexity and mode. This indicates that the effects of cognitive demands on syntactic complexity were not moderated by the interaction mode.

Table 24

Mixed-design ANOVA Results for Task Complexity and Interaction Mode Effects on Syntactic Complexity Measures

Source	SS	df	MS	F	*p*	Partial η^2
Task complexity						
# of clauses per AS-unit	0.53	1	0.53	5.35	0.03	0.12
# of dependent clauses per AS-unit	0.15	1	0.15	2.55	0.12	0.06
# of words per AS-unit	18.49	1	18.49	4.93	0.03	0.11
Mode						
# of clauses per AS-unit	0.32	1	0.32	1.03	0.32	0.03
# of dependent clauses per AS-unit	0.34	1	0.34	2.01	0.16	0.05
# of words per AS-unit	4.57	1	4.57	0.28	0.60	0.01
Task complexity * Mode						
# of clauses per AS-unit	0.00	1	0.00	0.00	0.99	0.00
# of dependent clauses per AS-unit	0.00	1	0.00	0.03	0.87	0.00
# of words per AS-unit	0.82	1	0.82	0.22	0.64	0.01

4.4.3.3 Lexical Variation

Table 25 demonstrates the results from the multivariate analysis for the effects of task complexity and interaction mode on lexical variation measures. Once again, given the fact that the length of learners' output varied greatly and the sensitivity of TTR to text length, we mainly considered the results from the D-value and MTLD. As shown in Table 25, no statistically significant differences were found for task complexity, mode, or the interaction between the two independent variables, suggesting that changes in cognitive demands or the interaction mode did not impact the lexical richness of learners' L2 production.

Table 25

Mixed-design ANOVA Results for Task Complexity and Interaction Mode Effects on Lexical Variation Measures

Source		SS	df	MS	F	*p*	Partial η^2
Task complexity							
	TTR	0.34	1	0.34	96.07	0.00	0.72
	D-value	0.01	1	0.01	0.00	0.99	0.00
	MTLD	2.26	1	2.26	0.17	0.68	0.00
Mode							
	TTR	0.03	1	0.03	3.38	0.07	0.08
	D-value	0.89	1	0.89	0.01	0.93	0.00
	MTLD	265.08	1	265.08	3.00	0.09	0.07
Task complexity * Mode							
	TTR	0.01	1	0.01	2.00	0.16	0.05
	D-value	1.57	1	1.57	0.03	0.86	0.00
	MTLD	7.64	1	7.64	0.17	0.68	0.00

4.4.3.4 Fluency

Table 26 demonstrates the results from the multivariate analysis for the effects of task complexity and interaction mode on fluency measures. There was no statistically significant effect for task complexity or for the interaction between the two independent variables. The only significant effect was found for the interaction mode as measured by verbatim repetitions, with a medium effect size, $F(1, 40) = 3.98$, $p = 0.05$, Partial $\eta^2 = 0.09$. The results suggested that task complexity did not play a significant role in fluency, but the mode of interaction moderately affected the frequency of learners' verbatim repetitions.

Table 26

Mixed-design ANOVA Results for Task Complexity and Interaction Mode Effects on Fluency Measures

Source	SS	df	MS	F	*p*	Partial η^2
Task complexity						
# of syllables per minute	28.01	1	28.01	0.26	0.61	0.01
Repeated words/phrases (%)	0.00	1	0.00	1.39	0.25	0.04
Verbatim repetitions (%)	0.00	1	0.00	0.04	0.84	0.00
Substitutive repetitions (%)	0.00	1	0.00	1.10	0.30	0.03
Mode						
# of syllables per minute	1423.86	1	1423.86	1.29	0.26	0.03
Repeated words/phrases (%)	0.01	1	0.01	2.58	0.12	0.06
Verbatim repetitions (%)	0.01	1	0.01	3.98	0.05	0.09
Substitutive repetitions (%)	0.00	1	0.00	0.14	0.72	0.00
Task complexity * Mode						
# of syllables per minute	77.78	1	77.78	0.73	0.40	0.02
Repeated words/phrases (%)	0.00	1	0.00	2.79	0.10	0.07
Verbatim repetitions (%)	0.00	1	0.00	1.74	0.19	0.04
Substitutive repetitions (%)	0.00	1	0.00	0.81	0.37	0.02

In sum, these findings suggested that the effects of task complexity were not greatly moderated by the interaction mode. The only significant interaction effect was found for accuracy. The SCMC group had more accurate text production in the complex task. In contrast, the FTF group had more accurate text production in the simple task. As for syntactic complexity, NNSs performed slightly better on the more cognitively demanding task in both modes. Nevertheless, the differences were not statistically significant. Regarding lexical richness, both groups performed equally well on the two tasks. Finally, task complexity did not significantly affect speech fluency. For repair fluency, increased task complexity led to fewer repetitions for the FTF group, with no effect for the SCMC group. The interaction mode, however, appeared to influence verbatim repetitions – the FTF group produced significantly more verbatim repetitions on both tasks than the SCMC group.

4.4.4 Validation of Task Complexity Manipulation

The last research question concerns whether the assumed task complexity effects actually posed different cognitive demands on the participants as intended. The two independent measures of task complexity employed in the study were self-ratings of task difficulty and retrospective time duration judgment. For learners' perception of task difficulty, an overall difficulty score for each task was calculated for each person. For time duration judgment, the subjective-to-objective ratio (i.e., the ratio of estimation to actual time duration) for each task was calculated for each person. Separate paired *t*-tests were performed on the two dependent variables – perceived task difficulty and time duration judgment.

4.4.4.1 Perceived Task Difficulty

The descriptive statistics for NNSs' perceived task difficulty are presented in Table 27. As we can see, learners from both FTF and SCMC groups reported that the complex task was more challenging for them to perform. The FTF_C and SCMC_C conditions had the highest scores, followed by the SCMC_S condition. The FTF_S condition had the lowest score.

The results from the *t*-tests confirmed the above observations that the scores on the complex task were significantly higher than those on the simple task for both the FTF group, *t*

(20) = 6.01, $p < .001$, and the SCMC group, t (18) = 2.69, p =0.02. The findings suggested that the NNSs perceived the complex task as more difficult than the simple task, regardless of the task modes.

Table 27

Descriptive Statistics for Self-Ratings of Task Difficulty

Group	n	M	SD
FTF_S	21	1.67	0.80
FTF_C	21	2.90	1.14
SCMC_S	19	2.16	1.17
SCMC_C	19	2.89	1.10

4.4.4.2 Retrospective Duration Judgment

The descriptive statistics for time duration judgment are reported in Table 28. It appeared that the subjective-to-objective ratios were always greater than one across task versions and interaction modes, which indicated that the estimated time was always greater than the actual time learners spent on the task, regardless of the complexity level and the interaction mode. The descriptive statistics also suggested that learners in the FTF group judged the task as taking more time than the SCMC group, regardless of the cognitive level.

Results from the *t*-tests indicated that changes in task complexity levels did not lead to significant differences in the duration judgments for both the FTF group, t (20) = 0.44, p = 0.67, and the SCMC group, t (18) = 0.34, p =0.74. The findings suggested that task complexity operationalization was not confirmed by the data from the duration judgments.

Table 28

Descriptive Statistics for Time Duration Judgment in Minutes

	Estimated time		Actual time		Ratio	
	M	SD	M	SD	M	SD
FTF_S	3.10	1.72	1.76	0.71	1.73	0.62
FTF_C	5.62	3.85	3.39	1.78	1.67	0.81
SCMC_S	1.98	1.39	1.42	0.59	1.36	0.63
SCMC_C	3.68	1.91	2.89	1.06	1.30	0.58

In sum, the two independent measures of cognitive task complexity revealed different results. According to the task difficulty questionnaire, the complex task was perceived as more difficult and challenging to perform by both groups. On the other hand, task complexity manipulation did not lead to significant discrepancies in retrospective duration judgment.

4.5 Discussion

4.5.1 Task Complexity, Interaction Mode, and Linguistic Performance in Mathematical Word Problem-Solving

This study investigated the impact of cognitive task complexity and mode on learners' L2 production related to solving math word problems, assessed in terms of accuracy, syntactic complexity, lexical variation, and fluency. With regard to task complexity effects, the results revealed that increasing task complexity led to significantly greater accuracy, as measured by the percentage of error-free clauses, in the SCMC mode, but not in the FTF mode. For syntactic and lexical complexity, the results did not show any significant effects of cognitive demands in either FTF or SCMC setting. As for fluency, the only significance was found for the general repair fluency, measured by the percentage of repeated words and phrases, in the FTF mode but not in the SCMC mode. Concerning the effects of mode, the FTF interaction induced significantly more verbatim repetitions than the SCMC interaction, with no significance for other indices of L2 performance.

