University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Complexity, Accuracy, and Fluency (CAF) Development in L2 Writing: the Effects of

Proficiency Level, Learning Environment, Text Type, and Time among Saudi EFL Learners

Talal Musaed Alghizzi

A Dissertation Submitted in Fulfillment of the Requirements for the Degree of Doctor of

Philosophy in Applied Linguistics in the College of Arts, Celtic Studies and Social Sciences

at National University of Ireland, Cork

January, 2017

Supervisor: Dr. Martin Howard

Table of Contents

Declaration

I hereby declare that the thesis I am submitting is my own work and has not been submitted for another degree, either at UCC or elsewhere. All external references and sources are clearly acknowledged and identified within the contents.

Signed: _____

Abstract

The purpose of this longitudinal exploratory research is to investigate the influence of four factors: proficiency levels, text types, times, and learning environments on the writing complexity, accuracy, and fluency of Saudi students majoring in the English language. Specifically, the study seeks to determine how and when the CAF constructs and sub-constructs of low- and high-proficiency Saudi EFL undergraduates in three learning contexts: traditional learning context (TLC), blended learning context (BLC), and online learning contexts (OLC), are affected longitudinally across two writing tasks (classification and argumentative) that differed in their level of complexity. Also, it intends to specify when and which of the three learning contexts: TLC, BLC, and OLC, will lead to the most/least increases or decreases in the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates across the two composition tasks. To answer such questions, 75 Saudi EFL university students were recruited from the pool of two proficiency levels (low and high). Six groups of equal number of students were generated from dividing randomly the 45 Low-proficiency participants and the 30 high-proficiency participants. Each of these groups was exposed to one of the previously mentioned learning contexts and undertook three tests: pre-test, mid-term test, and post-test. The 450 students' writings were analyzed according to 45 measures of CAF constructs and sub-constructs and by using two statistical tests: t-test and ANOVA. For the first question, the t-test results showed that the similarities and differences of effect on CAF constructs between the two writing tasks were observed to be group-specific as they were based on the proficiency levels, learning contexts, and timescales (i.e., short term and long term). In other words, depending on whether a construct in the two text types was influenced similarly or differently, such influence did not generally occur in a systematic way and across the same number and types of metrics for the same group, or even across the groups of the same or different proficiency levels in the short term and long term. The findings only lent partial support to Skehan and Foster's Limited Attentional Capacity Model and Robinson's Multiple Attentional Resources Model since some constructs increased (e.g., accuracy, lexical variation, and syntactic complexity) or decreased (e.g., lexical density, lexical sophistication, lexical variation, syntactic complexity, and fluency). There were many other cases which were beyond the predictions of the aforementioned researchers and their explanations on how the students' attention is deployed while performing the complex task(s). For instance, altering task complexity led some constructs to remain unaffected (e.g., syntactic complexity, lexical density, lexical sophistication, lexical variation, accuracy, and fluency), equal increases and decreases or only

increases (e.g., fluency), increases more than decreases (e.g., lexical variation), less increases (e.g., accuracy), or less/more decreases (e.g., syntactic complexity and lexical sophistication). In terms of the second question, the ANOVA test results indicated mixed findings because each of the three learning environments resulted in benefits in some ways. In the two proficiency levels, the TLC, BLC, and OLC had the same level of success/unsucess in enhancing all the measures of some CAF constructs in both writing tasks in the short term and long term. Nevertheless, in the other CAF constructs, there was no uniform linear development or deterioration of all measures across the six groups. In each of these constructs, the differences between these groups emerged from one or more measures, but not from all measures. Each of these learning contexts stood alone in being the most or least successful in increasing some constructs. Nonetheless, this was dependent on the participants' proficiency levels, text types, and timescales. This study provides several pedagogical implications and recommendations for academic research, EFL writing instructors at pre-university and university levels, and task-based investigators.

Keywords: writing, proficiency levels, teaching methods, learning environments, text types

Dedication

To my family

Acknowledgements

Reflecting on my long journey of choosing a topic, planning, conducting research, and actually writing up a dissertation, what stands out the most is how fortunate I have been to have such supportive mentor, family, and colleagues throughout my career as a doctoral student at University College Cork University. First and foremost, praise be to Allah the Almighty, the most Gracious and most Merciful, for paving the way for me to finish my dissertation. Second, since the production of this thesis has been lengthy and required an arduous effort, I would like to seize the opportunity to extend my gratitude to some people. My thanks must go to my supervisor, Dr. Martin Howard, for his steadfast and unlimited support, encouragement, and perseverance over almost four years. Without his help, the fact that his erudition has been always available to me and that he has allowed me to work in my own way, this dissertation would not have seen the light of day. I would also like to express my appreciation to Dr. Muhammad Abdel Latif for his abundant and invaluable assistance in teaching two of the research six groups. Moreover, I would like to convey my special thanks and my indebtedness to my friends, here and overseas, and to the participants of my research. Finally, so much gratitude is due to my family: my wife, Mashael, and my four children, Yazan, Yasser, Elias, and Taraf. All of them have always been there for me through hardships, have changed my life for the better, and have helped me bring this dissertation to a successful conclusion.

List of Tables

List of Figures

Chapter One

Introduction

1.0 Research Background and Rationale: English in a Saudi Educational Context

English is a widely used language as a means of communication between people from different ethnicities and language backgrounds. Therefore, it is a lingua franca (Crystal, 2005; Jenkins, 2006; Seidlhofer, 2001,2005) which represents a threat to the multilingualism for other languages (cf. House, 2003). The dominance of such language is attributed to many factors. For example, Al-Jarf (2008) noted that the language is important because it is:

> The language of 85% of international organizations and the main language of technology, business, finance and tourism. Most research, references, technical terms, international conferences, electronic databases are in English. World famous newspapers, T.V. stations, movies, airlines, multi-national companies, and 90% of the material published on the internet use English as a primary language. (p. 197)

Because of the rapidly-increasing dominance of English, some Arab countries manifested interest in teaching the language by implementing it as a compulsory subject in the pre-university (i.e., primary, intermediate, and secondary) school curricula, and by opening colleges/departments where students can specialize in the language. These steps were needed as the importance of English has been underscored in much research (e.g., Al-Tamimi & Shuib, 2009; Al Noursi, 2013; Bani-Khaled, 2014; Tahaineh & Daana, 2013; Tanni, 2015). These studies concluded that Arab EFL learners have developed the motivation to acquire English as a means of achieving instrumental goals, advancing their careers, reading technical materials, and translation. For example, Zughoul (2003, p. 2) stated that: "Despite the hegemonic and imperialistic nature of English, it is still badly needed in the Arab world for the purposes of communicating with the world, education, acquisition of technology and development at large." Tahaineh (2009) maintained that not only do better job opportunities, particularly in private companies, demand competence and knowledge in English, the fact that government officials are almost always bilingual in Arabic (First/Source language [L1/SL]) and English (Second/Target language [L2/TL]), but some people also may feel that the L1 is moved to a secondary status due to the frequent use of the L2. However, Syed (2003) contended that what painted a very unflattering picture of the education in the Arab world and gulf countries is that there are continuous challenges that have led to disastrous outcomes for learners in English. These challenges were the lack of or insufficient support system and of qualified teachers, outdated methodologies and curricula, students' gradual loss of motivation,

literacy, underachievement, and reliance on high-stake testing, memorization, and rote learning.

In the case of Saudi Arabia, all of the points mentioned are applicable. To clarify this, ever since the introduction of the English language there, the status of the language has fluctuated from one status to another.[1] Historically, it is unclear when exactly English was introduced in the education system in Saudi Arabia, but there was a mutual agreement of some researchers that the formal beginning was immediately or some years after the establishment of the General Directorate of Education, nowadays known as the Ministry of Education (MoE), in 1924, 1925, or 1928 (Al-Abed Al-Haq & Smadi, 1996; Al-Ahaydib, 1986; Alhajailan, 2015; Al-Seghayer, 2011, Faruk, 2014a). At that time, English was taught as a subject in the public primary schools (Al-Abed Al-Haq & Smadi, 1996). Unfortunately, after 13 years, the teaching of English at such a stage stopped for undeclared reasons (Alhajailan, 2015). Then, in the 1940s, the intermediate and secondary levels were established, and their English textbooks were similar to the ones in the Arab countries (Alhajailan, 2015). Alamri (2008) maintained that in the first application of English at the previous two levels, there were no single specific syllabi, but later, that is in 1953, the MoE imported Egyptian English textbooks. However, they were later deemed as unsuccessful and were changed due to many factors such as the differences between the Saudi students' customs and needs. In addition, Alamri (2008), Al-Seghayer (2011), Alhajailan (2015), and Mahboob and Elyas (2014) stated that the French language was also taught for some years at intermediate and secondary stages, but it was cancelled and English became the only registered subject in the education cycle. In 2004 —as stated by Alamri (2008)—there was a reintroduction of the English language into the curriculum by the MoE, which passed a law mandating that it be taught, starting at the sixth grade in primary schools. A year later, English was introduced, beginning in the 5th grade to primary school students, and, 6 years later (i.e., in 2011/2012), English began to be taught to students in the 4th grade. The decision by the ministry to make English an obligatory subject, commencing from grade 4 onward shows how important that language is deemed to be. As claimed by Liton (2013), and Rahman and Alhaisoni (2013), the status of English at these educational stages for both teachers and

---

[1] Many researchers (Al-Abed Al-Haq & Smadi, 1996; Al-Ahaydib, 1986; Alamri, 2008; Alhajailan, 2015; Alrashidi & Phan 2015; Al-Seghayer, 2011, Faruk, 2014a; Mahboob & Elyas, 2014) who addressed the topic of English in Saudi Arabia revealed discrepant points in terms of the language introduction date, its teachers, syllabi, textbooks, and hours of classes. Yet, the overview supplied is relatively an enough elaboration on how the status of the language changed gradually to become the only second language taught in the pre-university cycle. Cited studies are recommended for further reading.

students is no more than a compulsory subject to be learned for the purpose of passing examinations, and with almost no emphasis on acquisition efficiency. Yet, many studies (e.g., Al Asmari, 2013; Al-Jarf, 2008; Al Samadani & Ibnian, 2015; Faruk, 2014b) found that recently Saudi students started to have positive attitudes towards learning the language adequately for a variety of reasons. For instance, the investigation of Al-Jarf (2008) on the effect of English on Arabic showed that the majority of the 470 Saudi female participants majoring in medical fields and computer sciences adduced the superiority of the L2 in relation to the native language since it is the language of modernity, and a prerequisite of higher education and jobs. Also, most of these students indicated that the mastery of the language would contribute to better social status, self-esteem, and self-confidence. With such change of attitudes and perceptions, the MoE has had no choice but to continue its long-lasting scheme, which started since the early introduction of English, for the betterment of planning a purposeful curriculum, refining textbooks, and developing teachers and effective administration (Khan, 2011a, 2011b).

In other words, the first step by the MoE was to recruit foreign EFL instructors who according to Al-Awad (2002, as cited in "Ministry eyes ways," 2002) reached 1300 teachers. Al-Seghayer (2011) maintained that in 1970, the ministry designed a training program for secondary school graduates who wanted to become English teachers. Those graduates were required to study the language intensively for one year, passing a comprehensive exam, and finally, studying the language abroad for two years and one month in British universities to obtain a teaching certificate. Whoever successfully accomplishes such task would become an English instructor in intermediate schools (Al-Seghayer, 2011). Then, when universities such as Umm Al-Qura University (UQU), King Saud University (KSU), and Al-Imam Muhammad Ibn Saud Islamic University (IMSIU) were established in 1949, 1957, and 1974, respectively, their graduates were offered the possibility to go through the previous processes to become EFL instructors. Finally, the third major step by the MoE was to require universities to open English colleges/departments.

In these colleges/departments of English language, there are two options of curricula, either to study courses which belong to the language major fields such as literature, translation, theoretical and applied linguistics, or only to specialize in one of these fields and be exposed to various courses addressing its different branches. In all of these options, extensive teaching of the basic skills such as reading, writing, listening, and speaking is included in their curricula. The Department of English Language and Literature, in the College of Languages and Translation, at IMSIU, falls into the first option. According to this

college website (2016), there are ten objectives of the English Department. We present these below (our translation).

1) To provide a variety of educational and cultural services that meet the needs of students and their preparations.

2) To educate the student in terms of the diversity of linguistic system through studying variety of language subfields

3) To enable students to recognize the fields and subfields of the modern language, including the history of the English language, sounds, grammar, uses, styles, as well as literary works, theoretical and applied linguistics, and translation.

4) To encourage authoring, research and investigation in the areas of language, translation, literary and critical studies, and holding training courses, educational and scientific meetings.

5) To work on the preparation of scientific cadres specialized in the areas of English language teaching, translation, and literature by working on sharpening their translation skills, evolving their critical abilities, introducing them to different theoretical and literary styles, and promoting their educational and teaching potentialities.

6) To provide the student with the essences, concepts, theories, and recent trends of the field of English language.

7) To provide technical and practical consultations in the field of English language teaching and translation.

8) To work on strengthen the ties and links between the college and the community by offering academic programs, and specialized courses that contribute to raising the efficiency of members of the community in the field of English language, all of which are in line with the objectives of the development plans.

9) To make cadres acquire a good knowledge and command of English which will hopefully qualify them to enter working life efficiently and effectively, whether it is in the field of education, translation, or other developmental areas that require knowledge of English in public and private institutions.

10) To promote cooperation between the college and the centers/institutions of research and studies in similar local and international colleges.

However, graduates from the English colleges/departments who became EFL instructors have successfully fulfilled half of the MoE's aim to replace foreign teachers. The other half of the aim, namely improving pre-university students' English proficiency in general and their

writing skills in particular, is yet to be accomplished. The latter task seems to be far-fetched as it has been found that most of the Saudi EFL instructors are still weak, and ill-equipped to teach such skill (e.g., Al-Seghayer, 2015). For example, in the pre-university-level English classes, Alharbi (2015), Alresheed (2008) and Al-Seghayer (2014a, 2015) contended that the most commonly adopted teaching approaches are the *Grammar Translation Method (GTM)*, and the *Audio Lingual Method (ALM)*. Mart (2013) emphasized the benefit of the GTM in enabling the learners to understand how the language systems of the L1 and the L2 work, that is, by allowing them to notice the similarities and differences between the two languages. Al-Nofaie (2010) even reported general positive attitudes toward GTM by both Saudi teachers and students with a preference for limited usage of the mother tongue in certain situations and for specific reasons. In addition, Al-Hazmi (2006) maintained that what has had a negative effect on pre-university students' composition development is that the skill is taught with a focus on sentence-level structure and linguistic features. Alghizzi (2011, 2012) found that not only did most EFL instructors tend to use an unspecified approach, but they also required their students to memorize ready-made sample texts. These teachers taught learners how to recite and/or adjust samples according to the topic(s) they are required to write about (see Elyas & Picard, 2010; Shukri, 2014). The question to be asked is why do most EFL instructors in Saudi Arabia tend to use or focus on the aforementioned method and nothing else? The simple answer could be because they lack the adequate training to do so.

Al-Seghayer (2011) cited a study conducted by the MoE in 2004, which substantiated that the Saudi EFL teachers at intermediate and secondary school levels were "neither competent in English nor in the affair of teaching it" (p. 23). The study linked such weaknesses to the lack of proper training when these EFL instructors were undergraduates. It is true that other factors such as learners' psychology, experiences, perceptions towards learning the language, aspects of the curricula, pedagogy, administrative processes, language policy, methods, strategies, and assessments (Al-Seghayer, 2014a, Khan, 2011b), contributed to the weaknesses and unsatisfactory outcomes of the students, but these variables would have been resolved better if the teachers were proficient and trained to do so. Unfortunately, following the in depth investigation of the EFL instructors' pre-service and in-service preparation programs by Al-Hazami (2003), and Al-Seghayer (2014b), they called for the refinement, updating, and reform of such courses. The suggestions they made were that these programs should educate future/current EFL instructors on how to design lesson plans and activities, adopt teaching methodologies and technologies, use assessment and time management for each of language skills they teach accurately. Alhumaidi (2013) reported that

Dr. Sami Alshwairkh, the director of the English language teaching initiative in Tatweer Corporation for Educational Services (T4EDU), declared that the agency would launch a distance education project, aiming in the first stage at developing up to 2000 EFL teachers' language skills, training them on the different teaching methodologies, how to use technologies and the new pre-university levels' English textbooks once they become available to all levels in 2014/2015. In 2015, Alshwairkh announced that only 8700 EFL instructors of primary, intermediate, and high school educational levels, had successfully finished 450 training workshops held by 150 professional EFL trainers across the country on the new English textbooks (Alghamidi, 2015). He also stated that 1200 EFL instructors had attended 32 writing skill training workshops held by 64 trainers. The latter workshops were under the supervision of the Teachers College at Columbia University (Alghamidi, 2015). The issues here are that in order for such a project to reach all current EFL instructors, it will consume a lot of time and effort, let alone the fact that designing training workshops for writing are still not enough since it has always represented a challenge for both teachers and students. Consequently, the initiative to equip EFL teachers to teach composition effectively should start from the English colleges/departments themselves.

1.2 Statement of the Problem

No one doubts the importance and significance of teaching English writing skills as one of the fundamental aims of the colleges/departments of English in Saudi Arabia. The importance of such a skill arises from not only being one of the language's basic skills, but also because of the fact that high acquisition of it will most likely lead to achieving the above-mentioned objectives, and others. Thus, it is reasonable to expose EFL students, regardless of their fields in English, to intensive writing courses during their four-to-five-year bachelor degree(s). These courses can start from the first year (i.e., semester one), and last to the third year (i.e., semester six). The goal behind this, is to develop undergraduate EFL students' writing to the level of near mastery; in not violating the rules, norms, features, and conventions of the language (i.e., grammar, style, organization, spelling, word choice, punctuation, capitalization, text type norms), and to make them capable of teaching it to their future students.

Nevertheless, in spite of the keen interest and effort by these colleges/departments in developing writing skills, as well as other language skills of their students, it seems that Saudi students still generally lack the ability to compose effectively in English. Many studies concluded that the writing of these students is problematic, unsatisfactory, and lags behind

other skills such as speaking (Al Fadda, 2012). It is true that some extra efforts and effective interventions have been made, but they were limited, with temporary positive effects, and initiated personally by EFL writing instructors who are specialized or at least have profound expertise in the field of writing, and not by the faculties of English colleges/departments themselves. These ambitious instructors intended to develop their EFL students' composition by investigating, for example, the impact of applying and examining various types of tools, contexts, and supplementary materials such as mind mapping software, word processor, blogs, revision checklists and feedback, online courses, and cooperative learning in small groups, on writing ability (e.g., AbuSeileek, 2006, 2012; Al-Hazmi & Scholfield, 2007; Al-Jarf, 2002, 2004a, 2004b, 2006, 2007a, 2009a, 2013; Aljumah, 2012; Fageeh, 2011).

Regardless of how persistent writing weaknesses are, it is the belief of the present researcher that Saudi EFL undergraduate learners can improve their composing abilities and can rise to meet the English colleges/departments' expectations and objectives. Nonetheless, in order to do so, attention must be given to studies that examine and compare the effect of factors such as teaching and learning contexts, proficiency levels, and text types, on the development of CAF constructs (i.e., complexity, accuracy, and fluency [CAF]) in the written texts of under- and post-graduate students. The analyses of these studies may provide educators, L2 writing skills instructors, curricula designers, and faculties with relevant information on the most appropriate environment for teaching writing skills, the best methods to track its development and remediation. Should this be done, it can contribute to improving Saudi EFL undergraduate students' ability to a level of competence that is more likely to be satisfactory.

Finally, according to McMullen (2009), any EFL instructor who experienced teaching composition in Saudi Arabia may confirm that the skills "always represent…a unique challenge for Saudi EFL students" and to the EFL teachers themselves (p. 18). From such observation, there were three problems that inspired the current study. First is the fact that in Saudi Arabia, rarely have the faculties of English colleges/departments assigned the teaching of composition to well experienced specialists. The tendency is to allow professors or lecturers from other fields of interest such as literature, theoretical/applied linguistics, or translation to take such responsibility. The priority of such instructors is to teach the courses in regular traditional classrooms, and with no exposure to other contexts by which students' self-autonomy towards writing would be ignited and developed. Second, the evaluation of students' writing is predominantly undertaken by adopting one of two most common assessment methods: holistic or analytical. Charney (1984) stated that holistic rating is a set of

impressionistic quick qualitative procedures for ranking and categorizing a written text. The goal is not to perfect the production by identifying its weaknesses, or even correcting and editing it, but rather to assign a value to it based on a previously establish criteria. Hamp-Lyons (1995), Park (2008), Weigle (2007), and Wolcott (2004) similarly provided various points of criticisms toward this evaluation technique. Some of these points were that there is no consensus on the characteristics that constitute good writing, the lack of diagnostic power in identifying weaknesses and strengths, which is always demanded by L2 writing teachers, the fact that the assigned value of a text does not allow raters to distinguish between different aspects of composition like organization, vocabulary depth, and grammar mastery, and is given as if the text were a final product, with no credits given to the writing processes.

In the analytical assessment, it is most likely for university writing instructors to develop their own rubrics-based evaluation by developing several score levels, such as one to five, with a detailed description of what is expected at each level regarding some writing elements of their choice. Alternatively, these teachers may adopt some widely used rating scales such as '*ESL Composition Profile*,' which was developed by Jacobs, Zingraf, Warmuth, Hartfiel & Hughey (1981). The profile is a scoring procedure which divides writing into five major components each with designated points: content (30 points), organization (20 points), vocabulary (20 points), language use (25 points), and mechanics (five points). The total score can range from 34 minimum to 100 maximum. Also, each of these components' ratings is further broken down into numerical ranges which correspond to four levels of mastery: very poor, fair to poor, good to average, and excellent to very good (Jacobs et al., 1981). Haswell (2011) maintained that this rating scale "in its main features,…is no different than dozens of similar guides by which raters have decided, and continue to decide, the academic fate of thousands upon thousands of [L2] students" (p. 107). Therefore, he criticized it for having a limited number of component/traits, proficiency levels, and subtraits, and for disregarding other features of writing such as cleverness, logic, creativity, humor, suspense, tradition, and shock-appeal. Wiseman (2012) noted that the scale consumes both time and money, while Myford and Wolfe (2003, p. 395) adduced that it suffers from "halo effect," that is the fact that a rating on one scale will likely affect the rating on another scale. Becker (2011, p. 115) explained this by saying that "this rater effect reflects the tendency of raters to let one trait influence evaluation of another trait." However, whether the scale of assessing writing skills is holistic or analytical, there are still important issues as both of them lack the ability to specify the levels of development of students' writing ability and overlook indispensable and more general aspects (i.e., CAF constructs) of such skill. As a result, when

students perform different tasks, instructors will not be able to justify and understand why the former would prioritize some constructs such as grammar (i.e., complexity and accuracy) over others such as the sounding of the text (i.e., fluency) or vice versa, and how their attention towards these constructs is deployed during the production.

The third issue that stimulated this research is the fact that there is little or almost no information on tracing the writing development of Saudi EFL undergraduate students in terms of CAF constructs, and on how development of the latter constructs would be affected and/or differ depending on teaching and learning environments (traditional [TLC], blended [BLC], and online [OLC]), proficiency levels (low and high), text types (classification and argumentative), and timescales (short term and long term). The present study focusses on the last two problems and seeks to provide some information on such matters.

1.3 Research Questions

The primary agenda for this study is to examine, describe, and explore the differences between a traditional learning context, a blended learning context, and an online learning in the case of Saudi university students in English writing skills with regard to CAF, as well as the impact of four different factors: proficiency levels, learning contexts, text types, and timescales on such constructs. Specifically, the study addressed the following research questions:

1. How are the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates in the three learning environments: TLC, BLC, and OLC, affected longitudinally across two writing tasks (classification and argumentative) that differed in their level of complexity? And when?

2. Which of the three learning environments: TLC, BLC, and OLC, lead to the most/least increase or decrease in the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates across two writing tasks: classification and argumentative? And when?

1.4 Research Objectives

- To obtain information on when and how the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates in the traditional, blended, and online learning contexts will be influenced in the two writing tasks: classification and argumentative, that differ in the level of complexity.

- To obtain information on when and which of the traditional, blended, and online leaning contexts will lead to the most/least increase or decrease in the CAF

constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates across two writing tasks: classification and argumentative.

1.5 The Significance of the Research

The significance of this study derives from three reasons. First is the fact that it attempts to explore, describe, and capture longitudinally a new area in the writing skills field, that is, the role and effect of four factors: learning environments (i.e., TLC, BLC, and OLC), proficiency levels (i.e., low and high), text types (i.e., classification and argumentative), and timescales (i.e., short term and long term), on the writing development of Saudi EFL undergraduates. Second, it adopts, for the purpose of analysis, a large number of indices for complexity: syntactic and lexical, accuracy, and fluency, all of which underlie writing skill development. Third, it shows when, where, why, and for whom the increase/decrease of writing development aspects will occur. As such, this study provides information on the development of writing skills in a Saudi context.

Many studies in the Arab world in general and in Saudi Arabia in particular have addressed the writing skills from different angles. The topics of such research were: analysis of writing textbooks, academic features, and written mistakes/errors, incorporation of various text types, teaching approaches, and teaching and learning contexts, impact of L1/SL, planning conditions, and various correction types (i.e., teacher-correction, peer-correction, and self-correction) on L2 writing, and finally, the effect of and/or the relationship between other skills such as reading, listening, and speaking with writing. However, research which investigates the effect of four factors (i.e., proficiency levels, learning environments, text types, and timescales) on Saudi EFL undergraduate writing development is scarce. Therefore, the investigation of such a topic is urgently needed, due to the continuity of ineffective practices of writing skill instructors when assessing, tracing the development, and teaching EFL learners, especially to those majoring in English.

To elucidate this, composition is regarded as the last, and perhaps the most difficult, language skill acquired by native speakers of English language, and by EFL learners (Breland & Jones, 1982; Hamp-Lyons & Heasly, 2006). As mentioned before, in Saudi Arabia, the faculties of the colleges/departments of English language have always assigned the teaching of such skills to non-specialists. The issues emerged from the fact that most such instructors would embrace the traditional context of teaching (i.e., in regular on-campus classrooms) of such a difficult skill, and incorporate an ad hoc assessment method such as Jacob et al. (1981). Such a teaching environment is not necessarily suitable for all students, especially those with

different learning style preferences, while the adopted assessment method is limited in what it can tell about students' writing development levels. The assessment, as Khongput (2010) maintained, has also been prone to rater bias, subjectivity, perceptions of good writing, and cultural and professional background. What can be postulated is that most of these current students and prospective future EFL teachers will continuously have some deficits in their writing abilities and will lack the ability to teach the skills to their students.

Thus, this study is important because in order for EFL writing instructors and faculties of English colleges/departments in the Arab world in general and specifically in Saudi Arabia to help their EFL students to overcome their writing problems, and measure their development in the different writing tasks accurately, they need some information that allows them to choose the most appropriate teaching environment/contexts and assessment technique, and to understand how their learners deployed their attention when performing tasks. The research will be beneficial for not only L2 writing skill instructors, but also for task-based researchers, by extending the scope of investigation to reach a Saudi context, that has rarely been examined in relation to the above-mentioned factors; providing them with a picture of which of the CAF construct measures will effectively capture the significant increase/decrease, in which text types and learning contexts, and for which proficiency levels. Finally, since writing skills are essential to general academic success, it is hoped that the present results, along with their pedagogical implications, will contribute to more successful outcomes for Saudi EFL students and the refinement of the hypotheses proposed by language-task researchers who, majorly, overlooked the impact of learners' learning contexts on their prioritization and depririoritzation of CAF constructs and sub-constructs when performing tasks that are different in their level of complexity.

Chapter Two

Literature Review Part I: Writing Acquisition, Definitions, Teaching Environments, and Approaches

2.0 Introduction

This chapter presents existing literature on certain aspects of writing skills. The literature review includes two principal parts. The first part discusses composition acquisition/development studies, as well as the definition, teaching environments, and approaches of composition. The second part presents issues related to CAF, one of the assessment methods of composition, and its origin and research challenges (e.g., definitions and measurements, linguistic and psycholinguistic correlations and interconnections, and the factors affecting them).

2.1 Writing Skill Acquisition Studies

Most languages consist of four skills: listening, speaking, reading, and writing. The last is regarded to be the last step on the ladder of L1 and L2 acquisition, and it is extremely difficult to master for both native and nonnative speakers (Abedi, Latifi, & Moinzaden, 2010; Al-Khairy, 2013; Breland & Jones, 1982; Buckley, 1986; Fageeh, 2011; Gallavan, Bowles, & Young, 2007; Ghorbani, Gangeraj, & Alavi, 2013; Hammad, 2014; Hamp-Lyons & Heasly, 2006; Levy & Ransdell, 1995; Mourtaga, 2010; Osman & Abu Bakar, 2009; Shen, Tseng, Kuo, Su, & Chen, 2005; Tangpermpoon, 2008). For example, Javid and Umer (2014) contended that in the language learning natural order hypothesis, although composition skill is at the end, its importance is very high since classroom learners are obliged to complete written texts as a means of showing what they have learned. Both Hammad (2014) and Abedi et al. (2010) have emphasized that while the skill is difficult to master for L2 learners, it is also difficult for L1 learners (see also Breland & Jones, 1982). McNamara, Crossley, and McCarthy (2010) maintained that "writing well is a significant challenge for students and of critical importance for success in wide variety of situations and professions" (p. 58). Buckley (1986) stated that composition "is the most difficult and abstract form of verbal thinking" (p. 371).

Thus, it should be asked why "both professional and nonprofessional writers often lament that the process of writing is arduous and complex…[and think that] perhaps not coincidentally, writing is also one of the least understood language-production tasks" (Levy & Ransdell, 1995, p. 767). In other words, what are the reasons that writing is so hard to master?

For the last three decades, the field of writing has intrigued L1 and L2 researchers. The sole aim has been to determine the factors that could lead to or escalate the difficulty of acquiring and developing the skills. The accumulative studies of theoretical and applied L1 and L2 writing acquisition were categorized by specialists into different orientations. For example, McCarthy, Guo, and Cummins (2005) stated that the L1 and L2 composition research revolving around their processes could be divided into three kinds: studies that investigate the rhetorical and linguistic patterns between the processes of the L1 and the L2 compositions and how such patterns are similar and different in the two languages, studies that examine the role played by culture in differentiating between the writing in both languages, and lastly, studies that emphasize the role of instruction on the L1 and L2 writers. Barkaoui (2007) provided more general classifications, such as "text-focused, process-focused, and sociocultural" orientations (p. 35).

In the Arab world in general and in the Gulf countries in particular, some researchers divided L2 writing acquisition into different categories. For Ezza (2010), there are only two types of studies: "Cross-cultural…and instruction-oriented" research (p. 35). Alghizzi (2011) added a third category: "Area-specific studies" (p. 14). Both Abdel Latif (2011) and Alghizzi (2012) pointed out that there are various research clusters concerning the interference or transfer of the native language, discourse and error analysis, the assessment of writing, pedagogies and curricula, the evaluation of practices, weaknesses in L1 composition, other skills and writing, writing conventions and features, and teaching methodologies. It is true that factors such as writing and learning strategies and styles, technologies, teaching and learning contexts, feedback, and others (see Table 2 on page 66 for other variables) may not fall into any of the aforementioned clusters; yet, they are sufficient examples of how diversified the field of writing development is. Some of the studies exemplifying a variety of classifications and factors include AbuSeileek (2006), Ahmed (2012), Al-Ahdal, Alfallaj, Al-Awaied and Al-Hattami (2014), Al-Hammadi (2010), Al-Hazmi and Scholfield (2007), Al-Jamhoor (2001), Al-Jarf, (2002, 2004a, 2004b, 2009a, 2009b), Alkhawaldeh (2011), Alkubaidi (2014), Al-Mansour and Al-Shorman (2014), Almelhi (2014), Alnufaie and Grenfell (2012), Al-Quran (2010), Alsamadani (2010), Alshahrani and Storch (2014), Alsulayyi (2015), Bataineh (2005), Btoosh and Taweel (2011), Ezza (2010), Fageeh (2011), Glasgow and Fitze (2008), Javid and Umer (2014), Khan (2011a), Mahmoud (2011), McMullen (2009), Mourtaga (2010), Rajab (2013), Salebi (2004), Sawalmeh (2013), Sharma (2015), and Tahaineh (2010). The research papers found that these clusters and variables could contribute to learners' challenges in acquiring and developing the composition skill.

However, this discussion will focus on two factors—writing characteristics and teaching contexts and approaches—since these are the most relevant to the present research.

2.2 Definitions of Writing Skills

It has been argued that the definition of writing skill could be one of the reasons for students' struggles to master such a skill. The definition of composition, despite being meant to provide learners with a clear and precise picture of what the skill is, can actually exacerbate the difficulty of learning, acquiring, teaching, applying, and even assessing composition. In other words, when learners search for or teachers adopt an interpretation of the skill, they are really looking for something that, by its mere application, will lead to the production of written texts that meet the conventions of high-quality writings—a high level of rhetorical, syntactic, stylistic, and lexical language. However, according to Coulmas (2003), because of composition's long history, its importance, and the multiple meanings of English words, it is extremely challenging to define the construct clearly. He validated this assertion by providing six possible connotations of the word composition: (a) visible or tactile symbols that make up a system to record a language; (b) the use of such a system to record information; (c) the text that is recorded by employing such a system; (d) the specific format that such a text uses (block letters); (e) an author's creation; (f) the occupation of a professional or of an amateur. Cumming (2002) highlighted various ethical concerns that emerge "according to how the construct of writing is defined" (p. 73). White (1993) maintained that not only does a huge range of interpretations exist as a result of the sophisticated sequence of learning this skill, but when these interpretations (either in whole or in part) are used to analyze texts, they have to be specified. Otherwise, people may have "different conceptions of what is being measured" (p. 105).

Yi (2009) deduced that the range of proposed definitions for writing and the lack of consensus regarding a unified interpretation can be attributed to the fact that the different definitions can be based on teachers', educators', or researchers' philosophies and experiences in such skill; on the teaching approaches adopted; on the contexts (e.g., task, situation, organization, writer, setting, and materials); on the aim of the pedagogy; or on students' characteristics. In each of the above-mentioned cases, the interpretations not only varied but could also be criticized, thereby accelerating the difficulty in learning the skills.

First, with regard to defining writing in terms of the features of learners, Matsuda (2003) discussed how, for example, a "basic writer" (i.e., an amateur) had been interpreted in the literature, concluding that the epithet had been overgeneralized to include—and

sometimes to exclude—different types of learners without any reference to their educational, linguistic, or cultural backgrounds (p. 68). He criticized the reliance on definitions based on immigration status, language background, or the placement test scores of writing skill, suggesting the inclusion of "all students who are subject to the disciplinary and pedagogical practice of basic writing" (p. 84). On the other hand, Kellogg (2006) acknowledged the difficulty in defining expertise in composition (i.e., a professional writer), believing that it stemmed from "the task [being] ill structured…and the types of texts generated by professionals [being] so varied" (p. 389). In other words, the problem in interpreting 'professional writers' is linked to the lack of agreement and specification of what constitutes writing professionally, as well as the fact that its multiple aims are described with general terms, such as "cohesion" (p. 389). Kellogg also maintained that such a phrase could refer to other specialists in other fields who devote large amounts of time to writing but define their profession in different ways, such as engineers, businessmen, playwrights, novelist, professors, and scientists.

Second, a definition of writing may be based on the specific characteristics of writing. There are two issues when using such a technique. One is that teachers, researchers, and educators sometimes highlight characteristics that are shared by the other language skills. For instance, Byrne (1993) questioned the interpretation of writing as "graphic symbols: that is, letters or combinations of letters which relate to the sounds we make when we speak," since just as speech is more than the production, writing is more than the production of graphic symbols (p. 1). Writing involves a long process of joining symbols together to form words, which are then joined together to form sentences, each of which has to be based on language norms (Byrne, 1993). Therefore, many other researchers believe that writing is unique in both its mode of learning (Emig, 2011) and its features (Şahin, 2010). Mourtaga (2010) stated that such a skill demands a continuous productive process of active and intensive thinking to transfer ideas and thoughts to written communication—not merely letters and words on papers. Hammad (2014) believed that such a sophisticated process as composition "involves a series of forward and backward movements between the writer's ideas and the written text" (p. 1), which calls for an advanced level of generating ideas, planning, revision, controlling the language, goal setting, evaluation, and monitoring. In an attempt to accentuate the distinctiveness of writing, Byrne (1993) and Emig (2011) drew comparisons between writing and speaking. According to these authors, writing, in contrast to speaking, is a learned behavior, an artificial process, and a technological device—slow, stark, barren, and even naked as a medium. Its context involves an audience that is usually unknown and absent with

no immediate feedback. Writing usually leads to a visible graphic product, which is a more responsible and committed act related to the source and form of learning, since the written word is characterized by an aura, ambience, and mysterious meaning conveyed by such devices as punctuation, capital letters, sentence organization and construction, and, lastly, indications of sentences' boundaries.

The second problem in interpreting writing by its characteristics is that specialists, teachers, researchers, and even universities' websites, online writing centers, or their published writing guides emphasize different features. For example, in direct opposition to Byrne's (1993) artificial definition of writing—a production of a "sequence of [related] sentences arranged in a particular order and linked together in certain ways [coherently]" (p. 1)—Şahin (2010) interpreted writing as "a behavior including various closely interrelated complex skills such as punctuation, handwriting, spelling, creativity, and self-expression…as well as specific writing components such as grammar, mechanics, production, order of writing, linguistics, and understanding" (p. 777). The Writing Development Center of Newcastle University (2016) defined it as clear structure, logical development, analytical, filled with sources, explicit, objective, cautious, and formal language. Similarly, the book *Academic Writing*, published by the University of Technology Sydney (2013), interpreted good writing as "linear, informative, complex, formal, precise, objective, explicit, accurate, qualification/hedging, and responsibility" (pp. 3-4), while the University of Southern California's (2016) definition of good academic writing is as follows:

> A particular style of expression that researchers use to define the intellectual boundaries of their disciplines and their areas of expertise. Characteristics of academic writing include a formal tone, use of the third-person rather than first-person perspective (usually), a clear focus on the research problem under investigation, and precise word choice. Like specialist languages adopted in other professions, such as, law or medicine, academic writing is designed to convey agreed meaning about complex ideas or concepts for a group of scholarly experts.

Composition skills may also be interpreted based on the text types emphasized in classrooms or textbooks. In English, for example, as a consequence of there being various types of writing (of which, only the fundamental ones—narrative, expository, classification, descriptive, and argumentative—will be discussed here), the definition of the construct varies. This causes two issues. The first is that by observing the interpretations suggested, each type of text can be defined in two ways, as shown in Table 1.

Table 1

Text Type Definitions

Text Types	Succinct/Implicit Definition	Informative/Explicit Definition
Narrative	"A text in which events are related causally or thematically, and happen through time" (Wolfe & Woodwyk, 2010, p. 342).	"Relating a sequence of events which occurs over some period of time. Both what happens and the order in which the events occur are communicated to the reader. Effective narration requires a writer to give a clear sequence of events (fictional or non-fictional) and to provide elaboration for it" (Gallick-Jackson, 1997, p. 8).
Expository	"Texts that describe a system or even in terms of its processing or structure" (Wolfe & Woodwyk, 2010, p. 342).	"Presenting reasons, explanations, or steps in a process. Logical order should be used with appropriate sequencing of ideas or steps in a process. Effective expository writing should contain a main idea, supporting details, and a conclusion" (Gallick-Jackson, 1997, p. 8).
Descriptive	------	"The clear description of people, places, objects or events using appropriate details. An effective description contained sufficient and varied elaboration of details to communicate a complete sense of the subject being described. Details used were usually sensory ones selected to describe vividly what the writer saw, heard, smelled, touched, and tasted" (Gallick-Jackson, 1997, p. 8).
Argumentative	"…writing that takes a point of view and supports it with either emotional or logical appeals" (Crowhurst, 1990, p. 349).	"Writing that takes a position on an issue and gives supporting evidence to persuade someone else to accept, or at least consider, the position. [It] is also used to convince someone to take (or not take) an action…[It] helps you persuade people to see things your way, or at least to understand your position. Most of us have experienced the feeling of being a helpless victim − just standing by while something that we do not want to happen does occur. Although knowing how to argue will not eliminate all such situations, it will help you to stand up for what you want. You may not always win, but you will sometimes, and you will at least be able to put up a good fight" (Anker, 2010, p. 258).
Classification	"Dividing things by dividing into subcategories" (Saito, 1994, p. 50).	"Writing that organizes, or sorts, people or items into categories. It uses an organizing principle: *how* the people or items are sorted. The organizing principle is directly related to the purpose for classifying. For example, you might sort clean laundry (your purpose) using one of the following organizing principles: by ownership (yours, your roommate's) or by where it goes (the bedroom, the bathroom)" (Anker, 2010, p. 188).

The second issue is that some types of texts can be confusing when they encompass several subtypes (Wolfe & Woodwyk, 2010). For example, in the case of an expository text, there are several subtypes: time sequence, comparison and contrast, problem and solution, description, explanation or process, and cause and effect (Gunning, 2010, as cited in

Mongillo & Wilder, 2012, p. 28). Moreover, each kind of texts can be used in other domains. For instance, after defining descriptive writing, McCarthy (1998) argued the following:

> Descriptive writing is a 'maverick' sort of domain that—ideally—appears in other domains as well. In expository writing, we use description to present facts clearly. In narrative writing, we use description to show clearly what is happening, event-by-event. In persuasive writing, we choose strong descriptive words to present and support our opinions. (p. 5)

Finally, according to Swarts and Odell (2001), effective writing should not be defined by good punctuation or grammar, but, rather, by the context of its use. This means that the best interpretation of the skill is the one that describes it in "contextual terms, conventionalize it, and share it" (p. 26). Therefore, many teachers define writing by the approaches they use to teach it. However, this method of interpreting skills is not particularly simple, since it entails disputes among specialists regarding which of the approaches to use. For example, White (1993) refused to define writing as being concerned with the final, finished, correct, and neat product (i.e., the product approach), since composition will "remain…a matter of guessing about the rules which govern correctness" (p. 110). Instead, he preferred a definition (the process approach) that included the sophisticated judgments of writing competence—one that observed writing as a group of "overlapping activities, all of which have to do with critical thinking and problem solving: invention and prewriting, drafting, refining and rethinking, connecting, revising, and (finally) editing" (p. 111). When this is combined with other teaching approaches and contexts, it grows difficult to develop a single agreed-upon definition of writing or reach a mutual agreement among specialists regarding how to define the construct accurately.