Statistical significance, however, is not the only criterion to gauge the effects of the independent variables. Since *p* value is greatly affected by sample size (Moore et al., 2012), it is possible that the lack of significance was due to the relatively small sample size (n = 40) in the present study. In order to obtain a better understanding of the results, effect sizes were then utilized as an additional criterion to evaluate the impact of the independent variables. According to Synder and Lawson (1993), the effect size is independent of sample size and measures the magnitude of effects between variables.

Analysis of the effect sizes showed that task complexity had small to medium effects on accuracy, medium to large effects on syntactic complexity, and small effects on fluency. There was no significant effect size for lexical variation, however. The analysis also revealed medium interaction effects of the two variables on accuracy and small interaction effects on fluency. No interaction effects were found for syntactic and lexical complexity.

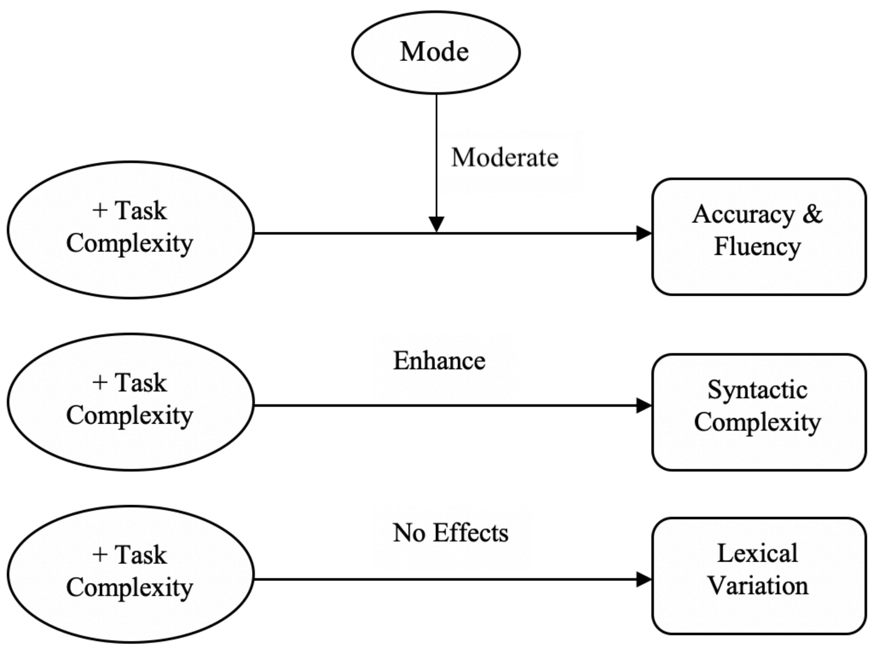
Combining the results of the *p* values and effect sizes, it can be concluded that task complexity had limited effects on accuracy and fluency, and the effects were moderated by the interaction mode. In other words, task complexity affected accuracy and fluency differently depending on the environment in which the interaction was conducted. Moreover, increasing task complexity led to enhanced syntactic complexity regardless of the interaction mode. Finally, neither task complexity nor the mode affected the lexical richness of learners' L2 production in math. The relationships between task complexity, mode, and L2 performance are illustrated in

Figure 7. These findings indicate that cognitive aspects of task design and interaction environment can affect learners' linguistic performance in content areas in some dimensions but not in others.

The findings in the present study did not support Robinson's (2005) Cognition Hypothesis, which predicts that increasing task complexity along resource-dispersing variables negatively affects complexity (i.e., syntactic and lexical), accuracy, and fluency of L2 output. The findings are also in contrast with those of previous research, which investigated the impact of cognitive demands on L2 performance within the resource-dispersing dimension (e.g., Gilabert, 2007a; Kormos & Trebits, 2012; Tavakoli, 2009).

Figure 7

The Relationships between Task Complexity, Mode, and L2 Production in Mathematics



The differences in the findings between the present study and previous research could be due to the specific characteristics of mathematical word problem tasks compared to regular language learning tasks used in classrooms, such as picture-story tasks and spot-the-difference tasks. Most experimental tasks in previous task complexity research are utilized to elicit learners' daily communicative language. In contrast, the tasks used in the present study required learners to use academic language appropriate for mathematics as a subject matter. While engaging in solving the complex version of the mathematical task, probably the NNSs needed to keep both the relevant elements in the problem and the strategies that were necessary for solving the problem in their short-term memory, and therefore had additional functional demands (for information integration) in performing the task.

This helps explain why syntactic complexity was enhanced in the more complex task in both FTF and SCMC modes. It is likely that the more complex task, though requiring the same mathematical knowledge, involved higher element interactivity (i.e., the levels of element connectedness) in the problem and therefore entailed a more complicated problem-solving process. In order to accurately depict the complicated process, the NNSs had to use longer sentences with more complex sentence structures, which led to higher syntactic complexity.

As for lexical complexity, highly similar values were found for the two versions of the task across modes. A possible explanation is that in both mathematical tasks, the NNSs' attention might have been dispersed to conceptualize the problem-solving process, which might have directed their attention to syntactic complexity and left insufficient attentional resources for lexical encoding. Consequently, the NNSs could only use vocabulary that was easily accessible from their existing language system. This process seems to indicate a trade-off effect between syntactic and lexical complexity in the L2 oral production for math word problems. Another likely explanation is that using a wide range of vocabulary might not be necessary for the explanation of the problem-solving process in the mathematical context. The two main features of the mathematical language are precision and unambiguity, which distinguishes it from daily communicative language. As long as the abstract and logical ideas could be accurately expressed, the NNSs might not feel obligated to take great effort to use different words or phrases in the explanation. Moreover, the language of mathematics involves a series of technical terms that are specific to mathematical discourse, which might be another restriction on lexical variation.

The study findings revealed that the interaction mode played a moderating role in the effects of task complexity on accuracy. Increasing task complexity led to greater accuracy but only for the SCMC group. As we know that the articulation of the problem-solving process was undertaken following the discussion of the word problem between the NSs and the NNSs. Whereas the FTF group conducted the oral discussion, the SCMC group interacted through text-based online chats. Previous research has suggested that SCMC can better facilitate noticing of self-errors and feedback (Lai & Zhao, 2006; Yuksel & Inan, 2014) than FTF communication. Therefore, it is possible that NNSs in the SCMC group had more instances of noticing during the discussion than their peers in the FTF group. This may in turn reinforce learners monitoring their L2 output during the narration. While performing the complex task, NNSs in the SCMC group may have contributed more attentional resources to both the content and the language to ensure an accurate articulation of the problem-solving process. In contrast, NNSs in the FTF group may have paid much attention to the content and little attention to, in particular, the accuracy aspect of, the language while performing both tasks.

Finally, task complexity had limited effects on fluency, and the interaction mode moderated these effects. Increasing task complexity led to fewer repetitions (i.e., the sum of verbatim and substitutive repetitions) for the FTF group and no effects for the SCMC group. The results from the FTF group corroborate those by Skehan and Foster (2008), who also found fewer repetitions in NNSs' narratives as task complexity increased along the resource-dispersing dimension (i.e., +/- planning). Since repair fluency has been viewed as an indicator of attention to form (Tavakoli & Skehan, 2005), it seems to be the case that NNSs in the FTF group paid less attention to the language while performing the cognitively complex task. This finding is aligned with the prediction of the Cognition Hypothesis that increasing task complexity along resource-dispersing dimension depletes attentional resources (Robinson, 2005). On the other hand, the finding for the SCMC group that no effects of task complexity were detected, is in line with our previous speculation that the NS-NNS online interaction may have had more instances of noticing, which may then have led to reinforced attention to form in the subsequent narrative task performance. The increased cognitive demands may have depleted attentional resources, but the NS-NNS interaction in SCMC may have promoted attention to form. As a consequence, no effects of task complexity on fluency could be detected for the SCMC group.

The interaction mode also appeared to be a significant factor for fluency. The FTF group had significantly higher percentages of verbatim repetitions than the SCMC group across cognitive conditions. In other words, NNSs in the FTF group spent more time thinking of the appropriate words to use during the narration. This finding implies that compared to their peers in the FTF group, NNSs in the SCMC group may have had better access to their interlanguage repertoire during the narration. This again supports our hypothesis that more instances of noticing may have occurred in the SCMC interaction prior to the narration. Since noticing is necessary for input to become intake (Schmidt, 2001), NNSs in the SCMC group may have processed input in the feedback to a great extent during the NS-NNS interaction. This may then have eased the demands on accessing mathematical vocabulary and expressions in their current L2 knowledge during the real-time performance. However, these assumptions need to be empirically investigated before any conclusion is drawn. As Sasayama (2016) has noted, “the relationship between task design and its cognitive complexity is more complicated than researchers previously seem to have assumed” (p. 248). Subject knowledge, which is cognitively demanding could further add complication to this relationship. Therefore, in order to provide pedagogical implications for teachers to design, sequence, and implement tasks with varying degrees of task complexity in content areas, future research is needed in this particular area.

4.5.2 Validation of Task Complexity Manipulation

The validity of task complexity manipulation has been largely overlooked in research that examines the relationship between cognitive demands and linguistic performance. Only a few studies have provided empirical evidence for the validation of task complexity (e.g., Baralt, 2013; Ishikawa, 2011; Lee, 2019; Révész et al., 2014; Sasayama, 2015). The present study employed self-perceived task difficulty and retrospective duration judgment to evaluate the cognitive loads of the experimental tasks. For self-ratings of task difficulty, as anticipated, the participants perceived the complex task as more difficult to perform than the simple task, suggesting that the manipulation of task complexity led to desired changes in cognitive load. However, there was no significant discrepancy in duration judgment between the simple and the complex tasks. In addition, learners in the FTF group judged the narrative tasks (both simple and complex) as taking more time than the SCMC group, although the differences were not significant. According to Block et al. (2010), when the duration of a task is judged retrospectively, the longer time estimation is, the more cognitively demanding the task is. In light of this, the findings suggest that the FTF group perceived both tasks as more cognitively challenging than the SCMC group.

Taken together, the findings from the two independent measures of task complexity suggest that the intended manipulation of task complexity induced different levels of cognitive loads, and self-ratings of task difficulty may be a more reliable way to validate task complexity manipulation than time duration judgment. These findings, however, were not aligned with Baralt's (2013) study findings as Baralt found that retrospective duration estimation was a more reliable strategy to validate task complexity assumptions than perceived task difficulty. Given the limited number of studies that employ retrospective duration judgment as a measure of task complexity, the inconsistent findings indicate a need for future research to explore the effectiveness of duration judgment in measuring the cognitive complexity of tasks.

4.6 Conclusion and Implications

The purpose of the study was to investigate the relationship between cognitive task complexity, mode (i.e., FTF or SCMC), and L2 performance in the mathematics content area. Overall, the findings indicate that both task complexity and mode affect learners' language performance in mathematics and the impact of each factor varied across different linguistic dimensions.

First, the effects of task complexity on accuracy and fluency were moderated by the mode, which suggests the potential of mode as a variable in pedagogical task design. Second, syntactic complexity was enhanced along with greater cognitive demands, regardless of the mode. This finding was interpreted by considering the inherent features of the mathematical context in the study. It appears that increasing cognitive demands in academic content induced additional functional demands, which pushed the learners to use sophisticated sentence structures in L2 production, especially when the cognitive load increased. Third, lexical variation was not impacted by task complexity or mode. This could be attributed to the two main features of mathematical language: precision and unambiguity. It seems that a great lexical richness is not a critical component of mathematical task performance. Taken together, the findings echo Ortega's (2015) arguments that L2 learning in content areas can be affected by both performative and contextual factors, and a careful design of pedagogical tasks can improve learners' L2 performance in content areas. More content-based explorations are needed to expand our understanding of the impact of task complexity on academic language performance in FTF and SCMC modes. The study also attempted to validate task complexity manipulation by employing independent measures of task complexity. The findings confirmed that the design-to-be more complex task was indeed more cognitively demanding, and suggested that self-ratings of task difficulty may be a more reliable measure of task complexity than retrospective duration judgment.

While the findings in the present study provide insights for future research, some limitations should be acknowledged. First, the study only investigated the effects of task complexity with mathematical word problems. The findings, therefore, cannot be generalized to other content areas. Future research that explores the impact of task complexity on academic language development in different subjects is warranted. The second limitation pertains to methodological concerns in the construct of task complexity. The present study only involved two levels of task complexity. Nevertheless, as Kim (2012) has pointed out, task complexity should be operationalized "as a continuum rather than as a dichotomous construct" (p. 652). Future research would benefit from including multiple degrees of task complexity to more closely examine the effect of tasks with different levels of task complexity on learners' performance. Finally, as previously mentioned, the lack of statistical significance in the study could be due to the relatively small sample size. A better understanding of the relationship between task complexity, mode, and L2 performance in content areas may be obtained if researchers include a larger sample size in future studies.

To conclude, the current study contributes to the field with both theoretical and pedagogical implications. In the theoretical perspective, the study tested the Cognition Hypothesis and found that the predictions of the hypothesis may not apply to the academic context. That is, task complexity may play a different role in developing academic language compared to conversational language. Moreover, the study also provided additional empirical evidence related to how task complexity operates in online environments. The study has pedagogical implications also. Although the study was conducted in a laboratory setting, it has implications for classroom teaching concerning task design for academic language development. The study suggests that appropriate incorporation of more cognitively demanding tasks in the SCMC environment may promote the development of accuracy and syntactic complexity of learners' L2 in content areas.

CHAPTER V

CONCLUSION

5.1 Summary

This dissertation aimed to: (a) validate the intended task cognitive demands of the experimental tasks by employing independent measures of cognitive load (i.e., self-ratings of task difficulty & retrospective duration judgment); and (b) investigate the effects of task complexity and modality (FTF vs. SCMC) on NS-NNS interaction and NNSs' L2 performance in the mathematical content area. The participants were 82 undergraduate and graduate students (55 males, 27 females) at a public university in the United States. Forty-one NS-NNS dyads were formed and then assigned to either the FTF (N = 21) group or the SCMC group (N = 20). The experimental tasks were two mathematical word problems, which required the same mathematical knowledge but varied in cognitive demands. Two levels of task complexity (i.e., simplex and complex) were operationalized along [+/- few steps] based on Robinson's (2007a) Triadic Componential Framework for pedagogical task design and sequencing. Self-ratings of task difficulty and retrospective duration judgment were adopted as the independent measures of task complexity. Each NS-NNS dyad discussed both word problems in FTF or online setting depending on the group they were assigned to. After the discussion, the NNSs then articulated the problem-solving process to the researcher. The NS-NNS interaction and NNSs' L2 output were recorded for data analysis. The NS-NNS interactions were evaluated based on the occurrence of LREs and self-initiated repairs. NNSs' L2 oral narratives were assessed in terms of accuracy, syntactic complexity, lexical diversity, and fluency.

The data analyses generated a series of findings in relation to the validity of task complexity operationalization and the effects of task complexity and modality on interaction and L2 performance in mathematical content.

With regard to the validity of task complexity manipulation, the results showed that (a) for the interaction part, participants perceived the designed-to-be more complex task as significantly more difficult and more time consuming than the designed-to-be simpler task; (b) for the narration part, learners also rated the complex task as significantly more difficult to complete, but no significant difference was detected in the duration judgments. Overall, the findings indicated that the designed-to-be more complex task was indeed more cognitively demanding than the simple task, affirming the validity of task complexity manipulation. Furthermore, the differences in the findings between the interactive and the narrative parts of the task performance suggested that task type might be a moderating factor that affects learners' retrospective time judgment. As such, self-perceived task difficulty is a more reliable measure of task complexity than retrospective time judgment across task types. This also highlights the importance of using multiple measures of cognitive load in task complexity research.

As for the effects of task complexity on NS-NNS interaction, the results showed that there were no significant differences in the mean rates of self-initiated repairs and LREs between the simple and complex tasks in either FTF or online modes. Nevertheless, the complex task yielded a higher rate of successful uptake than the simple task in both FTF and SCMC modes. The findings partially support the Cognition Hypothesis, which claims that increasing cognitive task complexity promotes negotiation of meaning, noticing, and uptake of more salient input in feedback (Robinson, 2011). Furthermore, the results also demonstrated that significantly more self-initiated repairs and LREs were yielded in the FTF mode than in the SCMC mode, irrespective of the cognitive condition. Overall, the findings suggest that task modality and task complexity play critical roles in facilitating noticing of one's own linguistic errors and noticing of feedback, respectively.