To sum up, the first challenge that instructors, researchers, and learners of L2 writing skill will mostly likely encounter is driven from the fact that writing can be interpreted in various ways depending on its possible meanings, characteristics, pedagogy, text types, and learners' features. There is also an issue of defining the skill according to its applied teaching environments and methodologies, which will be elucidated below. Seemingly, no matter how writing is interpreted, such interpretations will only contribute to the difficulty of acquiring and developing it by learners.

2.3 Teaching Environments and Approaches to Writing Skills

According to Farooq, Uzair-Ul-Hassan, and Wahid (2012), "Writing is not a natural activity, so explicit instructions are required for learning this skill" (p. 184). They maintained that because it depends on elements such as linguistics, conventions, mechanics, cognition,

and production, writing is regarded as the "best academic achievement" (p. 184). Along the same lines, Breland and Jones (1982) summarized the hurdles involved in teaching writing that prevent learners from improving. They stated, for example, that L1 writing instructors lack proper equipment and knowledge, that they are unwilling to exert effort in teaching the skills, that they lack agreement regarding the weight that should be assigned to various errors (i.e., similar errors may be assigned different marks depending on the teacher), and that their learners do not practice the skill sufficiently. These claims mirror the ones emphasized by Khan (2011a, 2011b), and Mourtaga (2010) in regards to L2 students.

For Barnett (1989), "Work[ing] on teacher approaches to both first and second language writing indicates that much of our share disappointment and sense of futility may well result from our view of writing" (p. 31). Johns' (1990) own observation of the research on writing pedagogy concluded that, "Unfortunately, there has as yet been little discussion of the development of coherent and complete theories of ESL[/EFL] compositions allied to—or separate from—the various theories of L1 composition" (p. 24). Therefore, EFL writing skills teachers—whether native or nonnative—encounter many problems when teaching it. For example, Matsuda (2011) stated that because of the vast number of L2 learners located in English-speaking countries, even native writing skill teachers often experience teaching such students. The issue here is that despite the similarities in instructing native and nonnative speakers, instructors could face "a unique set of challenges" (i.e., cultural and linguistic differences) inherent in such learners. Consequently, due to these factors and others that can interfere with learning processes, "Instructors [have] to become more sensitive to the unique needs of ESL writers" (p. 674). Likewise, nonnative instructors of L2 writing are obliged to base their choice of classroom approaches on the trial and error method (Farooq et al., 2012). The true consequence of this struggle is that "the approaches of writing suffer from a drawback of either total control or total freedom…[and] dysfunctional writing styles can create obstacles through one's life" (Farooq et al., 2012, p. 185).

According to Silva (1990), "There is no doubt that development in EFL composition have been influenced by and, to a certain extent, are parallel to development in the teaching of writing to native speakers of English. However, the unique context of ESL composition has necessitated somewhat distinct perspectives, models, and practices" (p. 11). Ezza (2010) maintained that "a number of approaches have been proposed to provide guidelines for (successful) writing pedagogy" (p. 34). The four influential approaches to teaching writing are the product, process, genre, and process-genre approaches, each of which can be used in different learning contexts (i.e., traditional, blended, or online).

2.3.1 Traditional learning context (TLC).

The concept of the traditional learning context has achieved consensus regarding its definition. For example, Alkhatnai's (2011) interpretation of the TLC was that it

> Depends on face-to-face learning in a classroom setting, traditional learning is an engaged, active exploration of information guided by someone with experience, usually a teacher or professor. Learning can take place in small or large classes, lab sessions, or seminars. This instruction is sometimes called Face-to-Face Instruction and refers to activities carried out with the students and instructors meeting synchronously in the same room; it is also referred to as 'on-ground' or 'on campus' instruction. (p. 16)

For such context, many teaching methodologies for L2 such as *Audio Lingual Method*, *Suggestopedia*, *Silent Way*, *Communicative Approach*, *Community Language Learning*, *Comprehension Approach*, and others have been invented and adopted (Larsen-Freeman, 1990). Each of these possesses both advantages and disadvantages (Zainuddin, Yahya, Morales-Jones, & Ariza, 2011). Yet, in the Arab world, teaching English is undertaken by only applying one or two of these approaches and is described by Al-Hazmi (2006) as "dominated by a traditional top-down, textbook-oriented, teacher-led methodology" (p. 38). Additionally, Al-Seghayer (2014a, 2015) contended that two teaching methods—the Grammar Translation (see also Alharbi, 2015; Alresheed, 2008) and the Audio Lingual—have been largely employed in Saudi Arabia's education. He also provided a detailed depiction of the techniques EFL instructors would employ to undertake these two approaches there. These ineffective techniques include explanation and translation of texts, vocabulary, and grammar; copying; corrections and comments; repetition; structural analysis; chorus work; reading passages; and memorization. For him, there are other issues concerning L2 teaching summarized by the EFL teachers' incompetency in language testing, strategies, and teaching language elements, as well as their negligence and abandonment of using supplementary teaching materials and of developing their students' communication ability in both speaking and writing skills.

Regarding the writing skill, Alghizzi (2011, 2012), along with other Arab researchers (e.g., Ahmed, 2010; Al-Khasawneh, 2010; Mourtaga 2010), emphasized that some of the previously mentioned techniques were the reasons why Saudi EFL learners' written productions are not proficient. To clarify this, he found that the excessive use of the mother tongue in English classes has made students think in Arabic and translate their ideas into English, which produces vague sentences. Even worse, depending on these students' proficiency levels, textbooks, schools, and teacher methods, learners would be required to do a variety of insufficient writing tasks such as answering short sentence exercises; copying

tasks of short stories, plays, English textbooks, and paragraphs; translating words and writing their antonyms; and writing the English alphabet, words, sentences, short paragraphs, essays, and conversations on their own or by relying on a provided table of information. For final exam purposes, students were provided with sample sentences, paragraphs, and essays and were asked to memorize them. The availability of choosing to either recite the same memorized sample or to adjust them as learners had previously been taught to do depends on the topic of the exam. Finally, the description provided for teaching writing skill in Saudi Arabia shows that the EFL instructors do not apply the teaching methods specified for writing skills. These methods are explained below.

2.3.1.1 Product approach(es).

This method is the oldest of the composition teaching methodologies. The product approach focuses on the mechanical aspects of composition, such as grammatical structures and imitations of models written by so-called good writers. The primary concerns of this method are correctness and the final product. The product approach comprises two levels: the "sentence level" and the "discourse level" (Tangkiengsirisin, 2006, pp. 3-4). In each of these levels, the approach is called something different (and other factors change, as well). For example, at the sentence level, the product method is called controlled/guided composition (Raimes, 1991; Silva, 1990). The firm basis for controlled composition is the notion of structural linguistics "that language is speech" and of behaviorism psychology "that learning is habit information" (Silva, 1990, p. 12; Li, 2013). Therefore—and due to the dominance of the audiolingual method—this method views writing as a secondary, supporting skill of the other skills in language, such as speaking, and thus focuses on teaching only some aspects of writing, such as grammar (Al-Khairy, 2013; Hammill, 2014; Nordin & Mohammad, 2006; Silva, 1990; Tangkiengsirisin, 2006).

The role of the teacher here is that of a corrector, commentator, proofreader, and editor whose interest is in the correct usage of the language rather than in the expression of ideas (Tangkiengsirisin, 2006; Silva, 1990). Silva (1990) stated that learners manipulate language structures that they have learned previously and that the context of the writing is their own classroom, such that "there is negligible concern for audience or purpose" (p. 13). Tangkiengsirisin (2006) maintained that since the focus is on the sentence level, the exercises used include copying, substituting, and combining in order to facilitate the mastery of the level.

According to Li (2013), Hassan and Akhand (2010), and Badger and White (2000), in any class that adopts the product approach, students have to undergo four stages—familiarization, controlled writing, guided writing, and free writing—to learn how to write adequately. Badger and White (2000) explain these stages by stating that in the very beginning, instructors provide their students with a group of—for example—house descriptions, most likely modified to meet the students' levels and the teaching goals. Then, students are required to specify the names of the rooms and the adjectives and prepositions used in the texts. In the stage of controlled writing, students are asked to write simple sentences by relying on "a substitution table" (p. 153). At the stage of guided writing, students may write text based on a picture of a farm. Lastly, students are asked to write freely to describe their own (or their families') farms.

However, Silva (1990) maintained that researchers, educators, and others realized that composing grammatically correct sentences (as emphasized by controlled writing) is not enough for students—particularly EFL students—and that these students need to be capable of producing lengthy texts. In short, "What was needed was a bridge between controlled and free writing" (p. 13). Therefore, a shift towards the discourse, paragraph, or essay level (Tangkiengsirisin, 2006)—the current-traditional rhetoric (Silva, 1990; Raimes, 1991)—was suggested. Both Silva and Tangkiengsirisin believed the major objective of this shift was to raise learners' awareness and improve their ability regarding the "logical construction and arrangement of discourse forms" (Silva, 1990, p. 14). This meant that students were to be exposed to and learn the factors constituting a paragraph (i.e., topic, supporting and concluding sentences, and transitions) and essay (i.e., introduction, body, and conclusion), as well as the elaboration of options (e.g., defining, exemplifying, partition, comparing, contrasting, causal analysis, classifying, and illustrating) and, finally, organization norms and modes (e.g., exposition, description, argumentation, and narration).

Furthermore, Silva and Tangkiengsirisin concluded that there are three levels of applications for such an approach: simple, moderate, and complex/advanced. In the simple application, learners are asked to write a paragraph based on information provided in a table, to order sentences in scrambled paragraphs, or to choose from alternative sentences that fit the context of a given paragraph or essay. At the moderate level, students are assigned model texts and are asked to analyze these texts with regard to words, idiomatic expressions, and grammar rules. They are then asked to apply what they have learned in a new writing of their own. At the advanced level, learners are simply assigned a topic and are told to write down topic sentences, supporting sentences, and concluding sentences on a separate piece of paper.

Then, they are asked to execute their writings based on these outlines. The application of such exercises is only possible if students are exposed to the features, organization, and modes of essays beforehand. This version of the product approach views the skill of composition as the ability to organize and arrange sentences in paragraphs and essays—that is, as the skill of "identifying, internalizing, and executing…patterns" (Silva, 1990, p. 14; Tangkiengsirisin, 2006, p. 5).

Although supporters of the product approach have claimed that it develops students' writing abilities (Nordin & Mohammad, 2006), many critics believe that it suffers from many disadvantages. Barnett (1989) emphasized that amateur and professional writers are likely to be unable to submit a first writing attempt as a final draft. Moreover, what teachers consider to be final drafts are—for students—first drafts that need to be embellished. Instructors of the product approach also have inaccurate ideas of what constitutes good writing—that is, instead of teaching their students the importance of coherently "communicating a message," they focus on "surface-level fine tuning" (p. 32). In other words, "This approach disregards the fact that communication, not absolute accuracy, is the purpose of writing," thereby leading students to believe that the skill regards only the evaluation of text (Tangkiengsirisin, 2006, p. 4). Even with the indispensable focus on syntactic accuracy, a lot of students who are able to compose grammatical and accurate sentences may not be capable of presenting written texts appropriately. In fact, their lower rates of grammatical mistakes could be attributed to the fact that they are "avoid[ing] taking risks and will not be able to reach beyond a current level of competence" (Tangkiengsirisin, 2006, p. 4). Both Li (2013) and Palpanadan, Salam, and Ismail (2014) have explicitly asserted that "this teaching method guided by behaviorism goes for standard writing norms leading to mechanic and stereotyped writing products without creativity and individuality" (Li, 2013, p. 1). Tong (2007) also suggested that the approach failed "to develop learners' interpersonal competence and creativity" (p. 53).

Finally, given the many disadvantages of the product approach, researchers and teachers have been forced "to reassess the nature of writing and the ways writing is taught" (Nordin & Mohammad, 2006, p. 76). Interest in the processes and stages students undertake when writing a text grew (Al-Khairy, 2013) with the revolutionizing shift of seeing composition as a non-linear skill involving recursive and complex processes (Nordin & Mohammad, 2006).

2.3.1.2 Process approach(es).

Although the process approach has been considered to be one of the most successful methods for teaching English writing since it regards the skill as a developing, complex,

cognitive, and recursive process (Farooq et al., 2012; Silva, 1990), "The term process writing has been bandied about for quite a while in ESL classrooms" (Seow, 2002, p. 315). Perceptions of the process method changed as a consequence of the dissatisfaction regarding the product approach (Silva, 1990). There was a shift from seeing compositions as finished products to seeing compositions as techniques and stages by which composing can be improved (Hassan & Akhand, 2010). According to Li (2013), the process approach depends "on communicative theory and stresses the writers' interactive principle" (p. 1), and psycholinguistic cognitive theory (Guleff, 2002). The issue here is that there is no universally accepted definition of such an approach (Sun & Feng, 2009), nor is there agreement regarding the major processes that students must undergo before handing in their written texts. For example, Montague (1995) defined the process method as the following:

> …a teaching approach that focuses on the process a writer engages in when constructing meaning. This teaching approach concludes with editing as a final stage in text creation, rather than an initial one as in a product oriented approach. The process oriented approach may include identified stages of the writing process such as: pre-writing, writing and re-writing. Once the rough draft has been created, it is polished into subsequent drafts with the assistance of peer and teacher conferencing. Final editing and publication can follow if the author chooses to publish their writing. (p. 15)

Murray (2011) interpreted the process approach generally as the following:

> ...the process of discovery through language. It is the process of exploration of what we know and what we fell about what we know through language. It is the process of using language to learn about our world, to evaluate what we learn about out world, to communicate what we learn about out world. (p. 4)

With respect to the structure of the process approach, Badger and White (2000) contended that there were various opinions regarding the stages. For example, various authors suggested two, three, four, or even eight stages in the process approach (cf. Alodwan & Ibnian, 2014, Barnett, 1989; Bayat, 2014; Farooq et al., 2012; Flower & Hayes, 2011; Hassan & Akhand, 2010; Li, 2013; Murray, 2011; Sun & Feng, 2009; Tangkiengsirisin, 2006). The reason for this disparity could be that some researchers consider sub-stages to be major ones. In any case, for the purposes of this paper, the major stages are four: prewriting, writing, revising, and rewriting.

In this approach, the process of learning to write revolves around students' status as subjects. This means that in the classroom, the students are the center of attention. The teachers' role, on the other hand, is that of a facilitator whose aim is to guide the students through the stages of writing (Li, 2013; Nordin & Mohammad, 2006; Tangkiengsirisin, 2006). A typical class using such an approach—as explained by Badger and White (2000) and

Hassan and Akhand (2010)—would begin with the teacher assigning students a topic. The students would then be asked to generate ideas, organize them, and plan the structure of their writing. As soon as students finished writing, they would be required to self-correct and/or peer-correct their texts, and then to rewrite them based on feedback. Ultimately, these drafts would be corrected by the teacher, who would then ask the students to follow the instructor's suggestions and comments when redoing their writing for the last time.

Barnett (1989) summarized some of the advantages of the process approach by stating that students eventually change their perspective of teachers in such a way that they stop seeing teachers as proofreaders and correctors, whose concern is primarily the proper or grammatical usage of the language, and begin to see them as specialists who read their writings in order to understand them and to help the students fix the issues that might impede communication. The hope is that since this process means more to the students, too, the students will write more. The teachers will also find the process gratifying, since their suggestions and comments will encourage the intellectual evolution of the students, thereby improving the students' critical thinking and reasoning skills. In other words, in the short term, students will be able to more intelligibly express their thoughts and ideas, while, in the long term, they will be capable of using critical thinking.

However, this approach has many disadvantages. Silva (1990) argued that this method does not tackle the fundamental issues of L2 writers properly, nor does it prepare them for academic work. According to Reppen (2002), the process method is focused on the stages of writing, and thus usually excludes compositionally correct forms, as well as the conventions and writing styles of various genres. This problem is accelerated when the learners are characterized as EFL and have different cultural backgrounds or when the learners' knowledge reflects poor mastery of modes of writing. The irony is that while even though principal writing features (and sometimes grammar rules, as well) are excluded from the teaching, students' evaluations are still often based on their mastery of such things (Reppen 2002). Seow's (2002) observation was that the method "in the classroom is highly structured as it necessitates the orderly teaching of process skills, and thus it may not, at least initially, give way to a free variation of writing stages" (p. 316).

Johns (1986) emphasized that when strictly following the major stages of such an approach, instructors could be doing a disservice to the students, since the approach must be tested based on the tasks learners are required to execute. Breuch (2011) stated that post-process researchers believe that the processes of writing (i.e., prewriting, writing, and rewriting) no longer accurately explain the act of writing, since the act of writing is

minimized by this paradigm to sequences of codified stages that can be taught. Flower and Hayes (2011) highlighted the fact that the method has no "clean-cut stages," which means that learners can engage in the stages simultaneously and that when teachers make sharp distinctions between the operations of these processes, they "may seriously distort how these activities work" (p. 255). Li (2013) asserted that the weakness of this approach is that it will partly limit the production of writers' free creativity, as instructors are viewed as the role models of standard answers. Hammill (2014) maintained that the two aforementioned approaches (product and process) "are not mutually exclusive," since the emphasis on composition procedures "does not necessarily preclude a consideration of the nature of the final product" (p. 32). Nonetheless, because the process method ignores the academic, social, and cultural settings in which writing takes place (Al-Khairy, 2013; Nordin & Mohammad, 2006), and since it lacks a focus on teaching students that certain types of writing are performed for specific purposes (Farooq et al., 2012), the genre approach was developed.

2.3.1.3 Genre approach(es).

Hammill (2014) maintained that the genre approach is regarded "as a reaction to the dominance of process approach" (p. 36). For Johns (2002), the former method is a shift from the "learner-centered classroom" to the "contextual approach, to analyses of the situations in which writing takes place" (p. 3). This approach is dependent on genre analysis theory (Li, 2013), and it is consistent with English for Specific Purposes (ESP) approach (Dudley-Evans, 1997; Silva, 1990). This means that it shares the same assumptions as the ESP method, in the sense that both approaches believe that there is a "'short-cut' method" of improving learners' writing ability by extensively exposing them to various types of texts in a relatively short period of allotted time (Dudley-Evans, 1997). However, the concept of genre has been applied by teachers differently and is loosely defined due to its evolution "in ways to reflect real uses of written (and oral) discourse among cultures, social groups, and communities of users" (Grabe, 2002, p. 250). For example, both Dudley-Evans (2002) and Macken-Horarik (2002) discussed the various types of such approaches, the issues facing their applications, and how to apply them appropriately (if possible). Along these lines, Johns (2002) cited the multiple meanings of the genre provided by Freeman and Medway (1994b) to be "(a) primarily literary; (b) entirely defined by textual regularities in form and content, (c) fixed and immutable, and (d) classifiable into neat and mutually exclusive categories and subcategories" (p. 3).

Badger and White (2000) stated that the genre approach is seen as an extension of the product approach; however, Nordin and Mohammad (2006) maintained that although the two approaches "view…writing as predominantly linguistic…the genre approach…places a greater emphasis on the social context in which writing is produced" (p. 78; see also Cope & Kalantzis, 2012; Paltridge, 2014). Farooq et al. (2012) also offered an interpretation of the genre method:

> A text either spoken or written that serves a particular purpose in a particular context and is composed of a series of conventions. The idea behind this approach is that writers write not only for their own enjoyment, but also in different contexts, for different purposes and in different ways. (p. 185)

Paltridge (2002) highlighted another problem with the field of the genre approach: the fact that in the literature, genre and text types have been used interchangeably. He argued that a genre is a subordinate/hypernym of the text type—a hyponym. To clarify this, when a journalist, for example, writes a letter to the editor (genre), the journalist has to choose the purpose for writing such a letter (e.g., argumentative or problem-solution), comprehend the expectations, and follow the conventions constituting these text types (p. 74).

Li (2013) maintained that the genre approach procedures include the modeling of text analysis, imitative composing, and independent composition. Badger and White (2000) explained that in a typical class using such an approach, the instructor will first supply his or her students with a model text of, for instance, a farm, as described by an estate agent for the purpose of selling the property. Learners are requested to look carefully through the text to analyze the grammar rules, adjectives, and prepositions, as well as the names of the properties, machines, and animals used in the text. Social context is critically important for learners to consider; this includes whether the description provided is based on a real visit to the farm, the identity of the person selling the property, and the prospective buyers. Then, students are required to write a short text with the partial help of the teacher. Finally, students are asked to write, individually, complete essays/paragraphs describing the farms of their families.

Like its predecessors, the genre approach is seen to have positive and negative effects. On the positive side, Reppen (2002) stressed that the approach improves learners' "metalinguistic awareness [which] empowers…and gives them tools to manipulate information and accomplish different purposes through writing" (p. 321; Paltridge, 2002). Hyon (2002) explained this by saying that through the explicit teaching of the genre approach, learners' production and processing of texts are facilitated, they become equipped to

sensitively analyze various texts effectively, and they improve the organization of their compositions and their confidence and reading speed. In relation to the negatives, Silva (1990) asserted that the approach's focus on composing in different fields is questionable. Hyon (2002) stated that teaching all genres to students adequately is impossible. In fact, too much exposure to the various genres is "counter-productive," since it does not leave space or allow students to articulate their thoughts and ideas, forcing them to rely instead on the instructor to provide them with suitable models and materials (Nordin & Mohammad, 2006, p. 79). The genre approach also leads to prescriptivism and limits students' own experience with and creativity in writing texts—a severe problem that may make their texts appear similar (Li, 2013; Tangkiengsirisin, 2006). One possible solution for such issues is the process genre approach.