Turning to the effects of task complexity on learners' L2 production, overall, the results showed that task complexity had different effects on various aspects of linguistic performance, and the effects were moderated by task mode. The results showed that increasing task complexity led to greater accuracy for tasks in SCMC mode, as measured by the percentage of error-free clauses. There was, however, no significant effect on accuracy for the FTF mode. As for fluency, the only significance was found for the general repair fluency – fewer repetitions on the complex task – in the FTF mode, but not in the SCMC mode. The results also showed that task complexity did not significantly affect lexical variation across interaction modes. Finally, syntactic complexity was enhanced along with increased cognitive demands in both FTF and SCMC modes. Concerning the effects of the interaction mode, the FTF group had significantly more verbatim repetitions than the SCMC group. There were no significant effects for other indices of L2 performance. These findings indicate that cognitive aspects of task design and task performance mode can affect learners' linguistic performance in mathematical problem-solving tasks in some linguistic dimensions but not in others.

5.2 Implications

5.2.1 Theoretical Implications

This study has implications for the Cognition Hypothesis (Robinson, 2001a, 2001b, 2005, 2011), which claims that increasing the cognitive demands of tasks will "promote interaction and negotiation work, and heightened attention to, noticing of, and incorporation of forms made salient in the input..." (Robinson, 2005, p. 3). The study confirmed that increasing task complexity leads to enhanced incidental noticing of linguistic forms as measured by successful uptake. Nevertheless, the amount of meaning negotiation, as evidenced by LREs, seemed not to be significantly affected by changes in cognitive complexity. This implies that some predictions of the Cognition Hypothesis concerning the effects of task complexity on interaction, may not be applicable to subject matter content, at least in the case of mathematical word problems. Since learners' previous subject knowledge can, at least partially, compensate for their language deficiency (Li & Chen, 2019; Usó-Juan, 2006), negotiation of meaning due to linguistic errors may not be significantly influenced by simply increasing cognitive demands of the task. By contrast, increasing task complexity led to higher percentages of more explicit feedback (e.g., metalinguistic), as evident in both the FTF and SCMC conditions in the present study. This may help explain why the more cognitively demanding task yielded higher percentages of successful uptake in both FTF and SCMC interactions. Based on these findings, I propose that increasing the cognitive demands of pedagogical tasks in content areas does not necessarily induce a greater amount of meaning negotiation due to linguistic errors. Rather, increasing cognitive task complexity promotes attention to, noticing of, and incorporation of forms in corrective feedback.

Robinson (2005) has also made claims pertaining to the relationship between task complexity and L2 performance in the Cognition Hypothesis. Robinson has proposed that increasing task complexity along resource-dispersing dimensions negatively affects the complexity (syntactic and lexical), accuracy, and fluency of learners' L2 production. Findings in the present study, however, did not support this proposal. Analyses of NNSs' L2 output indicated that task complexity had significantly positive effects on syntactic complexity, limited effects on accuracy and fluency, and no effect on lexical variation. These findings should be interpreted by considering the inherent features of the mathematical content in the study.

First, it appeared that the more complex task, which was designed by involving more steps in problem-solving based on the simpler task, led to higher element interactivity (i.e., the levels of element connectedness) or intrinsic cognitive load as suggested in Sweller's (2010) Cognitive Load Theory. In order to articulate the sophisticated problem-solving process in the academic content, the NNSs may have had to keep both the relevant elements in the problem and the strategies that were necessary for solving the problem in their short-term memory, and therefore induced additional functional demands (for information integration) or germane cognitive load as mentioned in Sweller's (2010) Cognitive Load Theory. To meet these demands, the NNSs may have needed to use longer sentences with more complex sentence structures, which led to greater syntactic complexity. Second, it appears that the NNSs did not pay much attention to the accuracy of the language during the articulation of the problem-solving process, even if the cognitive complexity of the task increased. This may be due to the fact that the academic content was already highly cognitively demanding (compared to that in traditional language learning tasks such as picture-story tasks) and thus left little attentional resources for the language. This is consistent with Van Patten's (1990) argument that when attentional resources are limited, learners tend to prioritize meaning over form. Third, the fluency of learners' L2 output was likely to be influenced by learners' attention control and lexical access (Segalowitz, 2007). Based on our hypothesis that the NNSs would not pay much attention to the language during task performance, lexical access then may have been the main underpinning variable for fluency. The finding that task complexity did not affect most of the measures of fluency then could be interpreted as increasing task complexity along resource-dispersing dimensions, in particular, the [+/- few steps] variable, did not impede the access to learners' existing interlanguage system in the mathematical problem-solving tasks. Finally, the lack of significance on lexical variation across cognitive conditions may be due to three possible reasons: (a) the limited attentional resources devoted to language only allowed learners to use vocabulary that was easily accessible from their existing language system; (b) a wide range of vocabulary might not be necessary for the explanation of the problem-solving process in the mathematical context; (c) the language of mathematics involves a series of technical terms that are specific to mathematical discourse. To sum up, based on these findings, I propose that increasing task complexity along the [+/- few steps] variable promotes syntactic complexity, and has limited or no effects on other aspects (accuracy, lexical variation, and fluency) of linguistic performance in mathematical content.

5.2.2 Methodological Implications

The findings of this study also have methodological implications. First, the study highlights the importance of including different measures of cognitive load for task complexity validation. Only with more than one independent measure of task complexity, can the results be triangulated for a more accurate interpretation.

Second, collecting qualitative data is imperative for task complexity research. While quantitative data promotes the objective investigation of the relationships between the variables, qualitative data such as observation notes enables us to gain a more panoramic view of the phenomena. Without the observation notes in this study, it would not have been possible to reveal that learners were more cautious about what they sent, for example, reading and checking the messages before sending them out, while performing the complex task than the simple task in the SCMC mode. In this sense, the observation notes provided unique insights into learners' internal processing during task performance, and thus offered in-depth explanatory information regarding the relationships between the variables in the study.

5.2.3 Pedagogical Implications

Although more explorations are needed to make a conclusive argument regarding the relationships between task complexity, task modality, and L2 learning, the study makes valuable contributions to L2 pedagogy. First and foremost, the findings suggest that teachers should sequence pedagogical tasks from cognitively simple to complex in an effort to enhance students' attention to linguistic forms in subject matter classrooms. Second, pedagogical tasks can be performed through online interaction and should also be incorporated in a sequenced order from cognitively simple to complex, as in the regular FTF classroom. Third, since academic content may add additional functional demands on learners, teachers should be cautious when manipulating task complexity levels to avoid cognitive overload. Finally, learners' academic lexical knowledge may not be significantly reinforced by simply manipulating cognitive task complexity. Therefore, teachers may need to operate on other task design features to expand learners' academic vocabulary.

5.3 Limitations and Future Research

5.3.1 Limitations

Although the study offers valuable insights into task complexity and L2 development research, it has certain limitations that need to be addressed. First, the participants were studying various majors in the university. Even though their mathematical knowledge levels were appropriate for the level that was required by the experimental tasks, and they all were able to solve the mathematical problems with their partners, their areas of study could have moderated the results. It is, therefore, suggested that future researchers include participants from one field of study and examine if the findings would be similar to our findings or not.

Second, learners' incidental noticing during the NS-NNS interaction was assessed by discourse moves (i.e., LREs and self-initiated repair instances) rather than direct measures such as stimulated recall. Although the concurrent data (i.e., the researcher's observation notes) helped explicate the results, the measures still could not unveil learners' internal processes during task performance. Direct measures of noticing are needed in future research to help us obtain insights into the internal processes during task performance.

Third, only two levels of task complexity were involved in the study. Since cognitive task complexity is considered to be a continuum rather than a dichotomous construct (Kim, 2012), multiple levels of task complexity should be operationalized in order to examine the distinctiveness among tasks and the effects of different levels of task complexity on L2 performance.

Finally, the study only examined the effects of task complexity on language use and development in a mathematical problem-solving task. The findings, therefore, cannot be generalized to other content areas. Research evidence is needed regarding the effects of task complexity on L2 learning across different content areas.