2.3.1.4 Process genre approach(es).

Badger and White's (2000) article was an attempt to analyze the strengths and weaknesses of the above-mentioned methods (product, process, and genre) of teaching writing. What they concluded was that these approaches complement one another, and by relying on their positive aspects, they can be combined with the process genre approach or the eclectic approach, as used by Farooq et al. (2012), Nordin and Mohammad (2006), and Cutler and Graham (2008). Nordin and Mohammad maintained that "such an…approach offers advantages such as a more focused use of texts as models without sacrificing the flexibility to acknowledge elements of other approaches" (p. 79). For Li (2013), this method is an integration of the communicative theory, the genre analysis theory, and behaviorism (on which the previous methods of process, genre, and product were based). The definition proposed for the process genre method is based on Badger and White's view of composition and development, which suggests that a classroom using a process genre method recognizes the following:

> Writing involves knowledge about language (as in product and genre approaches), knowledge of the context in which writing happens and especially the purpose for the writing (as in genre approaches), and skills in using language (as in process approaches) writing development happens by drawing out the learners' potential (as in process approaches) and by providing input to which the learners respond (as in product and genre approaches). (pp. 157-158)

For Cope and Kalantzis (2012), the genre approach demands the "need to analyze critically [and learn] the different social purposes that inform patterns of regularity in language – the whys and hows of textual conventionality" (p. 2). In other words, for each text,

there are certain elements that need to be followed and addressed, such as the mode, the audience and writer relationships, the organization, and the channels (Badger & White, 2000). The trick here lies in combining the aforementioned elements with the process stages that writers experience when producing texts.

Furthermore, Li (2013) maintained that the application of the process genre approach requires four procedures: demonstration analysis, imitative analysis, skill training, and independent composition. For Sari and Saun (2013), the procedures are six: preparation, model and reinforcement, planning, joint writing, individual writing, and revising. Badger and White (2000) further hypothesized the application of the process genre approach in writing skills classrooms, stating the need for hierarchal steps.

The first step, which is undertaken by the instructor, involves duplicating situations as thoroughly and intensively as possible. After that, the instructor supplies students with sufficient help to specify the writing's purpose, along with the other factors of the social context. For instance, students who envisage themselves as the estate agents must guide their attention towards writing descriptions meant to sell farms (e.g., goal, reason, or purpose), meaning that they must convince a specific group of people (i.e., buyers and tenants). Also, their descriptions must contain specific information (i.e., field), and they must recognize that various methods and techniques exist for how to write these description of farms (i.e., modes). The second step involves the instructor and students working collaboratively to analyze and supply the grammar, vocabulary, styles, and organization appropriate for the genre they are addressing. Lastly, students begin composing their own text, using the required knowledge to go through multiple steps such as drafting, revising, editing, and rewriting before they can submit their finished products to their instructors. Badger and White (2000) acknowledged that their approach is subject to change, since various genres require various kinds of knowledge and applications and depend on students' composition proficiency levels. For example, if students know the characteristics of a particular genre, they require no more input except on the elements they lack. These areas, along with students' other deficits, can be addressed through three potential sources of input: teachers, students, and model texts.

Given its tremendous flexibility, Li (2013) stated that the process genre approach cannot be used with learners who have relatively low levels of proficiency or in classrooms with 40 or more learners. Farooq et al. (2012) contended that the approach is like any other method of teaching writing, suggesting that it can be based on the assumption that it is "appropriate for all students in all settings" (p. 185). Meanwhile, Adas and Bakir (2013) deduced that the traditional methods and context of teaching EFL composition would not help much, and that

if teachers keep insisting on applying them, there would be a risk of turning the activities that take place inside classrooms into monotonous and passive ones. One tangible solution, as they suggested, is to embellish these approaches and the environment with technologies that will hopefully turn the learners into better achievers.

2.3.2 Writing Skills and Technology

According to Saddler and Asaro-Saddler (2013), not only might composition be one of the hardest "facets of the language arts," but its "instruction is not as effective as it should be" (pp. 20-21). The reason for the lack of effectiveness and the limited success in improving L2 learners' writing abilities, which is not at the point that educators had hoped for, is the fact that the skill is taught in traditional classrooms. The task of teaching these learners the basic requirements of composition skill—finding a thesis, supporting, organizing, revising, and editing, and presenting at a high or errorless level—is challenging, and it will be more difficult if the learners have different learning style preferences that such a context cannot handle (Al-Hammadi, 2011). Thus, the only way to address this issue is to enhance the teaching environment through technology or to change the environment to a completely new one.

In the field of education and Second Language Acquisition (SLA), there have been numerous advances in technologies that are seen as convenient and reliable means of promoting learning (Yamada, 2009). The ubiquity of technology, as seen in people's usage of technological tools in every aspect of their lives (e.g., finance, communication, and transportation), has led to the use of technology in education, as well (Gaudelli, 2006). Johnson (2006) and Hrastinski (2008) contended that current education technologies can be categorized into two groups: synchronous, in which the technologies exist in immediate and real-time formats, such that interactions and responses occur between teachers and students simultaneously (e.g., audio-video conferencing, chatrooms, text messaging, discussion groups, joint web browsing, desktop sharing, and electronic whiteboard), and asynchronous, which include delayed response formats in which interactions and responses occur between the teachers and students non-simultaneously (e.g., Listserv, bulletin board, wikis, blogs, Blackboard, WebCT, iWRITE, Moodle, and computers.)

Selfe (1999) was one of the first researchers to call for a change in perspective towards technology among writing specialists. She maintained that despite teachers having justifiable reasons to rely on writing textbooks and ignore technology, world-leading countries, such as the US, had changed their position to view literacy as a topic that encompassed knowing how

to use technology rather than viewing literacy as only knowing how to read and write. Hence, significant amounts of these countries' budgets were designated towards the implementation of technologies in the education cycle (Selfe, 1999). However, the current level of adaptation of such demand, including conducting experiments testing the effectiveness of various technologies in the field, is still poor, questionable, and progressing at a slow pace. For example, Miller (2001) reviewed the studies on distance education and writing in *Computers and Composition Journal* over a period of six years (from 1994 to 1999) and found that only 12 articles (divided between theoretical and practical) had been written during that period. Similarly, Moran (2003) presented an extensive analysis, overview, and discussion of the same journal over the course of 20 years and showed that despite the total of 74 articles tackling writing and technologies—which sought to show the journal's enthusiasm and optimism towards the current teaching of writing—the journal was "naive" in addressing "antique hopes," particularly in terms of determining whether various types of programs or technologies (e.g., word-processing, personal computers) would foster learners' writing (p. 354). There are also other more general reasons that can be said to have affected the poor implementation of technology in composition.

Although technology can be invented or even modified to accommodate a variety of learning styles of students (Chrichton & Kinash, 2013) while also promoting various kinds of interactions (e.g., student-to-student, student-to-technology, student-to-teacher, student-to-content, and others), there are some issues that could limit technology's use—or even its effects—in education (Dennen, 2013; Kim, Kozan, Kim, & Koehler, 2013). For example, Shearer (2013) emphasized that "there is no technology that addresses all the needs of learners or the results expected by the constituents involved in…education enterprise" (p. 251), and Moore (2013) suggested the possibility of having instructors "and others who subscribe to the 'if not invented here it cannot be good enough' syndrome" (p. 248). Even with the rapid increase in technological advancement, some teachers will remain skeptical of the true value of technology (Gillam & Wooden, 2013), possibly because they know little about technology's potential to improve learning (Garrison & Akyol, 2013). Moreover, instructors' use of technology is affected by their long-lasting belief that such innovations exploit, rather complement, the fundamental principles of the writing instruction: pen and paper and face-to-face instruction (Anson, 1999). It is possible that instructors could use a particular technology because it is new rather than because of its determined effectiveness (Arafeh, 2004) or to make courses more efficient. However, teachers and students "care more about minimizing their effort than mastering content" (Sapp & Simon, 2005, p. 477).

Finally, there is another major reason why educators, especially writing skill instructors, may feel reluctant to incorporate technologies into their courses. It is true that developments in technology have opened the door to new teaching and learning environments, but how are teachers supposed to use and apply technologies effectively for students when the theories building such contexts are still being developed? In other words, with more technologies, not only are different contexts possible, but different terms also appear to make the building of the various theories of education and technology more problematic regardless of the fact that certain terms have received acceptance by certain groups of practitioners (F. Saba, 2013). The issue with these constructs is that they are either undefined or are defined poorly through the assigning of various meanings (F. Saba, 2013). Moore and Kearsley (2012) and Kanuka and Conrad (2003) asserted that what causes the most difficulty for people seeking to understand technology integration is that many terms have been used interchangeably, despite meaning different things—as is the case of blended learning and online learning (i.e., distance education).

2.3.2.1 Blended learning context (BLC).

BLC, or any of its synonymous terms, such as hybrid learning or even mixed-mode delivery (Delialioglu & Yildirim, 2007; Greener, 2008; Kanuka & Conrad, 2003), is believed to be one of the prerequisite steps to transferring whole classrooms and learning and teaching environments to a new level—that is, a distance education context or electronic virtual classrooms. The concept of BLC is based on the constructive theory in which learners' knowledge is built through a reliance on both present and past knowledge (Mohar, Sraka, & Kaučič, 2012). In other words, the *ad hoc* roles of both instructors and students have changed; rather than teachers being the sole source of information and students being the receivers of said information, teachers now serve as consultants, and students are responsible for self-autonomy (Mohar et al., 2012). The popularity of this educational method for use at the pre-university, undergraduate, and graduate university levels in a variety of fields, along with the proven benefits of this method, suggest that the approach will continue to be used in the foreseeable future. Garrison and Kanuka (2004) maintained that BLC "has the proven potential to enhance both the effectiveness…efficiency [and convenience] of meaningful learning experiences" (p. 95; see also, Vaughan & Garrison, 2006). However, with regard to the construction of different applications and usages in experiments, various definitions have been proposed (e.g., Akkoyunlu & Soylu, 2008; Finn & Bucceri, 2006; Hastie, Hung, Chen,

& Kinshuk, 2010; Khan, Noor-ul-Quyyum, Shaik, Ali, & Bebi, 2012; Klein, Noe, & Wang, 2006; So & Bonk, 2010).

2.3.2.1.1 Definitions and types of BLC.

For Oliver and Trigwell (2005), defining BLC is problematic due to the fact that the term BLC is still ill-defined and is currently used incoherently and inconsistently, which makes it impossible to build a reliable theoretical framework through which data can be analyzed and research results can be synthesized. On the other hand, Driscoll (2002) emphasized that the reason why the term is still gaining popularity is the various interpretations that only "illustrate the untapped potential of blended learning" (p. 1). For example, Marsh (2012) deduced that the construct has developed to include a wide range of learning environments and approaches, such that it "refer[s] to any combination of different methods of learning, different learning environments, different learning styles" (p. 3). Verkroost, Meijerink, Lintsen, and Veen's (2008) interpretation of BLC was "the total mix of pedagogical methods, using a combination of different learning strategies, both with and without the use of technology" (p. 499). Yoon and Lee's (2010) also offered a general definition of the term:

> …bringing together the positive attributes of online and offline education, including instructional modalities, delivery methods, learning tools, etc., in relation to language teaching and learning approaches and methods in order to reinforce learning process, to bring about the optimal learner achievement, and to enhance the quality of teaching and learning. (p. 180)

In all of the above general definitions of BLC, it may be possible to take one aspect of the term and generate more specific interpretations that meet the purposes of individual studies. To illustrate this, Singh (2003) identified five dimensions of BLC, some of which have "overlapping attributes:" blending self-paced and live collaborative learning, blending offline and online learning, blending learning, practice and performance support, blending structured and unstructured learning, and blending custom content with off-the-shelf content (p. 54). Johnson and Graham (2015) identified only three types of BLC (enabling blends, enhancing blends, and transforming blends), whereas Verkroost et al. (2008) specified four: self- or teacher-directed, individual or group, structured or unstructured, and face-to-face or remote. Driscoll's (2002) four types of BLC involved mixing or combining instructional technology with actual job tasks, any form of instructional technology, modes of web-based technology, and various pedagogical approaches (p. 1).

In each of these types, BLC is defined differently. It is true that the common trend of most educators and researchers is to view the approach as a combination of technologies and traditional face-to-face instruction—such as Kanuka and Conrad's (2003) and Moore and Kearsley's (2012) definitions—and in fact, some authors would go even further in specifying the name of the technology in their proposed definitions, as in the case of Hastie et al.'s (2010) definition of an "integration of physical classroom and cyber classroom settings using synchronous learning to enable unlimited connectivity for teachers and students from any part of the world" (p. 10). However, these represent only one type of BLC. Nevertheless, Graham (2013) not only stated that some interpretations of the term were vague in terms of failing to identify the differences between BLC and other teaching formats (i.e., TLC and OLC), he also attributed the diversity of proposed definitions to other factors. He, for example, maintained that some interpretations could indicate—sometimes by specific numerical amounts—the reductions in traditional classes and the time allotted for integrated online learning. This is the case in the following two definitions of BLC: BLC "as learning that combines instruction lead learning with online learning activities leading to reduced classroom contact hours" (Saleh, El-Bakry, Asfour, & Mastorakis, 2010, p. 293) and BLC as "situations where an instructor personally interacts with learners during occasional face-to-face meetings, but the vast majority of learning occurs via Web-based distance learning where the learner interacts with course materials, the instructor, and other trainees using Web-based electronic media" (Klein et al., 2006, p. 666). Other interpretations specify BLC's prospective quality or outcomes (Graham, 2013), as in Finn and Bucceri's (2006) definition of BLC as the "effective integration of various learning techniques, technologies and delivery modalities to meet specific communication, knowledge sharing and informational needs. It is a means to an end" (p. 2).

2.3.2.1.2 BLC's benefits, drawbacks, and challenges.

The topic of BLC has been discussed thoroughly (i.e., from different angles, including advantages, disadvantages, challenges, assessments, misconceptions, teachers' and students' perceptions, and its framework and design) by various types of specialists in different fields. In the L2 field, researchers and educators have also covered all of the aforementioned angles—with little to no elaboration on weaknesses and challenges—to determine what, when, where, how, and why BLC will contribute to better learning opportunities than the traditional learning context. In other words, BLC's topics that authors discussed were, for instance, its description (Shaykina, 2015), interactions (Liang & Bonk, 2009), teachers and

students' attitudes, perceptions, and experiences towards it (e.g., Adas & Abu Shmais, 2011; Al Zumor, Al Refaai, Bader Eddin, & Aziz Al-Rahman, 2013; Bijeikienė, Rašinskienė, & Zutkienė, 2011; Dashtestani, 2014; Gleason, 2013; Hong & Samimy, 2010; Kobayashi & Little, 2011), impact on reading, writing, listening, speaking, vocabulary, and grammar (e.g., Al-Jarf, 2007b, 2013; Al-Shaer, 2013; Bañados, 2006; Ghahari & Ameri-Golestan, 2014; Grgurović, 2011), teachers' pedagogical knowledge (e.g., Badawi, 2009), and language proficiency (e.g., Obari, 2012; Obari & Lambacher, 2014).

Regarding the advantages, many researchers found that BLC has many positive sides for learning all language skills in general and writing skill in particular. For example, Marsh (2012) summarized the prospective strengths of the method for EFL learning by stating that it would provide an environment for the less stressful practice of L2 outside class, the flexibility to study anywhere at any time, and more personalized and individualized learning support and experience. BLC also accommodates a variety of learning styles, increases the engagement of students in learning, and aids students in improving their necessary and valuable updated learning skills (see also Klímová, 2008). Pardo-Gonzalez (2013) added that the context would make learners recognize the value of face-to-face teaching because when students are able to practice their language online, their self-confidence and "language ego" would most likely increase and lead them to use the language inside classrooms (p. 59). Aborisade (2013) maintained that since BLC allows the incorporation of various communication and collaborative platforms and learning contexts, it would therefore aid students in establishing a sense of community and collaboration for sharing their experiences of learning, foster reinforcement, and promote their accessibility of extra learning materials. Eventually, the learning experiences of such learners would be enhanced. Peachey (2013) stated that with the ability to teach part of the course online, BLC offers an opportunity to increase the time allotted for learning over a much greater timescale than would be possible with traditional classroom instruction alone. Such a feature is needed for a difficult skill such as writing. Eydelman (2013) particularly contended that BLC's use of technologies has many advantages for writing skill in motivating L2 students to write and address a wide range of audiences, embellish their general ideas, and refine their texts by allowing different types of editing (peer-and-teacher corrections). For her, the context also supplies learners with additional channels to interact and collaborate more, share their learning experiences and the difficulties they encounter, and to communicate informally.

With the practical application of BLC, many L2 writing researchers reported positive effects on different aspects of writing skills and possibly on factors essential for L2 learning,

regardless of technology or instruments used. For example, Miyazoe and Anderson (2010) found that Japanese EFL learners' writing ability in forums and wikis progressed in terms of distinguishing between styles of English writing. Yoon and Lee (2010) concluded that BLC in composition skills led to effective learning of written content, organization, structure, and mechanics. Roy (2012) emphasized that the analysis of website exercises might contribute to getting Japanese EFL learners involved in constructive writing practice and to an increase in critical thinking. Liu's (2013) study showed that the application of BLC in the Chinese EFL context contributed to an increase of various types of interactions—student-to-student and student-to-teacher, development of learners' motivation into becoming autonomous and independent learners, and development of writing ability—as well as a reduction of students' anxiety in communication. Both Janfaza, Shahsavari, and Soori (2014), and Niazi and Pourgharib (2013) found that emails exchanged between Persian EFL writing instructors and students in traditional classes resulted in the significant improvement of experimental groups in comparison to control groups. Also, Erkan (2013) concluded that Turkish EFL students' email exchanges with foreign counterparts not only promote the formers' writing self-efficacy, but it also motivated them to have positive attitudes towards writing skill and classes (cf. Lin, 2002). Finally, Shih (2011) maintained that the application of peer assessment on posted EFL undergraduate students' writing assignments on Facebook can be both effective and interesting for writing classes and therefore develops such learners' composition ability.

Finally, the weaknesses and challenges of BLC are one of the relatively little-tapped into topics in the L2 field and unfortunately have never been addressed in terms of composition skill, although many educational studies (e.g., Ocak, 2010) believed that they could emerge from the context reliance on different kinds of technologies. This means that the negative sides of BLC could potentially arise from faculties, teachers, students' lack of technical support, motivation and knowledge towards technologies, rigid school textbooks and curricula, and the large workload and timescale resulting from their incorporation (Ocak, 2010). Almost all of these weaknesses were also reported by Al Zumor et al. (2013) and Park and Son (2009) for EFL learners, while Bijeikienė et al. (2011) added the disadvantage of reducing face-to-face classes, a feature that is needed for guiding and assisting low-level learners. Alpala and Flórez (2011) maintained that EFL instructors will most likely encounter many challenges when implementing BLC with their students, such as adjusting the incorporated textbooks and materials equally between the face-to-face and the virtual features of the course, conducting a permanent evaluation of the course, being creative and

resourceful, and focusing on the students' learning processes, needs, expectations, and how to make them autonomous rather than emphasizing the importance of the technologies used.

2.3.2.1.3 BLC in Saudi Arabia.

In Saudi Arabia, the application and investigation of BLC, regardless of the different technologies available, is still primitive and scarce (Al Ebaikan & Troudi, 2014). The field suffers from a lack of studies addressing the future of the area, but it has the potential for developing Saudi EFL learners—in particular, women (Alebaikan, 2012; Al Ebaikan & Troudi, 2014). According to Al Ebaikan and Troudi (2014), one of the first implementations of BLC initiated by an organization was conducted by the College of Applied Studies and Community Services (CASCS) at King Saud University. This college provided BLC courses to those students with extremely low GPAs in order to help them raise their GPAs. It also provided an online diploma program to increase students' enrollments in the university; in such courses, 70% of students' time was allotted to classes online, and 30% was allotted to face-to-face instruction. Similarly, Arab Open University provided blended learning, which was divided as follows: 25% of time was devoted to traditional in-class instruction, and 75% was allotted to online instruction. Moreover, the National Center of E-learning and Distance Education currently offers a variety of multimedia resources that can be used in blended learning and online. Finally, King Khalid University has established three kinds of e-courses: one is purely online, and the others are supplementary and blended (Al Ebaikan & Troudi, 2014).

However, the so far minimal attempts to incorporate BLC into Saudi universities could be attributed to the challenges that educators have to overcome and the principles they have to take into consideration in order to apply the method effectively or to employ online discussion within BLC. The challenges include the difficulty in adopting the method within traditional university culture, issues with managerial and organizational support, the need for instructor and learner self-discipline, perspectives on being accustomed to only giving and receiving instruction inside the classroom, comfort in using the technology, the values and norms of the society, the choice of the best design of BLC, and, finally, time constraints (Alebaikan & Troudi, 2010a). The identified principles included time demands, e-plagiarism, e-pedagogy (i.e., the methods and structures of online instruction), learning management system tools, and infrastructure (i.e., Internet availability, computer laboratories, and technical support) (Alebaikan & Troudi, 2010b).

On the other hand, in Saudi Arabia, EFL writing skills instructors at the university level—who are likely to access reputable journals and books that report studies indicating the effectiveness and positive impacts of adopting different kinds of technologies and social media tools in traditional composition classes around the world, as well as EFL students' positive attitudes towards such technologies (e.g., AbuSeileek, 2006; Alwi, Adams & Newton, 2012; Blackmore-Squires, 2010; Dalir, Jafarigohar, & Soleimani, 2013; Kitchakarn, 2014; Liaw & Johnson, 2001; Liu & Sadler, 2003; Yoon, 2011; Zarei & Al-Shboul, 2013)—have taken the initiative to experiment with such technologies and social media tools. In other words, EFL writing skills instructors in Saudi Arabia have tested the use of emails, Facebook, Wikis, blackboards, Nicenet, computers, word processors, blogs, online courses, multimedia, mind-mapping software, the WhatsApp application, and Twitter, among others, as well as seeking to determine participants' attitudes towards these technologies. Some authors (e.g., Abdelrahman, 2013; Ahmed, 2015; Al Fadda, Almasri, & ALShumaimeri's, 2011; Al-Hammadi, 2010; Al-Jarf, 2002, 2004a, 2004b, 2007a, 2009a, 2013; Aljumah, 2012; Al-Menei, 2008; Alsaleem 2014; Alshumaimeri, 2011; AbuSeileek; 2006; Fageeh, 2011; Mahmoud, 2014; Zaid, 2011) found that not only did Saudi EFL learners benefit from BLC, such that those using technologies and social media significantly outscored the control groups in writing tests, but they also had more positive attitudes towards learning and wanted to continue using technologies in the designated skill areas—and even in other skill areas. Finally, with such positive results, the question of whether teaching EFL students writing remotely (i.e., online learning context) would lead to more or less benefit than BLC is still open for further investigations.

2.3.2.2 Online learning context (OLC)[2].

Irele (2013) contended that "the worth of distance education [DE] program is still not universally accepted despite the fact that it has been around for well over a century" (p. 494). The reasons behind this lack of acceptance could be the lack of historical perspective, the existence of confusing terminologies, and the absence of term validity—all of which have caused theory building surrounding DE to be problematic (F. Saba, 2013, pp. 49-50). The issue with these challenges is that they may lead researchers and educators to base their experiments on pure improvisation rather than on theories that have been authenticated and tested for validity and applicability.

---

[2] Although the online learning context is a hyponym of the distance education, both of them were used interchangeably.

To elucidate this point, F. Saba's (2013) own reading of the literature, for example, brought him to the inevitable conclusion that most researchers, reviewers, and authors of articles in DE not only lacked proper knowledge of both the conceptual growth and the historical origin of the field (which tends to grow once a person becomes interested in the topic), but some even went further by attempting to narrow and reduce the view of the whole field to a specific "mode of information transmission (e.g., online learning)" or "medium of communication (e.g., web-based learning)." Such a "reductionist approach to understanding the field is in sharp contrast to the holistic systems approach that is essential for developing a comprehensive theory of the field" (p. 50). Some researchers (e.g., Black, 2013; Diehl, 2013; Kanuka & Conrad, 2003), through intensive reading of the literature, have tried to compile the generations through which the field has gone to reach its current state. For instance, that DE has passed through several stages was shown by Kanuka and Conrad. In the first stage, instruction was carried out only by post. The second stage was conducted through two forms of media: radio and television. The third stage involved an innovative way of structuring education: open universities. The fourth stage, which took place in the 1980s, utilized advances in technology to deliver real time interaction through audio/visual teleconferencing accessed by telephone, satellite, cable, and computer networks. The current stage is totally online and students attend virtual classrooms and universities worldwide.

2.3.2.2.1 Definitions and types of OLC.

With regard to the other issues related to DE—that is, multiple semi-synonymous constructs and the lack of empirical validity—F. Saba (2013) asserted that these issues were the consequences of advancements in technologies. Many people perceive, for example, BLC to be an alternative synonymous term for DE/OLC, even though the only factor shared by the two is technology. In BLC, technology is integrated in and/or with regular traditional classrooms, whereas, in DE, technology is the only method of teaching and learning outside campus. The core difference between the two terms is the reliance on or incorporation of technology with or without traditional on-campus classes for teaching and learning. In their discussion of which term to use to identify the field, Moore and Kearsley (2012) and Kanuka and Conrad (2003) discussed the issue of multiple terms. Of all terms, it was decided that Distance Education should be the name of the field, since it is a loose construct that includes other specific terms, as well as terms that may arise in the future. Moore and Kearsley further justified their suggestion by saying that the DE concept is superior to others because it "is a

multiple dimensional [one]; it is a pedagogy different from that of the classroom; it has a long history…; and it has distinctive organizational forms" (p. 3).

Both Moore and Kearsley (2012) and Kanuka and Conrad (2003) deduced that, in the field, many existing terms are believed to be problematic for one or more reasons. Such terms include distance learning, e-learning (i.e., computer programs/applications including virtual classrooms, computer-based learning, digital collaboration, and web-based learning), online learning (i.e., technologies that depend on the Internet, extranets, or intranets), synchronous (i.e., communication occurring at the same time through the use of computer programs), and asynchronous (i.e., communication occurring at different times through the use of computer programs). Some other terms include distributed learning (a new and radical path for education that incorporates flexibility; new learning techniques; created, modified, and new learning resources; and modified best traditional teaching practices), flexibility learning, and open and distance learning (i.e., the ability to learn whatever and whenever using technology). Finally, there is also correspondence education (via mail), tele-education (via TV and, to some extent, radio), and home study (Kanuka & Conrad, 2003; Moore & Kearsley, 2012).

The latter authors also stated that there were two issues related to these terms: the fact that they could indicate learning that only occurs from home and not from other places (such as work), and that when a phrase contains the word 'learning,' it creates a student-centered domain (i.e., one that looks at the outcomes of DE from mainly one side: the students' side), thereby neglecting that learning is the result of teaching (i.e., the teachers' side). Some of the terms also suffered from the fact that they were driven by method of teaching/learning technologies that are now outdated or minimally used or that do not include other forms of fully fledged technologies (Kanuka & Conrad, 2003; Moore & Kearsley, 2012). Other terms were formed based on incorrect theories, such as the idea that learners "can distribute [their] own learning," as well as theories that "really [hold] no explanatory power in describing any kind of learning, and is inherently confusing" (Kanuka & Conrad, 2003, pp. 388-389).

Given the relative acceptance of the DE term, different interpretations have been proposed. For instance, Allen, Omori, Burrell, Mabry, and Timmerman (2013) defined DE "as instruction in which there is no expectation for the physical copresence of the learner and instructor" (p. 143; see also Kaya, 2012; Mehrotra, Hollister, & McGahey, 2001), whereas Kanuka and Conrad (2003) interpreted it as "the organizational apparatus and process of providing educational experiences to learners at a distance" (p. 388). However, in order to capture the multifaceted nature of the field, Moore and Kearsley (2012) defined DE as "teaching and planned learning in which teaching normally occurs in a different place from

learning, requiring communication through technologies as well as special institutional organization" (p. 2).

2.3.2.2.2 OLC's benefits, drawbacks, and challenges.

In teaching foreign languages, many of the OLC's benefits are said to be true for other types of contexts such as the BLC. Yet, to achieve them—especially the advantages that are only specified for the OLC—teachers have to go through some challenges. Singhal (1997) pointed out that the Internet has a lot of potential to be used for educational purposes—especially for EFL learners—as it comprises many innovative resources and technological tools that can be incorporated to explain the most difficult aspects of the language along with other advantages. For example, Hurd (2005) stressed that the use of asynchronous technologies by EFL learners would increase their time to focus on grammar and develop their linguistic accuracy and to create a special relationship between interaction and reflection, since such technologies allow students to pause and reflect while interacting. Synchronous tools, on the other hand, would make learners feel less stressed, and with peer support and the ability to send voice and video messages, students would most likely have positive attitudes towards learning, participate more, lessen their anxiety and sense of social isolation, increase motivation, and increase their responsibility of their own learning. Teng, Chen, Kinshuk, and Leo (2012) asserted that computer-mediated communication has helped learning and teaching "take place across geographical boundaries [and that the] online synchronous learning environment with cyber face-to-face features affords students the sense of learning together online" (p. 918). Arnold and Paulus (2010) maintained that the above-mentioned technology and any other forms of electronic communication tools (e.g., emails, chat, Facebook, and Ning) would provide L2 learners with genuine contextualized language input that is not restricted by the borders and boundaries of the traditional classrooms and would also allow them to effectively interact with either native speakers or other learners. The result of the latter would be the development of students' pragmatic, linguistic, and cultural competence. Ilter (2009) contended that not only do computers and other types of technologies provide learners with a sense of encouragement and freedom, but they also aid students to become involved, motivated, and active in language learning process. Pandya (2013) deduced that chatting with native speakers would allow L2 learners to exchange and practice new structures of the language and test their sociolinguistic appropriateness. It would also allow the latter to observe how native speakers interact, lessen their fears in learning the language, get them involved in learning it, and develop their understanding of the role of culture.

Yang and Chen (2007) reported that Internet-based teaching activities such as chatroom discussion, videoconferencing, group emailing, English home page designing, web-based course, and email writing programs applied to high school Taiwanese students who were part of "Advanced Joint English Teaching" (p. 860) resulted in more opportunities to experience new technologies, which promoted participants' learning possibilities by means of experiencing the pleasure of learning. Roux, Guzmán and González (2014) maintained that videoconferencing tools help students to access expert knowledge and input remotely. Wu, Yen, and Marek (2011) found that the application of the aforementioned tool with native speakers on Taiwanese EFL learners led to more motivation, confidence, and learning ability, and that long-term changes in the ability are attributed to enjoyment of the learning experience. They also found that what made learners more confident in what they learned, more comfortable in applying their skills, and more inspired to make global and cross-cultural connections was just a small amount of authentic interaction in the L2. Wang, Lefaiver, Wang, and Hunt (2011) cited many benefits of using multi-user virtual environments, such as Second Life, for EFL students. These benefits include the reinforcement and fostering of verbal communication within a social context; increasing effective, as well as active, interactive instruction and learning opportunities; allowing collaboration and creativity through discussions, field trips, presentations, guest lectures, and virtual classes; providing a context for culture and language immersion; reducing learners' tentativeness, which often appears in traditional classrooms; and finally, promoting the sense of fun.

Nevertheless, the application of OLC on EFL learners is not without its shortcomings, since the environments and its dependent technologies can have some weaknesses and challenges that instructors have to address. For instance, Buckley (1997) indicated that the absence of her image when teaching students online helped her teach them more effectively; however, Yamada (2009) contended that—along with the other drawbacks of OLC—the absence of EFL instructors could lead to two issues. One is the lack of immediate feedback. The other is the reduction of students' consciousness regarding the importance of their study, as well as the prioritization of their personal issues (since their courses can be completed via networked computers rather than through face-to-face classroom visits). Deveci (2015) reported that both Turkish EFL teachers and students' viewed the context as ineffective because it does not provide an atmosphere similar to the traditional classroom that students are accustomed to, because of students' lack of technological competence, and because of students' inability to become self-autonomous. Hurd (2005) asserted that both teachers and students may find this mode of instruction as "depersonalizing, fragmentary, and lacking the

humanity and intimacy that the face-to-face environment affords," because of the absence of paralinguistic elements (e.g., body language) and the overload of information (p. 13). For her, there is also a challenge in guiding learners into becoming dependent on themselves and in training them to use the technologies properly, as with inadequate guidance the reverse can occur. Singhal (1997) emphasized some issues such as the fact that teachers have little or no experience in incorporating technologies or even working in such a context, the lack of censorship that allows students to access any website, and lastly, the fact that some learners located in more isolated rural areas have limited or no access to the Internet. Roux et al. (2014) added three more challenges surrounding the context: the unavailability of alternatives for learners once the technology adopted fails, limiting the autonomy of learners to limited resources, and developing only social interaction rather than cognitive interaction, which is fundamental for promoting learning. Therefore, Zahedi and Dorrimanesh (2008) recommended EFL instructors in OLC prepare, educate, and provide opportunities for their learners to practice metacognitive learning strategies, such as "planning, monitoring, and evaluating" (p. 164). The mastery, as they believed, of these strategies in terms of when and where to use them and how to coordinate them with one another would maximize the benefits and lead to students becoming independent and effective learners.

Regarding writing skills, Miller (2001) noted that given the expansion of OLC in the field of education, some composition specialists call for critical investigation of modern technologies to determine their limitations and potential abilities in improving such skills. Yang (2015) believed that "the big data era, [availability on Internet and web-based appliances], provides teaching writing to EFL [distance] college students with abundant resources, technology-based teaching aids and new ideas" (p. 777). Singhal (1997) stated that email exchange with native speakers and learners from other countries would encourage learners to use computers in authentic realistic situations that would contribute to the stimulation and development of thinking and communication skills in persuading, arguing, and defining their points. Email would also get them involved in various types of profitable interactions, promote writing; generate ideas about topics, encourage discussion and writing freely, and facilitate learning about L2 culture, history, economy, geography, and politics. Despite the few studies that determined the effectiveness of teaching the skill remotely and its procedures, OLC still suffers some major issues. For example, Al-Abed Al-Haq and Al-Sobh (2010) concluded that the pre-university EFL learners taught by a means of a web-based writing program outscored those students who were instructed on campus. Hayati and Gooran (2014) examined the effect of teaching how to write a paragraph in a classroom and via email

and found that there was a significant positive correlation between incorporating email and Persian EFL learners' writing production. In addition, Alzu'bi and Sabha (2013) found that mobile-based email teaching led to significant improvements of both Jordanian EFL students' writing skills and vocabulary acquisition. Sayadi and Khoshsima (2016) contended that the experimental EFL learners who received their writing skill instructions on PowerPoint via emails and those who were instructed in traditional classrooms developed significantly in the skill, but the development of the former group was more than the latter group. Xiao's (2008) results showed that paired peer review in distance-taught EFL writing courses had beneficial effects on developing Chinese learners' composition, promoting their transferable skills—collaboration, problem-solving, and self-management skills—and self-efficacy.

However, Gillam and Wooden (2013) emphasized that in OLC, the very basic elements of writing pedagogy, such as immediate feedback, comments, and suggestions on students' writing, are still missing and that the mastery of content is assessed through simplified assignments or "user-friendly testing modules" (p. 27). Furthermore, both Blakelock and Smith (2006) and Sapp and Simon (2005) have argued that despite gaining prominence in writing pedagogy over the last decade, OLC still causes concern and skepticism for writing specialists. These authors maintained that administrators and writing teachers still question the reliability and validity of teaching writing to students remotely. This hesitation has likely led to at least some abandonment of such a method. They also stated that administrators and writing instructors do not know the significance of OLC for solving problems related to teaching composition, such as insufficient classroom space, lagging enrollment, and dwindling resources. The challenges in implementing OLC for writing courses emerge from the fact that writing instructors—who tend to endorse such types of teaching—are not willing to undertake such measures by themselves. These issues are related to the fact that writing instructors lack sufficient knowledge of the workloads of online courses and incorporated technologies and are afraid that OLC may negatively affect their teaching abilities. Many also fear the lack of control over classroom student size; the difficulty in commenting on, correcting, and grading students' writing assignments; the fact that the incentives and rewards of teaching such classes are not sufficient; the problems that may arise with course delivery (i.e., Internet systems); the possibility that OLC will leave students with unfinished learning objectives and decrease their engagement in the learning process; and, finally, that both students and teachers will have difficulty expressing themselves affectively and appropriately online (Blakelock & Smith, 2006; Sapp & Simon, 2005).

2.3.2.2.3 OLC in the Arab countries and in Saudi Arabia.

The Arab countries in general are classified to be "later adopters of" OLC (Alsunbul, 2002, p. 54). However, the stages through which the field has gone are similar to those experienced by Western OLC adopters. In other words, according to Alsunbul (2002), the Arab world underwent four phases of OLC developments. The first stage was the correspondence study, in which printed materials were sent to students via post. The second stage involved printed materials, together with alternative teaching media (i.e., through television and radio broadcasting). The third was the "inauguration of Al-Quds Open University" and, later, some similar universities (e.g., Arab Open University in most Arab countries and in Saudi Arabia in particular)—all of which incorporated multimedia, such as computers, books, CD-ROMs, blogs, the Internet, and audio and video cassettes (p. 66). Finally, the establishment of the Arab Network for OLC (ANODED), which comprises and supervises 60 institutions working directly or indirectly on OLC, represents the fourth stage. The development of OLC in the Arab world is slow for the following reasons: teachers and students had no knowledge of it, teachers' and students' experiences are based on the catastrophic and unsatisfactory outcomes of students enrolled in the first phase of OLC, and there is a current lack of technological and student-oriented support and updates. There are also other reasons, such as the lack of orientation towards using technology, the design of materials based on traditional, in-class instruction, subsidiaries, language accreditation, licensure, quality assurance, students' habits, cost consideration, and workloads (Alsunbul, 2002).

According to Al-Khalifa (2009), Saudi Arabia has been slower than many countries in implementing OLC. Al-Harbi (2011, p. 45) stated that "e-learning [there] is still in its infancy with a paucity of information on its use." Even knowing that the OLC medium has gone through the very same generation experienced in other Arab countries, the statements of the previous researchers are not accurate in the case of Saudi Arabia because the education there is segregated—women are separated from men in schools (Aljabre, 2012). Since the establishment of the first female college in the 1970s, the college was equipped with closed-circuit TVs to allow male lecturers teach female students (Al Ebaikan & Troudi, 2014). This indicates that the principle of OLC in the Kingdom of Saudi Arabia (KSA) was launched even before blended learning. Al-Shehri (2010) maintained that in the last five years KSA has witnessed a revolutionary growth and interest in OLC. This interest is observable in the remarkable projects initiated for both students and teachers in a variety of educational cycles, including higher education, undergraduate education, pre-universities, and vocational

institutions. For example, in 2005—in response to the objective of transferring education to the digital world—the *National Center for E-learning and Distance Education* was established (Al-Khalifa, 2009; Yamani, 2014).

In K–12 private and public schools, as elucidated by Al-Asmari and Rabb Khan (2014), the Ministry of Higher Education (MOHE) established and launched many projects for the purpose of introducing the new genre of electronic education. For instance, it invented and launched the Computer and Information Center (CIC), computer projects, and the Google Educational Program, all of which aimed to ensure all schools had Information and Instructional Technology (ICT) services that would allow students and teachers to access personal websites, emails, and office applications. The MOHE also signed contracts with Intel and Microsoft to provide teachers and students with e-learning orientation courses, training, and more. Furthermore, the launch of WATANI Schools' Net project aimed to link all educational directorates and schools through a wide area network. A local software company called Semanoor, in collaboration with Intel, transferred all school textbooks to electronic versions while launching a variety of tools, such as electronic classes, a multimedia library, and the Semanoor Browser. Similarly, in collaboration with Intel, Obeikan Education released a Skool website, which included interactive science and mathematics lessons for pre-university students. Finally, the Jehazi project sought to increase instructors' technological awareness by providing them with laptops (Al-Asmari & Rabb Khan, 2014).

In higher education, the inclusion of e-learning and the establishment of a Deanship and Faculty of e-learning have been advocated in many Saudi universities, such as Al-Imam Muhammad Ibn Saud Islamic University, King Saud University, King Khalid University, and others (Al-Asmari & Rabb Khan, 2014; Al-Dosari, 2011; Al-Khalifa, 2009). Moreover, the first Saudi Electronic University has been launched (Saudi Electronic University, 2011). Although the very basic element of OLC (the Internet) is available and its advantages related to evolving learners' competence has been identified, the use of this OLC is still facing resistance from certain teachers and students. However, Simsim (2011) noted that ever since the public was allowed access to the Internet in KSA, which occurred around 1999, there has been a steady increase of users. For example, it was estimated that by the end of June 2016 there were 20.813.695 million Internet users in the KSA, with Internet service penetration of 64.7% (Miniwatts Marketing Group, 2016). Some educators and researchers (e.g., Al-Dosari, 2011; Al-Fahad, 2009; Al-Harbi, 2011; Al-Harthi, 2005; Ali, Sait, & Al-Tawil, 2003; Alswaier, Youssef, & Emam 2012; Altameem, 2013; Altowjry, 2005; Yamani, 2014) view OLC as a necessity, since it can help students from various educational cycle levels with their

in-class subjects, can increase enrollment (specifically in Saudi Arabia, whose own universities—combined with scholarships to study abroad—are still insufficient), and can reduce students' social embarrassment, since demographic features can remain anonymous. Alaugab (2007) maintained that "online learning goes beyond barriers of time, location, and culture and has created many opportunities for learners and instructors. Learners can learn anytime, anywhere, regardless of gender, age, geographical location and culture" (p. 1; cf. Hamdan, 2014).

Ali et al. (2003) reported the reasons Saudi students may take courses on the Internet: preference for using computers (41.3%), preference for learning from one's own place (29.7%), promotion (29.1%), fun or experience (28.5%), the unavailability of courses in schools (18.8%), and, lastly, the need to earn a higher degree (19%) (cf. Al-Asmari & Rabb Khan, 2014). However, some researchers mentioned the problems that may explain students' and teachers' lack of interest in becoming involved with online courses. For example, despite what people would anticipate, a lack of Internet experience was not correlated with students' intention to use e-learning; however, enjoyment, computer self-efficacy, and anxiety were (Alenezi, Abdul Karim, & Veloo, 2010). Other factors include as a lack of knowledge regarding the essence of OLC (Ali et al., 2003), a lack of experience and technical skills (Altameem, 2013), English proficiency level, cultural barriers (i.e., not being accustomed to engaging with members of the other gender from the same nationality), time consumption, and most importantly the fact that OLC courses—even when they are only taken as a part of a qualification (undergraduate or postgraduate)—are still not accredited (Al-Harthi, 2005). Mirza (2007) explained this situation by saying that "you cannot qualify with that degree to a governmental job, and you cannot pursue graduate-level education in Saudi Arabia based on your internationally earned degree" (p. 1).