5.3.2 Future Research

It would be interesting to compare the result of the present study with that of future research, which employs direct measures of noticing to explore the relationship between task complexity and incidental noticing. A review of current literature reveals four direct measures of noticing that can be used in future research, including (a) immediate cued recall (e.g., Egi, 2004; Kim et al., 2015), (b) stimulated recall (e.g., Egi, 2010; Gass & Mackey, 2000), (c) think-alouds (e.g., Kim & Bowles, 2019; Sachs & Polio, 2007), and (d) eye-tracking technology (e.g., Smith, 2010, 2012). Researchers need to select the proper measure depending on the design features of the study.

Whether the noticing occurrences indeed leads to subsequent academic language development is another topic that is worth exploration. In this study, significantly higher rates of successful uptake were detected upon the completion of the more cognitively demanding task. While successful uptake provides some evidence of L2 development, future research can use tailor-made assessments by including the actual linguistic items that arise from the interaction to further investigate the actual L2 learning outcomes from the completion of the tasks.

In addition, empirically investigating the operationalization of cognitive task complexity in academic contents is highly warranted. In this study, the academic content influenced how task complexity was operationalized - the manipulation of [+/- few steps] led to greater intrinsic cognitive loads and additional functional demands for information integration, which facilitated syntactic complexity. It is unclear whether this was due to the subject content or language demands. This, however, is not mentioned in the Triadic Componential Framework (Robinson, 2007a) for task design and sequencing. Therefore, caution is needed when applying the Triadic Componential Framework in discipline-related research contexts. Related to this, multiple levels of task complexity along a combination of different variables should be involved in future research to better simulate learning in a real classroom.

REFERENCES

Adams, R., Alwi, N. A. N. M., & Newton, J. (2015). Task complexity effects on the complexity and accuracy of writing via text chat. *Journal of Second Language Writing, 29*, 64-81. https://doi.org/10.1016/j.jslw.2015.06.002

Aoki, K. (1995). Synchronous multi-user textual communication in international tele-collaboration. *Electronic Journal of Communication, 5*(4), 1-17.

Baralt, M. (2010). *Task complexity, the Cognition Hypothesis and interaction in CMC and FTF environments*. (Publication No. 3412595) [Doctoral dissertation, Georgetown University, Washington, DC]. ProQuest Dissertations and Theses Global.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition, 35*(4), 689-725. https://doi.org/10.1017/S0272263113000429

Baralt, M. & Leow, R. P. (2016). Uptake, task complexity, and L2 development in SLA: An online perspective. In R. Leow, L. Cerezo & M. Baralt (Eds.). *A psycholinguistic approach to technology and language learning* (pp. 3-22). De Gruyter Mouton.

Barwell, R. (2005). Integrating language and content: Issues from the mathematics classroom. *Linguistics and Education, 16*(2), 205-218. https://doi.org/10.1016/j.linged.2006.01.002

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*(1), 111-125.

Blake, C. (2009). Potential of text-based internet chats for improving oral fluency in a second language. *The Modern Language Journal, 93*(2), 227-240. https://doi.org/10.1111/j.1540-4781.2009.00858.x

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*(3), 330-343. https://doi.org/10.1016/j.actpsy.2010.03.006

Breen, M. P. (1987). Learner contributions to task design. In C. N. Candlin & D. Murphy (Eds.), *Lancaster Practical Papers in English Language Education: Language learning tasks* (pp. 23-46). Prentice Hall.

Bret Blasco, A. (2014). *L2 English young learners' oral production skills in CLIL and EFL settings: A longitudinal study*. [Doctoral dissertation, Universitat Autònoma de Barcelona, Spain].

Brown, J. D., Hudson, T., Norris, J. M., & Bonk, W. (2002). *An investigation of second language task-based performance assessments*. Technical Report No. 24. University of Hawaii, Second Language Teaching & Curriculum Center.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge University Press.

Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-146). McMillan Education.

Bygate, M., Skehan, P. and Swain, M. (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing*. Longman.

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

Chen, W. C., & Eslami, Z. (2013). Focus on form in live chats. *Journal of Educational Technology & Society, 16*(1), 147-158.

Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System*, *72*, 85-98. https://doi.org/10.1016/j.system.2017.10.010

Chun, D. M. (1994). Using computer networking to facilitate acquisition of interactive competence. *System, 22*, 17– 31. https://doi.org/10.1016/0346-251X(94)90037-X

Chun, D. M. (1998). Using computer-assisted class discussion to facilitate the acquisition of interactive competence. In J. Swaffar, S. Romano, P. Markley, & K. Arens (Eds.), *Language learning online: Theory and practice in the ESL and L2 computer classroom* (pp. 57–80). Labyrinth Publications.

Coniam, D., & Wong, R. (2004). Internet Relay Chat as a tool in the autonomous development of ESL learners' English language ability: An exploratory study. *System, 32*(3), 321-335. https://doi.org/10.1016/j.system.2004.03.001

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism, 10*(5), 543-562. https://doi.org/10.2167/beb459.0

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.

Crandall, J. & Tucker, G. R. (1990). Content-based language instruction in second and foreign languages. In S. Anivan (Ed.), *Language teaching methodology for the nineties* (pp. 83-96). SEAMEO Regional Language Centre.

Cummins, J. (1979). Cognitive academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism, 19*, 121-129.

Cummins, J. (2001). BICS and CALP. In M. Byram (Ed.), *Encyclopedia of language teaching and learning* (pp. 76-79). Rutledge.

Dale, T. C. & Cuevas, G. J. (1987). Integrating language and mathematics learning. In Crandall, J.A. (Ed.), *ESL through content-area instruction: Mathematics, science, social studies* (pp. 9-54). Prentice-Hall.

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles?. *Annual Review of applied linguistics, 31*, 182-204. https://doi.org/10.1017/s0267190511000092

Darhower, M. (2002). Interactional features of synchronous computer-mediated communication in the intermediate L2 class: A sociocultural case study. *CALICO Journal, 19*(2), 249-277. https://doi.org/10.1558/cj.v19i2.249-277

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology, 6*, 348-361. https://doi.org/10.3389/fpsyg.2015.00348.

de la Fuente, M. (2003). Is SLA interactionist theory relevant to CALL? A study on the effect of computer-mediated interaction in L2 vocabulary acquisition. *Computer Assisted Language Learning, 16*(1), 47–81. https://doi.org/10.1076/call.16.1.47.15526 de la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research, 10*(3), 263-295. https://doi.org/10.1191/1362168806lr196oa

Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *The Modern Language Journal, 94*(1), 1-21. https://doi.org/10.1111/j.1540-4781.2009.00980.x

Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning, 51*(1), 1-46. https://doi.org/10.1111/j.1467-1770.2001.tb00013.x

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press. https://doi.org/10.1177/003368820303400105

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

Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research*. Routledge.

Eslami, Z. R., & Kung, W. T. (2016). Focus-on-form and EFL learners' language development in synchronous computer-mediated communication: task-based interactions. *The Language Learning Journal, 44*(4), 401-417. https://doi.org/10.1080/09571736.2016.1227219

Foster, P., & P. Skehan. (1999). The effect of source of planning and focus on planning on task-based performance. *Language Teaching Research, 3*(3), 185–215. https://doi.org/10.1177/136216889900300303

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

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing, 30*, 45-57. https://doi.org/10.1016/j.jslw.2015.08.009

Fujii, A., & Mackey, A. (2009). Interactional feedback in learner-learner interactions in a task-based EFL classroom. *International Review of Applied Linguistics in Language Teaching, 47*(3/4), 267-301. https://doi.org/10.1515/iral.2009.012

García Mayo, M. D. P., & Lazaro Ibarrola, A. (2015). Do children negotiate for meaning in task-based interaction? Evidence from CLIL and EFL settings. *System, 54*, 40-54. https://doi.org/10.1016/j.system.2014.12.001

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Routledge.

Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For the Learning of Mathematics, 16*(2), 36-45.

Ghanbarzadeh, S., & Gholami, J. (2014). The effects of task complexity on the complexity and accuracy of foreign language learners' essays. *Modern Journal of Language Teaching Methods, 4*(2), 25 - 38.