Finally, Alaugab (2007) asserted that the "implementation of online instruction in Saudi Arabia has been very limited" (p. 10). This is quite correct, especially knowing that there is only one university (King Faisal) that offers a bachelor's degree in English language via distance education. Moreover, there are only four studies that were conducted to determine the success of the aforementioned program as well as the effect of remote teaching of the writing skills on Saudi EFL learners. For instance, Dokhykh (2014) found that the program has many advantages, such as not being different from traditional in-campus workload, materials, textbooks, and interactions; possessing flexibility in terms of students being able to study whenever and whatever they want, which saves both energy and efforts; and the incorporation of a variety of learning aids, supporting materials, and practice. The only

disadvantage was the fact that there are no other alternatives when Internet connection is lost. Ahmed (2013) investigated the impact of teaching composition skills via email and concluded that the experiment group of Saudi participants outperformed their control counterparts in sentence construction, text organization, paragraphing, and style. Abdul Fattah (2015) maintained that his Saudi EFL undergraduates, taught via WhatsApp program, developed significantly in writing punctuation and vocabulary and outperformed undergraduates instructed in a traditional classroom in punctuation and sentence structure. Alhassan and Gashan (2013) emphasized that Saudi EFL university students who were taught mainly through WebQuest writing instructions improved significantly in text length, grammar, and vocabulary, and outscored their counterparts who received traditional writing instructions. Unfortunately, the context of KSA is still suffering from another issue concerning the scarcity of research investigating other, alternative types of composition assessments, such as Complexity, Accuracy, and Fluency (CAF). All of these issues will be addressed thoroughly in the second part of the literature review.

2.4 Summary

This chapter discussed in detail the challenges and issues confronting the teaching of English writing skills to EFL learners in general and to Saudi EFL students in particular. These challenges and issues stemmed from the multifaceted nature and components of the skill, which led scholars and specialists to propose various types of proposed definitions and characteristics and invent and adopt different teaching and learning approaches and environments (TLC, BLC, and OLC). Although the basic elements (i.e., Internet and technologies) of the two latter learning contexts are available, they are rarely implemented when teaching English and its four skills in Saudi Arabia.

Chapter Two

Literature Review Part II: Complexity, Accuracy, and Fluency

2.5 The Origin of Complexity, Accuracy and Fluency (CAF)

The individual elements of CAF went through a long process before they were combined into one triad and were considered to be potentially capable of illuminating language use and measuring language development and proficiency (Rausch, 2012; Towell, 2012). Historically, most researchers who traced CAF's origins stated that it emerged shortly after the birth of SLA in the 1970s (Housen, Kuiken, & Vedder, 2012; Larsen-Freeman, 2009). At that time, one of the areas that SLA researchers had disputes about concerned the level of L1 influence on an L2 process, "And how much of it was due to a 'creative construction process common to all learners'" (Larsen-Freeman, 2009, p. 579). For example, many authors (e.g., Bailey, Madden, & Krashen, 1974; Dulay & Burt, 1973, 1974a, 1975; Krashen, 1982) reported a minimal effect of the L1 on the L2 of adults and children implicitly. To clarify this, in a series of studies Dulay and Burt followed Brown (1973), whose longitudinal work showed that English native-speaking children's syntactic morphemes emerged in a consistent order of 14 functions. What Dulay and Burt found when applying 8, 11, or 13 of Brown's morphemes was that whether their investigation focused on participants with different English L2 proficiency levels with the same native language—Spanish (e.g., 1973)—or participants with different English L2 proficiency levels with different mother tongues—Spanish and Chinese (e.g., 1974a, 1975)—the result remained the same: there were salient similarities between all groups in the grammatical acquisition sequences. In a more explicit way, Dulay and Burt (1974b) extended their research by examining "513 unambiguous [English oral syntactic] errors" of 179 Spanish L1 children (p. 131). The aim was to determine how many of these errors could be attributed to L1 interference and/or cognitive strategies development. The findings revealed that only 4.7% of these errors were linked to the former, while 87.1% of them were because of the latter.

Such a claim caused significant controversy, since other studies (e.g., Dušková, 1969; Hakuta, 1976; LoCoco, 1975) showed significant impact of the L1 on the L2. However, what quickly became evident was that the discrepancies in the results between these studies were due to the fact that researchers were using different benchmarks (Larsen-Freeman, 2009), different tasks and instruments (Larsen-Freeman, 1975), and were conducting various research designs, such as cross-sectional and longitudinal designs (Ellis, 1994; McLaughlin, 1987). There were also other reasons why the effect of the L2 would vary in how similar it

was to the L1 (the more similar the two languages are, the lesser degree of impact will be observed); in L2 learners' proficiency and development levels (McLaughlin, 1987; Larsen-Freeman, 1978a); and in learning methods, communication strategies, performance (McLaughlin, 1987), or even their age (Mitchell & Myles, 2004).

In the mid-seventies, because of the issues mentioned above, many researchers acknowledged the lack of proper and suitable L2 developmental yardsticks and called for the construction of means by which such development could be measured, "Apart from the use of lengthy standardized proficiency tests, which serve other purposes" (Larsen-Freeman, 2009, p. 579; Larsen-Freeman, 1978a, 1978b; Larsen-Freeman & Storm, 1977). Hakuta (1975a, 1976), for example, was the first researcher to state that the field was not only "lacking any strong theoretical framework," but it also suffered from "a problem which plagues (and will probably continue to plague for some time) [SLA] research: the lack of an index of development, such as Mean Length of Utterance [MLU] in [L1] research" (1976, pp. 345-348). He elsewhere (1983) maintained that "there [were] still severe problems with measurement and appropriate testing [and] the first step in this area must be theoretical refinement followed by operationalization of the individual dimensions" (p. 50). Regarding L2 theory alternations, it must be remembered that in some facets, language development and proficiency are not the same, though development is part of proficiency. Two such differences can be found in their standards of measurement and in the language areas that are their foci. Wolfe, Quintero, Inagaki, and Kim (1998) mentioned these differences and stated that while language development includes elements of a learner's production that show incidents of progress—such as the number of clauses per T-unit (a number predicted to increase as development improves)—language proficiency is a more overall construct that separates language learners into clusters according to their language capabilities along a normal distribution. Therefore, any test of language proficiency would include an overall estimate of the learner's language proficiency established by adding a wider number of theoretically motivated linguistic categories.

For the L2 development index, Larsen-Freeman (1978b) believed that researchers started "to make some inroads in [the] task of constructing" such an index (p. 447). Eventually, they became more meticulous towards the goals of an index, the methods and conditions by which it can be achieved and regarded as effective, and its prospective benefits. For instance, Larsen-Freeman (1983) stated that the purpose was to find "a developmental yardstick" that meets the "criteria of objectivity, reliability, ease of application, and universality" (pp. 287-288). The index should be readily available regardless of the

instrument used in the research and could be used even if the data has already been collected. It should also work perfectly for various L2s (Larsen-Freeman, 1983) "in a nonlanguage specific way" (Larsen-Freeman, 1978a, p. 134) and in reflecting all learners' overall proficiency, regardless of their L1s, educational backgrounds, ages, or having been analyzed individually or as groups (Larsen-Freeman, 1978b, 1983). Larsen-Freeman (1978b) adduced that as learners progressed towards full internalization of a TL, such a yardstick ought to "allow us to give a numerical value to different points along a second language developmental continuum—numerical values which would be correlates of the developmental process and would increase uniformly and linearly" (p. 440). However, in order to create such an index, Larsen-Freeman (1983, p. 301) recommended that researchers go through a procedure of two steps to identify the best measure(s). The first step would be deducing the "gross estimate of a learner's overall L2 proficiency" (i.e., whether he/she is at a beginner, intermediate, or advanced level) by incorporating one measure and then deducing a more precise developmental level identification by using level-appropriate measures. The metric used in the latter step might vary based on the category given to learners initially. The prospective benefits, as Wolfe-Quintero et al. (1998) posited, were twofold. Not only could the index be used in situations where L2 researchers, language program administrators, and teachers lack access to language proficiency tests, but it would also lead research studies to produce more exact and accurate descriptions of learners' development level(s) beyond the usual practice of using ambiguous terms like intermediate (Larsen-Freeman, 2009, p. 580). Consequently, the comparability between such studies would be greater, perhaps, even between those that address developmental levels in various target languages (cf. Dewaele, 1998; Engber, 1995; Hyltenstam, 1977; Meisel, Clahsen, & Pienemann, 1981). Also, it would at least make the authors capable of knowing whether other studies' participants are comparable to theirs or not (Larsen-Freeman, 1978a; Larsen-Freeman, 2009; Larsen-Freeman & Storm, 1977). For Larsen-Freeman (1983), the advantages of the yardstick would be in aiding administrators to be capable of specifying the L2 learners' developmental/proficiency levels precisely and adequately and then assigning them to their designated classes. Teachers, as well, will have a reliable method to capture the fluctuations of their students' overall proficiency/development over the duration of the course.

Thus, with the intention to develop such an independent index, L2 researchers turned to L1 researchers' studies for metrics (Ellis, 1994, 2008; Housen et al., 2012; Larsen-Freeman, 2009). The reasons could be to find "a developmental yardstick against which global (i.e., not skill nor item specific)" L2 proficiency could be captured (Larsen-Freeman, 1983, p. 287) and

characterized as reliable, objective, accessible (i.e., availability of use "with a minimum of time or expense"), and instrument-free (i.e., leading to consistent results if used for the same participant across various testing situations, and texts [Harrington, 1986, p. 49; Ellis & Barkhuizen, 2005; Hakuta, 1976, 1983; Iwashita, Brown, McNamara, & O'Hagan, 2008; Larsen-Freeman, 1978a, 1983; Polio, 1997, 2001; Wolfe-Quintero et al., 1998]). For Ellis (1994, 2008), such a decision was "a good starting point," since the latter can provide the former with "useful methodological procedures" and "a baseline" of metrics for calibrating developmental and acquisitional sequences, patterns, and order (1994, p. 76). For instance, in the mid-1960s and early 1970s, grammatical complexity and accuracy measures (e.g., MLU, T-unit) in L1 were developed (e.g., Brown, 1973; Hunt, 1965, 1970a) and adopted by L2 researchers (Housen et al., 2012; Larsen-Freeman, 2009). As discussed by Larsen-Freeman (2009), Hakuta (1976) took the same steps as the L1 acquisition investigators (e.g., Brown, 1973) in recording longitudinally the English development of a Japanese child called Uguisu, or "'nightingale' in Japanese" (Hakuta, 1976, p. 321 [see also Hakuta, 1974, 1975a, 1975b]). Despite Brown and his associate students' usage of the MLU as a development index (Larsen-Freeman, 2009), the MLU might not be exact or completely efficient (Larsen-Freeman & Long, 1997; Larsen-Freeman & Storm, 1977), as it captures the development in question partially (Larsen-Freeman, 2009). Even worse is the fact that its application in L2 development research is problematic, since adult learners are described as more taxingly mature learners who depend on formulaic utterances and who are able to—even after few hours of exposure to the L2—produce longer utterances (Larsen-Freeman, 1978a, 2009; Larsen-Freeman & Long, 1997; Larsen-Freeman & Storm, 1977).

On the other hand, Larsen-Freeman (2009) stated that she was one of the first to call for an L2 developmental index (1976/1978a) and followed Hunt (1965, 1970a) in a group of her studies (Larsen-Freeman, 1978b, 1983; Larsen-Freeman & Storm, 1977). Hunt's goal was to determine the metrics by which L1 and L2 writing development can be captured. He used the "terminable unit," defined as "any statement containing one main clause, with or without subordinate clauses, to be punctuated with terminal marks at both ends: a capital at one end and a period or question mark at the end" (1970b, pp. 198-199). The justification for his use of T-units instead of sentence length was of the established facts that English native speaking schoolchildren tend to embed a great amount of "sentence constituents" (p. 195). In Larsen-Freeman's coauthored studies, she incorporated various measures of the T-unit (e.g., percentage of error-free T-units, average length of error-free T-units, number of T-units, and others). The results of these studies were somehow inconsistent (Larsen-Freeman, 1983) and

disappointing when the metrics were applied to spoken language (Larsen-Freeman & Long, 1997). In other words, Larsen-Freeman (1983) maintained that the very same metric (i.e., percentage of error-free T-units) that discriminated between participants' proficiency levels in one study (e.g., 1978b) failed to do so in another study (1983). Also, the total number of words per error-free T-unit metric differentiated between some adjacent proficiency levels: level one and three, but not between level one and two in the oral data of the 1983 study. In some cases, this metric could discriminate between levels with small numbers of participants such as the three groups of eight participants who performed picture composition task, but not between the three groups of eight participants who undertook the oral story-retelling task. Based on the above studies, Larsen-Freeman (2009) concluded that it was substantiated that error-free T-unit average length was an acceptable metric to discriminate between L2 developmental levels "at least at the group level and at least for English as a second language" (p. 580). For Vann (1979), the measure is the best indicator of the adult L2 learners' proficiency in writing. Later, it was proved that such a measure could be incorporated when analyzing oral or written English or other L2 languages by themselves, such as Chinese (Jiang, 2013), Japanese (Loschky, 1994; Stockwell & Harrington, 2003), and English and Japanese (Iwashita, 2010). However, because complexity, accuracy, and fluency are conflated by the previously mentioned measure, research studies had moved beyond only investigating accuracy to include complexity and fluency by the 1980s (Larsen-Freeman, 2009).

According to Housen and Kuiken (2009), the CAF triad originated through the research of L2 pedagogy. In the 1980s, researchers (e.g., Brumfit, 1979, 1984; Hammerly, 1990) made a basic distinction between L2 speech fluency and L2 usage accuracy for the purpose of studying the development of L2 oral proficiency in classroom contexts (Housen & Kuiken, 2009; Housen et al., 2012). For Housen and Kuiken (2009), Brumfit (1984, see also 1979) is considered to be one of the first to use this dichotomy when he contrasted fluency and accuracy pedagogically. For instance, he differentiated between activities that are fluency oriented and ones that are accuracy oriented (Housen & Kuiken, 2009). The former will promote L2 spontaneous oral production, while the latter will emphasize the L2 controlled production of grammatically correct syntactic structures. In other words, Wolfe-Quintero et al. (1998) believed that in Brumfit's opinion, fluency is the natural automatic and implicit language use for communication, whereas accuracy is the language use explicitly displayed for evaluation. The crucial issue for him was whether classroom focus is specified for communicating a message to achieve fluency, or for learning accurate L2 forms to achieve

accuracy. Thus, accuracy—rather than being the result of an unconscious restructuring process—is viewed as a result of a conscious focus on form (Wolfe-Quintero et al., 1998).

In the 1990s, complexity, the third norm of the triad, was added as a consequence of the L2 model proposed by Skehan in 1996 (1996a) and 1998 (Housen et al., 2012). In this model, the author used CAF for the first time as the three fundamental dimensions of proficiency (Housen & Kuiken 2009; Housen et al., 2012). Also working definitions of the three constructs were established at that time and have been used ever since (Housen & Kuiken, 2009; Housen et al., 2012; Lennon, 2000; Wolfe-Quintero et al., 1998).

From the 1990s onward, the three constructs have appeared prominently and predominantly—most often together—as dependent variables in SLA research and are viewed as the evaluated features of learners' performance to examine the impact of other factors (Housen & Kuiken, 2009; Housen et al., 2012). These factors were investigated separately or in a combination of two or more, as exemplified in Table 2.

Table 2

Factors Affecting/not CAF Research Studies

Factors	Research Studies
Age	Celaya, Torras, and Pérez-Vidal (2001); Miralpeix (2006); Navés,Torras, and Celaya (2003); Torras And Celaya (2001); Torras, Navés, Celaya, and Pérez-Vidal (2006)
Teaching Instructions, Learning Environments (Abroad/at Home)	Baró and Serrano (2011); Dubiner, Freed, and Segalowitz (2006); Freed, Segalowitz, and Dewey (2004); Freed, So, and Lazar (2003); Housen et al. (2011); Lord (2009); Mora and Valls-Ferrer (2012); Pérez-Vidal and Juan-Garau (2009); Sasaki (2004, 2007); Segalowitz and Freed (2004); Tavakoli, Dastjerdi, and Esteki (2011)
Feedback	Chu (2011); Grami (2005); Rahimi and Dastjerdi (2012); Robb, Ross, and Shortreed (1986)
Proficiency Levels and Program Types	Serrano (2011)
Task Features/Types	Armstrong (2010); Bonzo (2008); Dickinson (2014); Ferreira (2013); Pourdana and Behbahani (2011); Rezazadeh, Tavakoli, and Eslami Rasekh (2011); Skehan (1996a, 1996b); Skehan and Foster (1997); Sponseller and Wilkins (2015)
Task Complexity Manipulation, Proficiency Level, and/or Planning Types	Abdollahzadeh and Kashani (2011); Al-Humaidi (2008); Ellis (2009); Ellis and Yuan (2004); Farahani and Meraji (2011); Foster and Skehan (1996,1999); Ghavamnia, Tavakoli, and Esteki (2013); Ishikawa (2006, 2007); Khorasani, Pandian, and Ismail (2012); Kuiken and Vedder (2007a, 2007b, 2008, 2011); Kuiken, Mos, and Vedder (2005); Mohammadzadeh Mohammadabadi, Dabaghi, and Tavakoli (2013); Nikou and Eskandarsefat (2012); Ong and Zhang (2010); Piri, Barati, and Ketabi (2012); Seyyedi, Ismail, Orang, and Nejad (2013); Tavakoli and Skehan (2005)
Collaborative and Individual Writing	Dobao (2012); Elola and Oskoz (2010); Jafari and Ansari (2012); Jong (2009); Nassaji and Tian (2010); Pae (2011); Storch (2005); Tavakoli and Rezazadeh (2014); Wigglesworth and Storch (2009); Zabihi, Rezazadeh, and Dastjerdi (2013)

From the plethora of research mentioned above, CAF appeared to be "distinct components" with the potential to measure L2 performance and proficiency separately and the ability to develop variably by various types of participants under various types of learning conditions and manifest distinctively and variably under various L2 use conditions (Housen & Kuiken, 2009, p. 462).

From the mid-1990s onwards, because of the advancement in psycholinguistic and cognitive psychology, CAF has been featured as "the primary foci or even as the independent variables of investigation in SLA" (Housen & Kuiken, 2009, p. 462). Examples include the studies of Ågren, Granfeldt, and Schlyter (2012); Ellis and Barkhuizen (2005); Ferrari (2012); Foster and Skehan (1996, 1999); Gunnarsson (2012); Lennon (1990); Norris and Ortega (2009), Polio (1997, 2001); Robinson (2001a, 2001b, 2003, 2005, 2007, 2011a, 2011b, 2015); Skehan (1996b, 2003a, 2003b, 2009, 2013); Skehan and Foster (1997, 1999; 2001, 2005, 2012); Thornbury (2000); Towell (2012); Wolf-Quintero et al. (1998). In most of these studies, CAF was featured as the major phenomenon of the psycholinguistic mechanisms and processes underlying the L2 system/knowledge representation, internalization, and processing (Housen & Kuiken, 2009, Housen et al., 2012). For Housen et al. (2012), CAF has received the status of being the major, distinct dimensions of L2 proficiency and performance and has been vindicated theoretically and empirically. Theoretically, it is thought that the three following major phases of change in underlying L2 systems are typified by CAF constructs: (i) internalization of additional L2 structures or more complexity, as shown by the development of more complicated and refined L2 knowledge systems; (ii) the alteration of the learners' L2 ability, as indicated by their reorganizing and adjusting their L2 knowledge (this would include the deviant or non-target-like elements of their interlanguage [IL] so that the learners display more accuracy and more complex structures); and (iii) the incorporation and arrangement of L2 knowledge, which provides more performance control (fluency) as a result of routinization, lexicalization, and automatization of more complex L2 components.

Many researchers have applied different factor analyses, such as exploratory, confirmatory, Varimax, or even simple correlation (i.e., Pearson) on CAF constructs and their measures (e.g., Deane, 2014; Koizumi & In'nami, 2014; Kuiken, Vedder & Gilabert, 2010; Mojavezi, 2014; Mota, 2003; Puranik, Lombardino & Altmann, 2008; Robb et al., 1986; Sakuragi, 2011; Skehan & Foster, 2005; Tavakoli & Skehan, 2005; Vercellotti, 2012; Wagner et al., 2011; Yang, 2014). The purpose was to determine to what extent these metrics truly gauge their specified CAF components, how many of them are redundant, their correlations between each other, and their underlying structures (Koizumi & In'nami, 2014; Norris &

Ortega, 2009; Wolfe-Quintero et al., 1998). Some of the empirical evidence reported by studies that used the factor analyses maintained explicitly that the above-mentioned theoretical part was valid, that CAF constructs are identified to be competing and distinctive areas of L2 production, and implicitly that before general claims about L2 production and proficiency can be made, all the three constructs must be considered (Housen et al., 2012).

2.6 CAF Research Challenges

CAF has attracted the interest of many linguists, researchers, and practitioners and has had widespread application in L2 research. Still, none of the three constructs are uncontroversial, and many questions and issues still remain unanswered and unresolved. As a result, in the search for their refinement, some researchers (e.g., Bulté & Housen, 2012; Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Larsen-Freeman, 2009; Norris & Ortega, 2009, Pallotti, 2009, Polio, 1997, 2001) have highlighted several problems that should be addressed in future studies. Housen et al. (2012) believed that these problems stem from five areas:

> (i) the definition of CAF as scientific constructs, (ii) the nature of their linguistic correlates and cognitive underpinnings, (iii) their connections and interdependency in both L2 performance and L2 development, (iv) their empirical operationalization and measurement, and (v) the factors that affect the manifestation and development of CAF in L2 use and learning. (p. 3)

However, because of the correlation between all of CAF problem sources and the fact that one source of a problem can lead to another source of a problem, they were divided into two major categories: (a) the challenges in defining and measuring CAF constructs; (b) the challenges in determining their linguistic and psycholinguistic correspondences, factors, and correlations.

2.7 The Challenges in Defining and Measuring CAF Constructs.

Despite the manifold research in CAF, two of the questions that remain unanswered are concerned with how to interpret CAF constructs and how to reliably, efficiently, and validly gauge them (Bulté & Housen, 2012; Ellis & Barkhuizen, 2005; Housen & Kuiken, 2009; Housen et al., 2012; Norris & Ortega, 2009; Pallotti, 2009; Polio, 1997, 2001; Skehan, 1996a, 1996b, 2009; Wolfe-Quintero et al., 1998). Norris and Ortega noted that "researchers have not done sufficient thinking about what [they] are measuring and why" (p. 560). Admittedly, they, together with other researchers (e.g., Housen & Kuiken, 2009; Wolfe-Quintero et al., 1998), stated that various types of interpretations and definitions coexist because these constructs are multidimensional and multicomponential and cannot be explained in one single

definition. Nevertheless, most—if not all—of these definitions are also problematic for other reasons. First, they are not supported by linguistic theories and language learning (Housen & Kuiken, 2009). Second, they cannot be used interchangeably in two different modes: speaking and writing (Abdel Latif, 2013).

Some authors have maintained that many L2 studies examine CAF without any specification as to what these constructs mean. Bulté and Housen (2012) and Housen et al. (2012) noticed that even when some studies do so, they define them implicitly or explicitly but with vague, general, or even circular terms, or without any indication of the instruments by which such constructs can be captured. Pallotti (2009) deduced that there are two broader issues of the measurements. The first is that when researchers, scholars, and educators conducted studies, they would most likely try to use or specify the "best measures" (p. 590) that will aid them in describing language performances or levels of language acquisition (e.g., Knoch, 2008). For them, these measures are called such because they indicate differences between learners "both over time and across tasks, correlating with other equally varying proficiency measures"(Pallotti, 2009, p. 590). The issue here is that these researchers are falling victim to "the necessary variation fallacy," in the sense that some of them consider the measures that show similarities and constants and do not change over time to be uninformative, invalid, or poor ones, although they clearly indicate that the features they try to capture do not actually vary between learners (pp. 590-591). The second issue regards using many variables to compare groups of participants for the sake of having "significant results" (Pallotti, 2009, p. 591). The problem here arises when out of the number of measures used, only one shows significant differences among these groups of subjects. Such a result would indicate that the researcher(s) is dealing with groups that come from different populations who are similar in everything except one characteristic.

For Housen and Kuiken (2009) and Norris and Ortega (2009), CAF has been evaluated across many language-related domains by a range of various tools—subjective, objective, and holistic quantitative measures. The reason for such measurements of particular and general linguistic properties of second language performance is to determine "more precise, objective accounts of an L2 learner's level with each (sub)dimension of proficiency" (Housen & Kuiken, 2009, p. 464). The problem here is in deciding which one of the general or specific measures is more appropriate than the other in tapping into CAF. Bulté and Housen (2012) and Housen et al. (2012) contended that due to the lack of consensus towards the definitions of CAF constructs—which has been partly reflected by the daunting nature of many of CAF indices—many researchers have undertaken critical analyses and reviews of these available

measures. According to these scholars, as well as Housen and Kuiken (2009), the result of such analyses and reviews have shown that there are three problems: (a) the lack of straightforwardness of correlation between particular and general developmental CAF measures and between objective and holistic measures; (b) challenges in the analysis that is in the calculations of many of these metrics; (c) in terms of their reliability, comparability, validity, and sensitivity, both as measures of L2 development and as indices of L2 proficiency and performance.

Housen et al. (2012) added that many of these analyses and reviews emphasized the lack of viewing, not to mention treating CAF as a group of interrelated, regularly changing, and dynamic subsystems. Therefore, Norris and Ortega (2009) recommended researchers to "engage in a much more organic practice…to achieve a thorough understanding of CAF," and "that measurements must provide multivariate, longitudinal, and descriptive accounts of constructs in L2 performance…to capture the complex, dynamic, and developmental nature of CAF phenomena," and also "provide learner-, task-, and L2 form-sensitive accounts of the local SLA ecology, given the ways in which these factors moderate the observations we might be making about CAF" (p. 574). This means that because the nature of such constructs has been proved to be of non-linearity, multidimensionality, variability, and dynamicity, in their studies, researchers should define the construct in question theoretically, observe how it manifests in language productions behaviorally, and use the metrics by which such manifestations can be captured (Bulté & Housen, 2012, Bulté, Housen, Pierrard, & Van Daele, 2008; Housen et al., 2012; Norris & Ortega, 2009). Otherwise, their results will appear inconsistent.

2.7.1 Complexity.

Housen and Kuiken (2009) stated that "complexity is the most complex, ambiguous, and least understood dimension of the CAF triad" (p. 463), "A palimpsest" (Housen et al., 2012, p. 4), and "certainly the most problematic construct…because of its polysemous nature" (Pallotti, 2009, p. 592). These issues are probably why Bulté and Housen (2012) deduced that "there is no commonly accepted definition of complexity" (p. 22), which contributed to the contradictory and mixed findings of the studies that investigated it (cf. Abdollahzadeh & Kashani, 2011; Ishikawa, 2006, 2007; Rezazadeh et al., 2011; Robinson, 2001a, 2001b, 2003; Skehan, 2009, 2013; Skehan & Foster, 2005, 2012). In other words, there are many reasons why there is no agreement on one definition of complexity, such as interpretations-related issues, the references of the construct, its multidimensional and

multicomponential characteristics, and its multiple meanings. The discussion here will address L2 complexity in terms of two sub-categories of linguistic complexity: syntactic complexity and lexical complexity.

2.7.1.1 Syntactic complexity.

2.7.1.1.1 Definitions.

Syntactic or grammatical complexity is one of the most difficult CAF constructs to define. Several justifications have been proposed for such difficulty. Generally, Pallotti (2009) maintained that specifying the use of complexity to describe performance will retain many meanings depending on the various aspects of communication and language it has been applied to. In the L2 field, complexity has been interpreted variously based on two approaches: relative and absolute. According to Housen and Kuiken (2009) and Bulté and Housen (2012), both relative (also known as cognitive complexity or difficulty) and absolute (also known as linguistic or inherent) complexities essentially refer to the properties of language characteristics in terms of rules, constructions, items, structures, and patterns, and of sub-systems, such as syntax, lexis, morphology, and phonology. Nonetheless, complexity in the relative approach is defined from the perspective of cognition in the sense that in order to regard any of the language or system features as complex, it has to make language users employ their mental resources extensively and taxingly when processing or acquiring such features (Bulté & Housen, 2012; Housen & Kuiken, 2009). The issue here is that by determining what cognitive complexity means (see section 2.8 for more elaboration), researchers (e.g., Cho, 2015; Jiaxin, 2015; Rahimpour & Hosseini, 2010; Ruiz-Funes, 2014; Pourdana & Behbahani, 2011; Sample & Michel, 2014) almost always overlook defining the other aspects of complexity, such as the syntactic and lexical complexity used to analyze learners' production. On the other hand, Bulté and Housen (2012) contended that the absolute approach views complexity "as the number of discrete components that a language feature or a language system consists of, and as the number of connections between the different components" (p. 24). There are several facets that increased the difficulty in defining syntactic and even lexical complexity based on this approach and possibly made the validity of their proposed interpretations doubtful: (a) the approach contains multiple subcomponents of complexity that have to be specified along with their designated definition; (b) researchers tend to ignore them when trying to define the former constructs accurately. To illustrate, Bulté and Housen (2012) argued that the L2 complexity has many subcomponents that can be organized hierarchally to L2 complexity → absolute complexity → linguistic complexity →

system or structure complexity → formal or functional complexity → syntactic and lexical complexity. Each of these subcomponents address language complexity from different angles.

For these authors and Housen and Kuiken (2009), linguistic complexity can be interpreted in two different ways. The first is as a dynamic property of the learners' target language system at large (i.e., system/global complexity). Accordingly, when considered at the learners' TL level, system complexity has been interpreted as the degree of richness, breadth, size, width, or elaboration of the learners' TL linguistic system or repertoire that includes the diversity, range, variety, or number of various items and structures that the learner uses or knows (Bulté & Housen, 2012; Housen & Kuiken, 2009). The second one interprets linguistic complexity as the more stable property of the individual linguistic rules, structures, or items that make up learners' TL systems (i.e., structure/local complexity). When considered at the local level of the individual linguistic characteristics themselves, linguistic complexity is labeled as structure complexity, which focuses on depth rather than range or breadth (Bulté & Housen, 2012, Housen & Kuiken, 2009).

Bulté and Housen (2012) stated that structure complexity can be further broken down into two different sub-types: the functional and the formal complexity of a TL feature. Functionally, complexity concerns "the number of meanings and functions of a linguistic structure and to the degree of transparency, or multiplicity, of the mapping between the form and meanings/functions of a linguistic feature," and because "there is straightforward, one-to-one mapping of meaning onto form," the morpheme 's' used to indicate plurality is functionally less complex than the 's' used to indicate third person singular in the present tense (p. 25). Moreover, referring to many situations, formal complexity can be defined in terms of a linguistic feature's structural substance, which is determined by the number of discrete components of a linguistic form. Examples of such includes simple past versus present perfect forms and passive clauses versus active ones (Bulté & Housen, 2012). However, with all of the above-mentioned subcomponents of absolute complexity, there is no interpretation of syntactic complexity that could be said to be invented on the basis of good understanding of what these subcomponents mean or the language aspect(s) angle they focus on. Still, this is not by any means indicating that these are the only reasons why the syntactic complexity definitions are considered to be problematic.

Specifically, some researchers would adopt the general interpretations of complexity and use them for syntactic complexity, while they can be used for lexical complexity as well. For instance, Ellis (2003) defined complexity as "the extent to which the language produced

in performing a task is elaborate and varied" (p. 340). Inasmuch as elaborate and varied language can be in terms of syntax, it can be in terms of lexis. This is probably why Skehan (2009) questioned whether syntactic and lexical complexity are two separate areas or "different aspects of the same performance area" (p. 528). Also, the general interpretations of complexity could be themselves problematic. Pallotti (2009), to elucidate, criticized the complexity definitions provided by Skehan (2009, pp. 510-511)—"More advanced language leading to complexity" and "challenging language"—that of Ellis (2009, p. 475)—"The capacity to use more advanced language"—and that (though not cited in Pallotti's article) of Ellis and Barkhuizen (2005, p. 139)—"The extent to which learners produce elaborate language [determined by] learners[']…willingness to use more challenging and difficult language,…[and by their] preparedness to use a wide range of different structures." The reasons, as emphasized by Pallotti (2009), are that not only do these definitions equate difficult and cognitively demanding with complex, but they are also dependent on the assumption that advanced, challenging, or difficult language equals being acquired late, while the fact is that some linguistic forms are acquired late because they are irrelevant to communication and/or infrequent rather than being structurally complex. Therefore, Pallotti called for the separation of "progress, or developments" from the notion of complexity (p. 593). He further explained this by stating that the terms "advanced" and "developed" are chronologically-related—meaning 'appearing late'—and are not found in any dictionary definitions of "complexity." Moreover, if "complex" implies advanced and is conjoined with time, then complexity must necessarily increase over time. Nevertheless, to meaningfully apply the concept of complexity to development, the second construct cannot be included in the first; otherwise, redundancy results. The increase over time of structural complexity (e.g., subordination ratio or lexical variety) or of cognitive complexity (e.g., the distance across which grammatical features are unified in a sentence) needs to be an independent, verifiable, empirical proposition, not a major part of the definition of complexity itself.

In addition, other authors preferred interpreting syntactic complexity on its own, and this led to the existence of various interpretations. For example, Wolfe-Quintero et al. (1998) believed that the construct "means that a wide variety of both basic and sophisticated structures are available and can be accessed quickly, whereas a lack of complexity means that only a narrow range of basic structures are available or can be accessed" (p. 69). Ortega (2003) viewed syntactic complexity as "the range of forms that surface in language production and the degree of sophistication of such forms" (p. 492). Foster and Skehan (1996) interpreted the constructs as "progressively more elaborate language…[and] a greater

variety of syntactic patterns" (p. 303). The issue here is how to determine the level of structure sophistications and the ranges of the syntactic patterns. If such is undertaken based on native speakers, as it is explicitly indicated in Skehan's (1996b) definition—"Structuring is concerned with the process by which the interlanguage system becomes more complex, elaborate, and structured…, maybe more efficient and less circumlocuitous in communication…, more consistent with input data, and more native-like" (p. 47)—then it becomes problematic, as native speakers are different from nonnatives (Conti, 2015). Also, if the task in determining the structure sophistications and the variety of syntactic patterns are based on nonnative speakers, there is still no agreed upon criteria between scholars. For example, Bulté and Housen (2012) cited the dispute and the contradictory classification among scholars in terms of the third person singular 's'—as a formally simple; yet, functionally complex feature (Ellis, 1990), as a formally and functionally simple feature (Krashen, 1994), or as a formally and functionally complex feature (DeKeyser, 1998).

Moreover, syntactic complexity was defined by Nunberg, Briscoe, and Huddleston (2002) as "the way words are combined to form sentences" (p. 1728). Polio (2001) raised the issue of what constitutes a word, as studies appeared to have paradoxical perspectives towards it. The construct can sometimes be defined in relation to the language investigated, the syntactic complexity measures, or its sub-construct(s). For Stadie et al. (2008) syntactic complexity is "the amount of syntactic phrase structures necessary to generate specific German sentence structures" (p. 2011), while for Huiting (2008) it is the "great length and subordination of T-unit" (p. 11). Mata (2003) stated that in the her research context— Mexican women of different educational levels—the syntactic simplicity is regarded as the greater incorporation of coordinate and independent clauses, whereas syntactic complexity is regarded as the greater use of subordinate clauses. With such a specification of what syntactic complexity means, Norris and Ortega (2009) highlighted another sub-construct by which syntactic complexity can be defined—subclausal through phrasal elaboration. Finally, a solution to the aforementioned issues of defining syntactic complexity was proposed by Bulté and Housen (2012, p. 27) in Figure 1.

Figure 1. Different Levels of Syntactic Complexity Specification
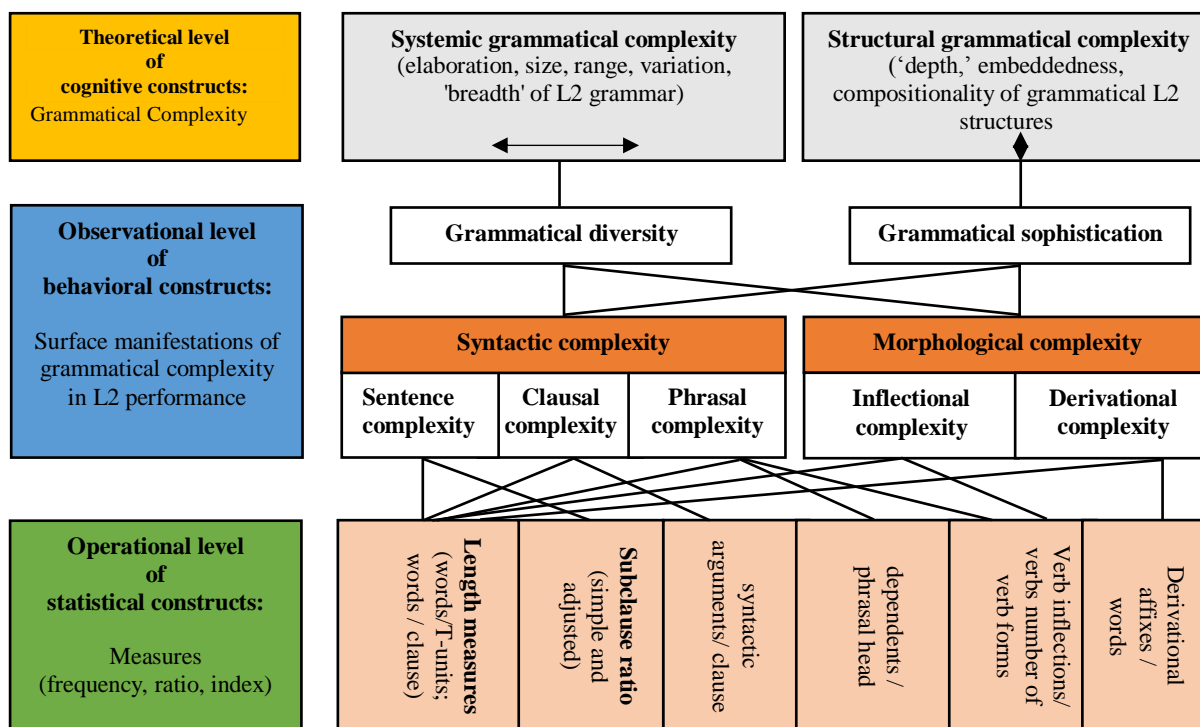
In such an overview, Bulté and Housen (2012) contended that there are three aspects of the concept of linguistic complexity that can be scrutinized. The first is as a speculative, hypothetical construct inherent in or part of the cognitive network. Using this framework, an analysis can be made of its total constituents, how deeply they are lodged, and the nature of the correlations that exist among them. The second is the less abstract, observable aspect of real language production. This relates to the first aspect in that the theoretical ideas can show themselves through usage in different ways and degrees, for example, in the joining and reduction of clauses, employing various verb forms, and utilizing more complex, specific vocabulary. The final aspect involves the measures and tools that have been created to indicate the amount of complexity in any actual sample of language use. By employing these instruments, more objective measurements of complexity can be arrived at, and different samples can be inspected and compared. Such measures are found at the functional, operational level. Lastly, once researchers established in a clear way the links among these levels by identifying, for example, the meaning of syntactic complexity theoretically, the ways in which it manifests itself in language production observationally, and how such manifestations can be measured, the metrics will not only be valid, but the interpretations of the studies will be meaningful (Bulté & Housen, 2012). Nonetheless, such an endeavor is yet to be accomplished.

2.7.1.1.2 Measurements.

One of the things observed in syntactic complexity is that it does not suffer from any shortage of measures (Lu, 2010). Wolfe-Quintero et al. (1998) maintained that the construct analysis is concerned with determining how sophisticated and varied the units of production are, not with calculating the number of units of production. For them, such measures—which are meant to capture the expansion of the construct—can be divided into two categories. The first category deals with measures that analyze sentences, T-units—or "terminable unit[s]" (Hunt, 1970b, p. 198)—and clauses across each other. Communication units (C-units)—units that code and isolate phrases with communicative value but have no verb (Crookes, 1990)— and Sentence nodes (S-nodes)—units which could be a clause, either main, subordinate, or embedded (Chiang & Dunkel, 1992) with a finite or infinite verb phase (Ishikawa, 2006)— can be added to the first category as well (e.g., Bulté & Housen, 2012). The second category concerns the measures that analyze the presence of specific grammatical structures in correlation to sentences, T-units, clauses, C-units, and S-nodes. Chen and Zechner (2011) stated that these categories can be calculated according to three ways: (a) the frequency that counts the appearance of some grammatical structure (e.g., reduced, dependent, adverbial, adjective, or nominal clauses; passives; prepositional phrases; pronouns; articles; and connectors); (b) the ratio that counts the percentage of one type of production unit divided by the percentage of another production unit (e.g., clauses per T-units, sentences, C-units, or S-nodes, T-units per sentences, clauses per error-free T-units, and passives per T-unit, clause, or sentence); (c) the index that "is computing numeric scores by specific formulae" (e.g., complexity index, complexity formula, and coordination index [p. 723]).

Bulté and Housen (2012) provided a more extensive table of syntactic complexity measures. They acknowledged that the measures they collected were classified "tentatively" and "loosely" in terms of the theoretical and behavioral syntactic complexity constructs that they capture. However, such looseness and tentativeness in their classifications were justified, as many of these measures are "hybrid," which can measure several sub-domains and sub-constructs of L2 complexity simultaneously (p. 29), and the fact that some of these suggested measures examine the uninvestigated sub-constructs (i.e., sentential-coordination: *coordinated clauses / clauses*; clausal: *syntactic arguments / clause*; phrasal: *dependents / [noun, verb] phrase*) of syntactic complexity. Nevertheless, there are other issues that stem from the definitions of the measures' production units and the application of such definitions, along with more general problems, such as some measures being repetitive and their effectiveness being restricted to specific proficiency levels, data, and sub-constructs.

Specifically, Ellis and Barkhuizen (2005), Polio (1997, 2001), and Wolfe-Quintero et al. (1998) agreed that there are conflictual opinions on what constitute some units of productions—words, T-units, clauses, and sentences—which resulted in various methods in calculating such units.