Gilabert, R. (2007a). The simultaneous manipulation of task complexity along planning time and +/−Here-and-Now: Effects on oral production. In M. D. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44–68). Multilingual Matters. https://doi.org/10.21832/9781853599286-006

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

Gurzynski-Weiss, L., & Baralt, M. (2014). Exploring learner perception and use of task-based interactional feedback in FTF and CMC modes. *Studies in Second Language Acquisition, 36*(1), 1-37. https://doi.org/10.1017/s0272263113000363

Gurzynski-Weiss, L., & Baralt, M. (2015). Does type of modified output correspond to learner noticing of feedback? A closer look in face-to-face and computer-mediated task-based interaction. *Applied Psycholinguistics, 36*(6), 1393-1420. https://doi.org/10.1017/S0142716414000320

Gurzynski-Weiss, L., Long, A. Y., & Solon, M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis?. *Studies in Second Language Acquisition, 39*(2), 213-224. https://doi.org/10.1017/S0272263117000080

Halliday, M. A. K. (1975). Aspects of sociolinguistics. In E. Jacobsen (Ed.), *Interactions between linguistics and mathematics education* (pp. 64–73). UNESCO.

Hardy, I. M., & Moore, J. L. (2004). Foreign language students' conversational negotiations in different task environments. *Applied Linguistics, 25*(3), 340-370. https://doi.org/10.1093/applin/25.3.340

Hsu, H. C. (2020). The impact of task complexity on patterns of interaction during web-based asynchronous collaborative writing tasks. *System, 93*, 1-13. https://doi.org/10.1016/j.system.2020.102328

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development, 35*(1), iii-67. https://doi.org/10.2307/1165818

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*(2), 172-176. https://doi.org/10.1016/j.learninstruc.2009.02.013

Ishikawa, T. (2011). Examining the influence of intentional reasoning demands on learner perception of task difficulty and L2 monologic speech. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 307–329). John Benjamins.

Iwashita, N. (2001). The effect of learner proficiency on interactional moves and modified output in nonnative–nonnative interaction in Japanese as a foreign language. *System, 29*(2), 267-287. https://doi.org/10.1016/S0346-251X(01)00015-X

Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition, 25*(1), 1-36. https://doi.org/10.1017/s0272263103000019

Jarrett, D. (1999). *The inclusive classroom: Teaching mathematics and science to English-language learners: It's just good teaching*. North West Regional Laboratory.

Just, M. A., & Carpenter, P. A. (1976). The role of eye-fixation research in cognitive psychology. *Behavior Research Methods & Instrumentation, 8*(2), 139-143. https://doi.org/10.3758/BF03201761

Kaivanpanah, S., & Miri, M. (2017). The effects of task type on the quality of resolving language-related episodes and vocabulary learning. *TESOL Journal, 8*(4), 920-942. https://doi.org/10.1002/tesj.311

Kelm, O. R. (1992). The use of synchronous computer networks in second language instruction: A preliminary report. *Foreign Language Annals, 25*, 441–454. https://doi.org/10.1111/j.1944-9720.1992.tb01127.x

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech Language and Pathology, 22*, 397–408. https://doi.org/10.1044/1058-0360(2013/12-0083)

Kern, R. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and quality of language production. *Modern Language Journal, 79*, 457–476. https://doi.org/10.1111/j.1540-4781.1995.tb05445.x

Kim, Y. (2009). The effects of task complexity on learner–learner interaction. *System, 37*(2), 254-268. https://doi.org/10.1016/j.system.2009.02.003

Kim, Y. (2012). Task complexity, learning opportunities, and Korean EFL learners' question development. *Studies in Second Language Acquisition, 34*, 627 – 658. https://doi.org/10.1017/s0272263112000368

Kim, Y., & Payant, C. (2017). Impacts of task complexity on the development of L2 oral performance over time. *International Review of Applied Linguistics in Language Teaching, 55*(2), 197-220. https://doi.org/10.1515/iral-2017-0066

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition, 37*(3), 549-581. https://doi.org/10.1017/S0272263114000618

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal, 99*(4), 656-677. https://doi.org/10.1111/modl.12273

Kim, Y., & Tracy-Ventura, N. (2011). Task complexity, language anxiety and the development of past tense. In P. Robinson (Ed.), *Second language task complexity: Researching* *the Cognition Hypothesis of language learning and performance* (pp. 287 – 306). John Benjamins. https://doi.org/10.1075/tblt.2.18ch11

Kitade, K. (2000). L2 learners' discourse and SLA theories in CMC: Collaborative interaction in Internet chat. *Computer Assisted Language Learning, 13*(2), 143-166. https://doi.org/10.1076/0958-8221(200004)13:2;1-D;FT143

Knight, L. N., & Hargis, C. H. (1977). Math language ability: Its relationship to reading in math. *Language Arts, 54*(4), 423-428.

Kormos, J. (1999). Monitoring and self-repair in L2. *Language learning, 49*(2), 303-342. https://doi.org/10.1111/0023-8333.00090

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

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 267-285). John Benjamins. https://doi.org/10.1075/tblt.2.17ch10

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

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon.

Krashen, S. (1985). The Input Hypothesis: Issues and implications. Longman.

Kung, W., & Eslami, Z. R. (2015). Learners of different language proficiency levels and incidental focus on form in synchronous text-based discussion. *International Journal* *of Computer-Assisted Language Learning and Teaching (IJCALLT), 5*(3), 42-59. https://doi.org/10.4018/IJCALLT.2015070103

Kung, W. T., & Eslami, Z. R. (2018). Focus-on-form and L2 learning in synchronous computer-mediated communication: Language proficiency and dyadic types. In B. Zou & M. Thomas (Eds.), *In handbook of research on integrating technology into contemporary language learning and teaching* (pp. 118-139). IGI Global. http://doi:10.4018/978-1-5225-5140-9.ch006

Lai, C., & Zhao, Y. (2006). Noticing and text-based chat. *Language Learning & Technology, 10*(3), 102-120.

Leaver, B. L., & Stryker, S. B. (1989). Content-based instruction for foreign language classrooms. *Foreign Language Annals, 22*(3), 269-275. https://doi.org/10.1111/j.1944-9720.1989.tb02746.x

Lee. J. (2000). *Tasks and communicating in language classrooms.* McGraw-Hill.

Lee, J. (2019). Task complexity, cognitive load, and L1 speech. *Applied Linguistics, 40*(3), 506-539. https://doi.org/10.1093/applin/amy011

Leow, R. P. (2012). Attention, awareness, and noticing in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Wiley-Blackwell.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104. https://doi.org/10.1016/0010-0277(83)90026-4

Li, Y., & Chen, Z. (2019). The effectiveness of content-based language teaching and task-based language teaching in teaching reading to learners of military English. *English Language, Literature & Culture, 4*(2), 39-43. https://doi.org/10.11648/j.ellc.20190402.12

Liu, H. (2019). An action research into task-based CLIL applied to education majors: From Chinese students' perspective. *English Language Teaching, 12*(3), 94-107.

Llinares, A., & Dalton-Puffer, C. (2015). The role of different tasks in CLIL students' use of evaluative language. *System, 54*, 69-79. https://doi.org/10.1016/j.system.2015.05.001

Loewen, S. (2005). Incidental focus on form and second language learning. *Studies in Second Language Acquisition, 27*(3), 361-386.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126-141. https://doi.org/10.1093/applin/4.2.126

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based Language teaching. In K. Hyltenstam & M. Pienemann (Eds), *Modelling and assessing second language acquisition* (77-99). Multilingual Matters.

Long, M. H. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115-141). Newbury House.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press. https://doi.org/10.1016/B978-012589042-7/50015-3

Long, M. H. (2005). *Second language needs analysis*. Cambridge University.

Long, M. H. (2007). *Problems in SLA*. Lawrence Erlbaum Associates.

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

Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly, 26*(1), 27-56. https://doi.org/10.2307/3587368

Lopes, A. (2020). Linking content and language-integrated learning (CLIL) and task-based language teaching (TBLT) in an effective way: A methodological proposal. *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile, 6*, 5-22. https://doi.org/10.7764/onomazein.ne6.01

Lyster, R. (2015). Using form-focused tasks to integrate language across the immersion curriculum. *System, 54*, 4-13. https://doi.org/10.1016/j.system.2014.09.022

Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition, 35*(1), 167-184. https://doi.org/10.1017/s027226311200071x

Lyster, R. (2017). *Content-based language teaching*. Routledge.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*(4), 557-587. https://doi.org/10.1017/s0272263199004027

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback?. *Studies in Second Language Acquisition, 22*(4), 471-497. https://doi.org/10.1017/S0272263100004010

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS–NNS and NNS–NNS adult and child dyads. *Language learning, 53*(1), 35-66. https://doi.org/10.1111/1467-9922.00210

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. *British Studies in Applied Linguistics, 12*, 58-71.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85-104. https://doi.org/10.1191/0265532202lt221oa

Marsh, D. (2008). Language awareness and CLIL. *Encyclopedia of Language and Education, 6*, 233-246. https://doi.org/10.1007/978-0-387-30424-3_152

Martin, J. R. and White, P. R. R. (2005). *The Language of Evaluation: Appraisal in English*. Palgrave/Macmillan.

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD).* (Publication No. 3199485) [Doctoral dissertation, The University of Memphis]. ProQuest Dissertations and Theses Global.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381-392. https://doi.org/10.3758/brm.42.2.381

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language Learning, 56*(4), 693-720. https://doi.org/10.1111/j.1467-9922.2006.00393.x

Meyer, O. (2010). Towards quality-CLIL: Successful planning and teaching strategies. *PULSO: Revista de Educación, 33*, 11-29.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching, 45*(3), 241-259. https://doi.org/10.1515/iral.2007.011

Monaghan, F. (1999). Judging a word by the company it keeps: The use of concordancing software to explore aspects of the mathematics register. *Language and Education, 13*(1), 59–70. https://doi.org/10.1080/09500789908666759

Moore, D. S., McCabe, G. P., & Craig, B. A. (2012). *Exploring the Practice of Statistics*. Macmillan Higher Education.

Moradi, A., & Farvardin, M. T. (2020). Negotiation of meaning by mixed-proficiency dyads in face-to-face and synchronous computer-mediated communication. *TESOL Journal, 11*(1), 1-17. https://doi.org/10.1002/tesj.446

NCELA. (2019, September). *English learners and high school mathematics*. https://ncela.ed.gov/files/fast_facts/FactSheet_Del4.4_HighSchoolMath_112619_508.pdf

Norris, J. M. (2010, September). *Understanding instructed SLA: Constructs, contexts, and consequences*. In Plenary address delivered at the annual conference of the European Second Language Association (EUROSLA), Reggio Emilia, Italy.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds), *The handbook of second language acquisition* (pp. 716-761). Blackwell. https://doi.org/10.1002/9780470756492.ch21

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

Nuevo, A. (2006). *Task complexity and interaction: L2 learning opportunities and interaction*. (Publication No. 3247335) [Doctoral dissertation, Georgetown University, Washington, DC]. ProQuest Dissertations and Theses Global.

Nuevo, A. M., Adams, R., & Ross-Feldman, L. (2011). Task complexity, modified output, and L2 development in learner–learner interaction. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 175-202). John Benjamins. https://doi.org/10.1075/tblt.2.13ch7

Nunan, D. (2004). *Task-based language teaching.* Cambridge University Press. https://doi.org/10.1017/CBO9780511667336

Ortega, L. (2015). Researching CLIL and TBLT interfaces. *System, 54*, 103-109. https://doi.org/10.1016/j.system.2015.09.002

Parlak, Ö., & Ziegler, N. (2017). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition, 39*(2), 257-285. https://doi.org/10.1017/S0272263116000310

Payant, C., & Kim, Y. (2017). Impact of task modality on collaborative dialogue among plurilingual learners: A classroom-based study. *International Journal of Bilingual Education and Bilingualism, 22*(5), 614-627. https://doi.org/10.1080/13670050.2017.1292999

Payne, S., & Ross, B. (2005). Synchronous CMC, working memory, and L2 oral proficiency development. *Language Learning & Technology, 9*(3), 35-54.

Payne, J. S., & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal, 20*(1), 7-32. https://doi.org/10.1558/cj.v20i1.7-32

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kern (Eds.), *Network-based language* *teaching: Concepts and practice* (pp. 59– 86). Cambridge University Press. https://doi.org/10.1017/cbo9781139524735.006

Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, 3rd ed., pp. 703–732). Wiley.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language research and instruction. In S. M. Gass &. G. Crookes (Eds.), *Task-based learning in a second language* (pp. 9-34). Clevedon, Avon: Multilingual Matters.

Pienemann, M., & Johnston, M. (1987). Factors influencing the development of language proficiency. In Nunan, D. (Ed.), *Applying second language acquisition research* (pp. 45–141). National Curriculum Resource Centre.

Prabhu, N. S. (1987). *Second Language Pedagogy*. Oxford University Press.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal, 95*, 162-181. https://doi.org/10.1111/j.1540-4781.2011.01241.x

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*(4), 703-737. https://doi.org/10.1017/S0272263115000339

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*(3), 615-650. https://doi.org/10.1111/lang.12061

Richards, J., Platt, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. Longman.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning, 45*(1), 99-140. https://doi.org/10.1111/j.1467-1770.1995.tb00964.x

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

Robinson, P. (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 University Press. https://doi.org/10.1017/cbo9781139524780.012

Robinson, P. (2003). The Cognition Hypothesis, task design and adult task-based language learning. *Second Language Studies, 21*(2), 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 in Language Teaching, 43*(1), 1-32. https://doi.org/10.1515/iral.2005.43.1.1

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

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

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

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 3-38). John Benjamins. https://doi.org/10.1075/tblt.2.05ch1

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language instruction. *International Review of Applied Linguistics, 45*(3), 161-176. http://dx.doi.org/10.1515/IRAL.2007.007

Robinson, P. Gilabert, R. (2012). Task-based learning: Cognitive underpinnings. In C. Chapelle (Ed.), *Encyclopedia of applied linguistics* (pp. 1-6), Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1143.pub2

Rouhshad, A., Wigglesworth, G., & Storch, N. (2016). The nature of negotiations in face-to-face versus computer-mediated communication in pair interactions. *Language Teaching Research, 20*(4), 514-534. https://doi.org/10.1177/1362168815584455

Sachs, R., & Suh, B. (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 324– 338). Oxford University Press.

Sasayama, S. (2015). *Validating the assumed relationship between task design, cognitive complexity and second language task performance*. (Publication No. 3742197) [Doctoral dissertation, Georgetown University]. ProQuest Dissertations and Theses Global.

Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal, 100*(1), 231-254. https://doi.org/10.1111/modl.12313

Sasayama, S., & Izumi, S. (2012). Effects of task complexity and pre-task planning on Japanese EFL learners' oral production. In A. Shehadh & C. Coombe (Eds), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 23–42). John Benjamins. https://doi.org/10.1075/tblt.4.05sas

Sasayama, S., Malicka, A., & Norris, J. (2015). *Primary challenges in cognitive task complexity research: Results of a comprehensive research synthesis*. In 6th Biennial International Conference on Task-Based Language Teaching (TBLT), Leuven, Belgium.

Sato, M., & Loewen, S. (2019). Towards evidence-based second language pedagogy. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy: A collection of instructed second language acquisition studies* (pp. 1-24). Routledge.

Sato, R., & Takatsuka, S. (2016). The occurrence and the success rate of self-initiated self-repair. *TESL-EJ, 20*(1), 1-15.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning and Technology, 13*(1), 96–120.

Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. *Interlanguage Pragmatics, 21*(42), 1-31.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning (Technical Report #9)* (pp. 1-64). University of Hawaii, Second Language Teaching and Curriculum Center.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. K. Bhatt, & I. Walker (Eds.). *Perspectives on individual characteristics and foreign language education* (pp. 27-50). https://doi.org/10.1515/9781614510932.27

Segalowttz, N. (2007). Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly, 41*, 181-186. https://doi.org/10.1002/j.1545-7249.2007.tb00047.x

Shehadeh, A. (2001). Self-and other-initiated modified output during task-based interaction. *TESOL Quarterly, 35*(3), 433-457. https://doi.org/10.2307/3588030

Shehadeh, A. (2018). Foreword: New frontiers in task-based language teaching. In J. Ahmadian, & M. P. García Mayo (Eds.), *Recent perspectives on task-based language learning and teaching* (pp. vii–xxi). Walter de Gruyter. https://doi.org/10.1515/9781501503399-015

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

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

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1-14.

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

Skehan, P. & Foster, P. (2008). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. In Van Daele, S., Housen, A., Kuiken, F., Pierrard, M. & Vedder, I. (Eds.). *Complexity, accuracy, and fluence in second language use, learning, & teaching* (pp. 207-226). Contactforum.

Smith, B. (2003). The use of communication strategies in computer-mediated communication. *System, 31*(1), 29-53. https://doi.org/10.1016/S0346-251X(02)00072-6

Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition, 26*(3), 365-398. https://doi.org//10.1017/S027226310426301X

Smith, B. (2005). The relationship between negotiated interaction, learner uptake, and lexical acquisition in task-based computer-mediated communication. *TESOL Quarterly, 39*(1), 33-58. https://doi.org/10.2307/3588451

Smith, B. (2009). The relationship between scrolling, negotiation, and self-initiated self-repair in an SCMC environment. *CALICO Journal, 26*(2), 231-245. https://doi.org/10.1558/cj.v26i2.231-245

Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology, 16*(3), 53-81.

Smith, M., & Stein, M. K. (1998). Reflections on practice: Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School, 3*(5), 344–350. https://doi.org/10.5951/MTMS.3.5.0344

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *The Journal of Experimental Education*, *61*(4), 334-349. https://doi.org/10.1080/00220973.1993.10806594

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition, 39*(2), 347-380. https://doi.org/10.1017/s0272263116000425

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning, 52*(1), 119-158. https://doi.org/10.1111/1467-9922.00179

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–253), Newbury House.

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

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

Swain, M. & S. Lapkin (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal, 82*(3), 320–337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Swain, M. & S. Lapkin (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99-118). Longman.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press.

Sykes, J. M. (2005). Synchronous CMC and pragmatic development: Effects of oral and written chat. *CALICO Journal, 22*(3), 399-431. https://doi.org/10.1558/cj.v22i3.399-431

Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System, 37*(3), 482-495. https://doi.org/10.1016/j.system.2009.02.013

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

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

Toyoda, E., & Harrison, R. (2002). Categorization of text chat communication between learners and native speakers of Japanese. *Language Learning & Technology, 6*(1), 82-99.

Usó-Juan, E. S. T. H. E. R. (2006). The compensatory nature of discipline-related knowledge and English-language proficiency in reading English for academic purposes. *The Modern Language Journal, 90*(2), 210-227.

Van den Branden, K. (2006). *Task-based language teaching in practice.* Cambridge University Press. https://doi.org/10.1017/CBO9780511667282

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

Wang, Y., Borst, S., Feng, J. L., & Chang, R. (2015). It does matter with whom you chat: Chinese learners' perspective on NS vs. NNS chat partners. *Studies in Chinese Learning and Teaching, 1*(1), 18-39.

Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal, 13*, 7– 26.

Whimbey, A., Lochhead, J., & Narode, R. (2013). *Problem solving and comprehension* (7th ed.). Routledge.

Wickens, C. D. (1989). Attention and skilled performance. In D. H. Holding (Ed.), *Human skills* (2nd ed., pp. 71-105). John Wiley.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). HarperCollins.

Wickens, C. D. (2007). Attention to the second language. *International Review of Applied Linguistics in Language Teaching, 45*(3), 177-191.

Widdowson, H.G. (1998). Skills, abilities, and contexts of reality. *Annual Review of Applied Linguistics: Foundations of Second Language Teaching, 18*, 323–35. https://doi.org/10.1017/s0267190500003615

Willis, J. (1996). *A framework for task-based learning*. Longman.

Xing, J. (2015). The effects of increasing task complexity on EFL learners' writing performance. *Studies in Literature and Language, 11*(4), 34-39.

Yanguas, I. (2012). Task-based oral computer-mediated communication and L2 vocabulary acquisition. *CALICO Journal, 29*(3), 507-531.

Yilmaz, Y. (2011). Task effects on focus on form in synchronous computer-mediated communication. *The Modern Language Journal, 95*(1), 115-132. https://doi.org/10.1111/j.1540-4781.2010.01143.x

Yilmaz, Y., & Granena, G. (2010). The effects of task type in synchronous computer-mediated communication. *ReCALL: The Journal of EUROCALL, 22*(1), 20-38. https://doi.org/10.1017/S0958344009990176

York, J. (2019). *Language learning in complex virtual worlds: Effects of modality and task complexity on oral performance between virtual world and face-to-face tasks.* (Publication No. 27745902) [Doctoral dissertation, University of Leicester]. ProQuest Dissertations and Theses Global.

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), 1-27. https://doi.org/10.1093/applin/24.1.1

Yuksel, D., & Inan, B. (2014). The effects of communication mode on negotiation of meaning and its noticing. *ReCALL: The Journal of EUROCALL, 26*(3), 333-354. https://doi.org/10.1017/s0958344014000147

Zabihi, R. (2020). The effects of task type on the resolution of grammatical cognitive conflict episodes and grammar learning. *The Language Learning Journal*, 1-13. https://doi.org/10.1080/09571736.2020.1795913

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction. *Studies in Second Language Acquisition, 38*(3), 553-586. https://doi.org/10.1017/S027226311500025X

APPENDIX A

BACKGROUND QUESTIONNAIRE (FOR NNS)

Name: ____________ Age: ____________ Gender: Male/Female

Major: ___________ First Language(s): ____________

1) How long have you been studying English? ___________________

2) Where (which country) have you studied English? _________________________

3) Have you ever visited an English-speaking country before (including the US)? If yes, which country and how long?

____

4) Are you taking any ESL courses right now? If yes, how many hours per week?

____

Technology Usage

5) Have you ever used the following technologies in English? If yes, how often?

Social networking	Never, barely, sometimes, often, frequently
Text messaging	Never, barely, sometimes, often, frequently
Online chatting (text-based)	Never, barely, sometimes, often, frequently
Online video chatting	Never, barely, sometimes, often, frequently
Listening to music	Never, barely, sometimes, often, frequently
Any other? ___________	Never, barely, sometimes, often, frequently

6) Have you ever used computers or other technology to learn English? If yes, please provide one or two examples (e.g., Google Translate, Grammarly, Thesaurus, YouTube, Spotify).

____

APPENDIX B

EXPERIMENTAL TASKS

A. The Simple Task

Megan and Sophia love flowers so they decide to plant some flowers in the spring. They have a total of 47 flowers. Megan can plant 6 flowers per hour and Sophia can plant 8 per hour. However, Sophia had a meeting at the time they were supposed to start, so Megan worked for 2 hours before Sophia came back. How long will it take them to plant the rest of the flowers?

B. The Complex Task

It takes Jack 5 hours to paint a room of 420 sq ft. It takes Alex 7 hours to paint the same room. They plan to paint this room together. However, Alex had an important meeting at the time they were supposed to start, so Jack worked for 2 hours before Alex joined him to paint. After painting together for one hour, Jack received a phone call and went out for one hour, during which Alex was painting by herself. How long will it take them to paint the rest of the room working together?

APPENDIX C

TASK DIFFICULTY AND TIME JUDGMENT QUESTIONNAIRE

(FOR INTERACTION)

Name: _____________ Task version: __________ (A or B)

I. How much time do you think you spent on this task? ___________(mins)

II. Please circle a number for each statement to specify your level of agreement or disagreement.

Strongly agree	Agree	Slightly agree	Slightly disagree	Disagree	Strongly disagree
6	5	4	3	2	1

1. Overall, this task was difficult.

6 5 4 3 2 1

2. I felt like I didn't have enough time to think before I had to respond.

6 5 4 3 2 1

3. I felt confident in my ability to discuss mathematics in English.

6 5 4 3 2 1

4. The task I had to do with my partner wasn't hard.

6 5 4 3 2 1

5. I was relaxed and comfortable completing the task.

6 5 4 3 2 1

6. I felt rushed during the task.

6 5 4 3 2 1

7. It was tough to communicate the main points of problem in English.

6 5 4 3 2 1

Please circle a number for each statement to specify your level of agreement or disagreement.

Strongly agree	Agree	Slightly agree	Slightly disagree	Disagree	Strongly disagree
6	5	4	3	2	1

8. This task was stressful for me.

6 5 4 3 2 1

9. I enjoyed communicating with my partner during this task.

6 5 4 3 2 1

10. What my partner asked me to do in this task wasn't so difficult.

6 5 4 3 2 1

11. Sometimes I struggled during this task.

6 5 4 3 2 1

12. This task didn't make me feel anxious.

6 5 4 3 2 1

13. It was easy for me and my partner to find the solution of the problem.

6 5 4 3 2 1

14. It did not bother me when I did not understand everything my partner was saying.

6 5 4 3 2 1

15. I felt tense having to communicate with my partner.

6 5 4 3 2 1

APPENDIX D

TASK DIFFICULTY AND TIME JUDGMENT QUESTIONNAIRE

(FOR NARRATION)

Name: _____________ Task version: ___________ (A or B)

I. How much time do you think you spent on this task? ____________(mins)

II. How difficult you think this narrative task was?

Very difficult	Difficult	Slight difficult	Slightly easy	Easy	Very easy
6	5	4	3	2	1