Regarding sentences, Maloney and Hopkins (1973, p. 426) defined them as "beginning with a capital letter and/or on a new line and/or having a period, question mark, or exclamation point at the end containing at least one subject and predicate," and in Oxford Advanced Learner's Dictionary (2010), the interpretation was "a set of words expressing a statement, a question or an order, usually containing a subject and a verb [and] in written English sentences begin with a capital letter and end with a full stop/period" (p. 1345). The sentence can also be viewed as "consisting of at least one main clauses, with or without subordinate clauses, and ending with a full stop" (Idol & Croll, 1987, p. 219), which could be used to define a T-unit as well. Yet, the application of these definitions when using a unit measure (i.e., mean length of sentences)—Wolfe-Quintero et al. (1998) used it for fluency— by researchers can lead to distorted results if they hold discrepant opinions of what a word is (Polio, 2001) or how a sentence is categorized. For example, EFL learners in general and Arabs in particular have always been found to be inconsistent and weak in punctuation (Alamin & Ahmed, 2012; Chen 2006; Lintunen & Mäkilä, 2014), which will most likely result in the production of very long run-on sentences or short incomplete sentences. Thus, some researchers would work out the run-on sentences by considering each pair of clauses as one sentence (e.g., Idol & Croll, 1987), whereas others would regard the incomplete sentences only if they are caused by an exclamation mark occurring in the middle of a sentence (e.g., Lu, 2010; Wang & Slater, 2016). Šuster (2011) pointed another weakness in the mean length of sentence measure, which is its lack of sensitivity in indicating the structural differences within sentences.

The T-unit has had widespread use as a measure in the studies of second language writing (e.g., Jiaxin, 2015; Kuiken & Vedder, 2007a, 2007b, 2008, 2011, Kuiken et al., 2005; Mohammadzadeh Mohammadabadi et al., 2013; Nikou & Eskandarsefat, 2012; Salimi & Dadashpour, 2012; Sasayama, 2011) with the goal of coding and revealing second language learners' CAF changes as they become mature (Breland, 1983; Elola & Oskoz, 2010), as well as revealing their writing quality and elaboration (Ha, 2001). The T-unit is defined by Hunt (1970a) "as one main clause plus any subordinate clause or nonclausal structure [phrasal] that is attached to or embedded in it" (p. 4). The term was generated from the fact that any sentence with a main clause—with or without subordinate clauses—allows, grammatically,

punctuation with terminal marks, such as a capital letter at the beginning and a period or question mark at the end. Nonetheless, researchers have different procedures for counting T-units. For example, a T-unit can occur within a sentence and can include sentence fragments punctuated by the learner (Wolfe-Quintero et al., 1998; e.g., Bardovi-Harlig & Bofman, 1988; Hunt, 1966; Lu, 2010); a T-unit can also be across sentences and can exclude sentence fragments (Wolfe-Quintero et al., 1998; e.g., Ishikawa, 2006, 2007).

However, there are other problems with T-units. Hirano (1989) maintained that T-units do not quantify some indispensable writing features, such as coherence, cohesion, organization, and relevance. Gaies (1980) highlighted such shortcomings, which are applicable to syntactic complexity other measures, in the pair of sentences from Moffett (1968, p. 174): "I don't like what is left in the cup after you finish drinking" and "I don't like the dregs." The T-unit analysis relies on the treatment of syntax and ignores the correlation with vocabulary. In other words, the first sentence has three T-units and almost triple the words and clauses than the second sentence that has just one T-unit, and although it could be superficially regarded as more syntactically complex and its writer could be considered to be more proficient in English, there is a possibility that the circumlocution of its T-units is a proof of the writer's "restricted vocabulary development" (p. 56). Gaies (1980) underscored another limitation: the fact that the T-unit would not be suitable to analyze the production of low-proficiency learners because such production would contain many frequent grammatical and lexical errors that will interfere with tabulating the T-units by researchers. Bardovi-Harlig and Bofman (1988) asserted that with the T-units' focus on subordination, it is not capable of capturing other specific features of performance—such as the excessive use of coordination—because in T-units "an independent clause divides all full clause conjuncts into separate T-units, it treats all conjoined and non-conjoined sentences equally, as they were non-conjoined sentences" (p. 5). Thus, they suggested analyzing learners' productions at sentential level (i.e., coordination via coordination index). The reason for the analysis level is that such a common unit of analysis will make teachers and researchers involved in more interactive conversations, and it can facilitate comparisons between learners of various stages of L2 development and characterize their knowledge more adequately. Ishikawa's (1995, as cited in Wolfe-Quintero et al., 1998, p. 70) argument, on the other hand, is that clauses may represent better production units for examining beginning level writing, since they are smaller than T-units and therefore present less context for analyzing language growth in a number of different ways.

In terms of clauses, a clause suffers from the same criticism and more as the T-units. The first issue with a clause is in its various definitions. For instance, Fischer (1984) interpreted it "as a syntactic unit which contains a finite verb" (p. 15), while Lintunen and Mäkilä's (2014) interpretation was "a structure which does not need to include a finite verb…a structure that consists of a verbal element plus an additional clause element, for example, an object or an adverbial" (pp. 382-383). Merriam-Webster.com (2015) defined the term as "a group of words containing a subject and predicate and functioning as a member of a complex or compound sentence," whereas Idol and Croll (1987) viewed it as "a set of words containing a verb, verb phrase, or verb form such as a participle, gerund, or infinitive" (p. 219). Nevertheless, Wolfe-Quintero et al. (1998) deduced that researchers have different methods of calculating a clause in the sense that it can include sentence fragments with no overt verb (e.g., Lu, 2010), or in the sense that it excludes fragments if they are not complete thoughts (e.g., Ishikawa, 1995). For Wolfe-Quintero et al., what is more problematic is that clause types—main clauses (independent), relative clauses (adverbial and adjective), and noun clauses that contain that-clauses and interrogative clauses (nominal)—can be defined individually or altogether in one definition, as well as the fact that they can be embedded and reduced. In the reduction of one sentence's complexity phrasal means—the reduction of the adverbial clause "Because he saved a child's life, he is heroic" to a participle "Having saved a child's life, he is heroic," Wolfe-Quintero et al. (1998, p. 72) emphasized that such a procedure will affect the word count of the measure of T-units and raised a question concerning the possibility of considering a long sentence or a reduced one as more grammatically complex than the other. The answer they provided was that defending a choice of one over the other on the basis of grammatical terms is impossible. They also believed that the judgment should be based on usage, as the tendency of advanced writers may be to use more reduced forms. Conversely, Gaies (1980) stated that no one would deny that the first sentence is syntactically more complex than the second one, but to claim that the former is better than the latter is questionable, as both sentences were written according to two stylistic options, each of which is suitable for a specific context.

Yet, Norris and Ortega (2009) maintained that there are two other general issues with respect to some of the aforementioned measures' validity. The first is that not only are some metrics redundant, since they are tapping into the exact sub-construct of syntactic complexity, but they may also be appropriately incorporated in specific type of data and for particular proficiency levels. For example, in the class of measures that address subordination (e.g., mean number of clauses per T-unit or C-unit, and the number of dependent or subordinate

81

clauses per total clauses), all of them exclusively measure complexification as a case of subordination. When researchers use all such measures in the analyses of the same data, this may result in replicating the findings. Therefore, authors have to employ the variables "depend[ing] only on which unit of discourse segmentation they consider more appropriate for the data at hand" (Norris & Ortega, 2009, p. 560). Norris and Ortega emphasized that the T-unit measures may be appropriate only for analyzing intermediate and advanced students' written samples, which are characterized as having full sentences and clauses (cf. Cronogue, 1985; Farahani & Meraji, 2011; Iwashita, 2010; Zyad, 2016). On the other hand, the C-unit could be suitable for examining oral data of low-proficiency learners (Norris & Ortega, 2009) as well as written data (Robinson, 2001a, 2001b).

The other issue mentioned by Norris and Ortega (2009) is that certain indices are capable of indexing different sources of complexification, implying that researchers have to be careful to use these measures for the designated type of complexity that such measures can capture and not to discuss the results in relation to the wrong types of complexity. The first type is the general syntactic complexity, which is captured by length-based metrics (e.g., sentence, T-unit, and C-unit). The reason for such a belief is driven by the fact that measures with multiple clauses can be lengthened by several methods. For example, adding prepositional phrases, adjectives (pre-or-post modifying nouns), and subordinate clauses can lengthen such measures. Nonetheless, none of these can be determined by the numerical outcomes yielded by these measures. Bulté and Housen (2012) added that when researchers calculate the number of words/morphemes found in the vague and/or hybrid length-based measures—two characteristics represented by the multiple lines connecting the behavioral constructs and statistical constructs measures in Figure 1 (p. 75)—they capture syntactic complexity at the structural compositionality level or substance level, and they inform about the morphological (derivational, inflectional) complexity of a language production. For them, the length-based measures can also capture at the same time various layers (i.e., sentential, clausal, and phrasal) of syntactic complexity.

Norris and Ortega (2009) maintained that the second type of syntactic complexity is one via subordination, a characteristic of intermediate and upper-intermediate learners captured by many measures (e.g., clauses per C-unit or T-unit, dependent clauses per clauses, number of subordinate clauses, and subordinate clauses per clauses, dependent clauses, or T-unit) variables. Conti (2015) contended that such measures only inform about numbers and not types of subordination, and therefore any writer who produces less but more complex subordination than others will score lower in syntactic complexity. Bulté and Housen (2012)

contended that the other issues of such subordination measures are again being hybrid in the sense that they could tap into linguistic compositionality, depth, diversity, and difficulty, as well as the fact that because subordinate structures are assumed to be cognitively difficult in the measurement of the construct, they are selected and assigned greater weight. Similarly, this is the case of the other measures of syntactic complexity (i.e., frequencies, such as passive forms and Wh-clauses, and indices, such as elaboration index, syntactic complexity formula, or coordination index) which give different syntactic structures different weight depending on what is considered to be different levels of difficulty. For Bulté and Housen (2012) difficulty and structural complexity are two different constructs "and the correspondence between the two [of them] still has to be demonstrated rather than a priori assumed, and there is no guarantee that it holds for all syntactic structures" (p. 36).

Norris and Ortega (2009) mentioned the third type of syntactic complexity, "Sub-clausal complexity via phrasal elaboration" (p. 562), which is a feature of advanced learners calculated by mean length of clause. The authors also explained that this kind of complexity differs from the other length measures, as clause length is not affected by variations in how much subordination is included in production. When calculating the average length of all finite clauses (counted irrespective of their being independent, dependent, or subordinate), any increase can be caused in only three ways: (a) the addition of pre- or post-modification within a phrase (by means of adjectives, adverbs, prepositional phrases, or nonfinite clauses); (b) as a result of the use of nominalizations; (c) by the reduction of clauses into phrases, thereby condensing information. Norris and Ortega (2009) also stated that the fourth type of syntactic complexity is clausal complexification via coordination, which is a feature of EFL beginners calculated by coordination index. Bulté and Housen (2012) asserted that both of these suggested measures (i.e., length of clause and coordination index) should not be regarded as pure measures. The reasons are that the application of the former depends or how the clause is defined, as well as the fact that the length could be increased by other means than what Norris and Ortega (2009) mentioned (i.e., at clausal level with the addition of manner, place, and time), while the latter is calculated by dividing the number of coordinate clauses by the total number of combined clauses (i.e., coordinates and subordinates). Wolf-Quintero et al. (1998) believed that the problem with complexity indices (e.g., coordination index, complexity formula, and complexity index) is that they are not significantly related to L2 development.

Finally, Norris and Ortega (2009) deduced that there is a fifth type of syntactic complexity: "Complexity defined as the variety, sophistication, and acquisitional timing of

forms produced" (p. 562). However, Wolfe-Quintero et al. (1998) maintained that the validity of such frequency measures is questionable "because of the lack of a fixed delimiter as found in ratio measures" (p. 75). Thus, Norris and Ortega (2009) called for the necessity of "aim[ing] for the creation of L2 indices that offer well-theorized and potentially better specified operationalizations of variety, sophistication, and acquisitional timing of particular forms, based on more fine-tuned L2 developmental findings" (562).

2.7.1.2 Lexical complexity.

2.7.1.2.1 Definitions.

Lexical richness, knowledge, proficiency, and competence are synonymous terms for lexical complexity, a vital area of investigation in CAF literature. This is not only due to it being considered one of the most illuminative predictors and indicators of learners' language proficiency in general (writing quality in particular), but also because it is one of the major areas in which complexity can be captured (Bulté et al., 2008; Polio, 2001; Wolfe-Quintero et al., 1998; Yu, 2007, 2009). Nevertheless, there are some issues that are "both nomenclature diffusions and confusions" in defining, using, and specifying the construct and its sub-constructs (Yu, 2007, p. 80, 2009; Bulté & Housen, 2012; Bulté et al., 2008; Polio, 2001). The confusion emerged from the fact that a term such as lexical diversity—one aspect of lexical complexity—could refer to learners' (writers or speakers) linguistic features or the production quality (texts or dialogues), or it could refer to four terminology-related problems (Yu, 2007).

First, numerous terms have been used loosely and interchangeably (Bulté et al., 2008) to allude to lexical diversity, even though they mean different things. Yu (2007, 2009) gave a list of these terms, such as lexical rareness or sophistication, lexical variation or variety, lexical or vocabulary density, lexical individuality or originality, vocabulary diversity, lexical complexity, lexical range and balance, and lexical or vocabulary richness. Second, there is not any agreed upon definition for lexical complexity or even its components, as researchers either provided different interpretations or at least added some features to the ad hoc interpretations they used. For example, Wolfe-Quintero et al. (1998) defined lexical complexity as "a wide variety of basic and sophisticated words are available and can be accessed quickly, whereas a lack of complexity means that only a narrow range of basic words are available or can be accessed" (p. 102). Bulté et al. (2008), on the other hand, interpreted the term as "the impression of someone's lexical proficiency created by…the ability to comprehend or use not only the prototypical or default semantic, collocational,

grammatical or pragmatic aspects of a specific word but also a variety of other, more specific, peripheral and less frequent properties" (pp. 279-280). Moreover, both Johansson (2008) and Laufer and Nation (1995) defined lexical density as "the proportion of content words (nouns, verbs, adjectives, and often also adverbs) to the total number of words" (Johansson, 2008, p. 65) or to "non-lexical items" (Mihaljević Djigunović & Letica Krevelj, 2011, p. 253). Yet, in Halliday's (1994) interpretation, he included "the number of lexical items to total running words…, to some higher grammatical unit, most obviously the clause; with or without weighting for relative frequency (in the language) of the lexical items themselves" (p. 56).

For Bulté et al. (2008), lexical variation means the amount of different words "or phonologically-orthographical different word forms" used in a written sample (p. 279). Johansson (2008) added that the words already used should not be repeated much. Malvern, Richards, Chipere, and Durán (2004) argued that the term means "the range of vocabulary and [complete] avoidance of repetition" (p. 3), while Laufer and Nation (1995) preferred to define the sub-construct in terms of one of its possible measurements: "The type/token ratio, i.e., the ratio in per cent between the different words in the text and the total number of running words" (p. 310). On the other hand, Read (2000) interpreted lexical sophistication as "the use of technical terms and jargon as well as the kind of uncommon words that allow writers to express their meanings in a precise and sophisticated manner" (p. 200). Nonetheless, for Laufer and Nation (1995) the sub-construct is defined as "the percentage of 'advance' [English] words in the text" (p. 309). Although ranking words as advanced or not could be left to researchers to decide, they could operate via wordlist programs (based on word families) that label words in terms of the level of frequency (i.e., less or more frequent) in the language. Therefore, Read and Nation (2002) viewed the sub-construct as the proportion of "low-frequency, or 'rare,' words used in a text" (p. 5). However, Lindqvist, Gudmundson, and Bardel (2013) contended that lexical sophistication means a combination of a low-frequency (calculated based on lemma rather than word families) Italian and French words and advanced thematic and cognate words, while in Norris' (2006) research, the sub-construct referred to German lexical range, originality, and non-Germanic words.

Furthermore, some authors used and defined some terms "exclusively to each other" (Yu, 2007, p. 81). The best example of such is the research conducted by Bulté et al. (2008). In that study, researchers claimed that lexical diversity, lexical sophistication, lexical complexity, lexical productivity, and lexical fluency are aspects of lexical proficiency development. That is why each of these supposed aspects was provided with a unique definition of its own. What exacerbates the problem is the fact that other researchers broaden

the use of one of the above terms to contain other terms (i.e., hierarchically). To clarify this, despite using lexical diversity as an equivalent of lexical complexity—as in the case of Daller, Van Hout, and Treffers-Daller (2003)—in the discussion by Mutta (2006), lexical complexity included number of words (which is believed to be a fluency measure), lexical density, lexical sophistication, lexical individuality, and lexical variation (see Read, 2000; Polio, 2001; Wolfe-Quintero et al., 1998). Other authors proposed other aspects of lexical complexity, such as lexical individuality/originality (Polio, 2001), compositionality (i.e., the percentage of semantic and formal components of lexical elements such as denotations, phonemes, and morphemes), collocational, and lexemic (Bulté & Housen, 2012). Some of these aspects are shown in Figure 2.

Figure 2. Different Levels of Lexical Complexity Specification
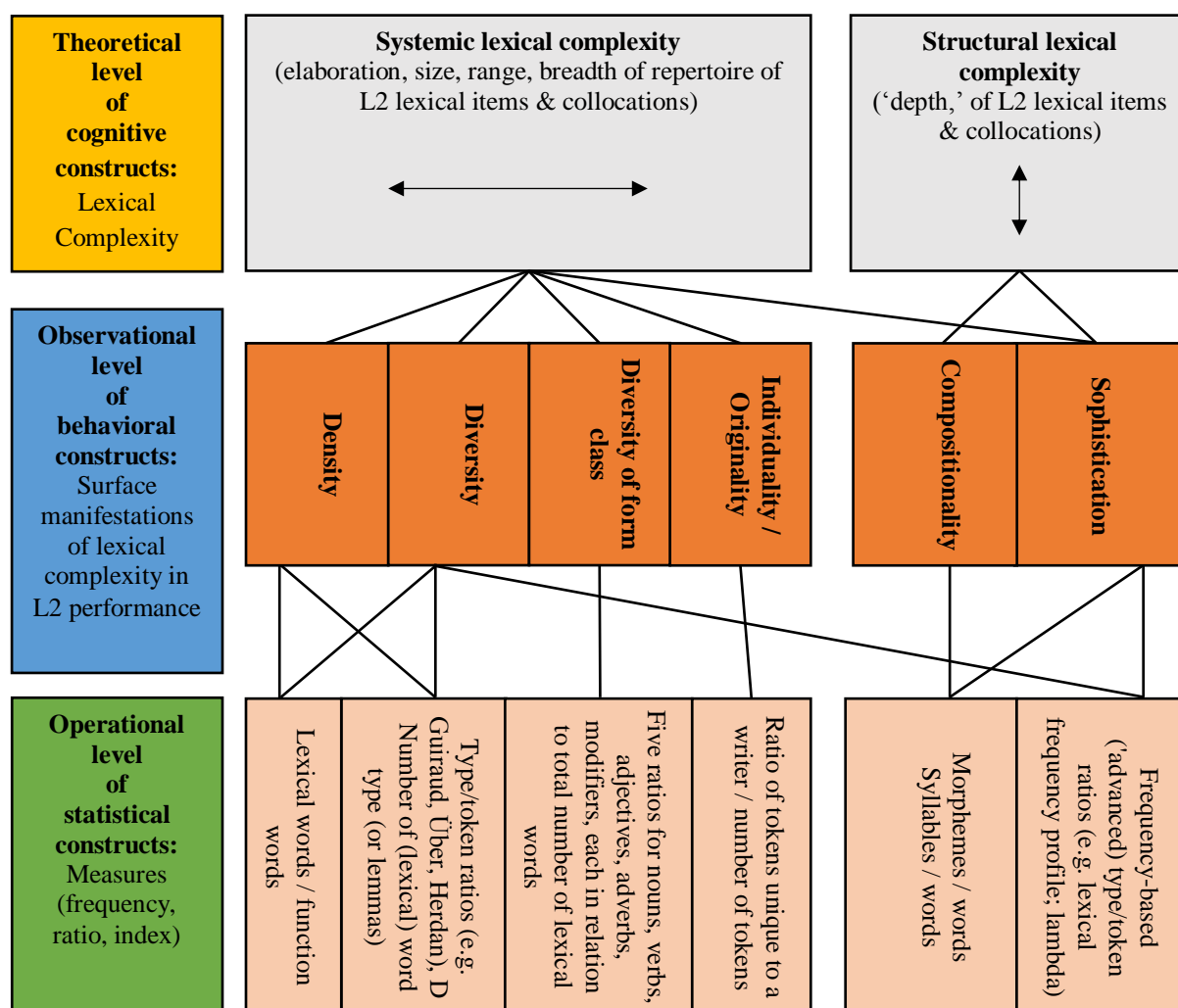
(Adapted from Bulté & Housen, 2012, p. 28, and Polio, 2001, p. 99)

Moreover, there is a fourth problem in the definitions of lexical complexity and its sub-dimensions. The issue is that—as emphasized by Bulté and Housen (2012)—by observing the

above mentioned interpretations, it is clear that these (sub-)constructs were defined mainly operationally (see Johansson, 2008, Halliday, 1994, 2009; Laufer & Nation, 1995) or behaviorally (see Bulté et al., 2008), but not theoretically (i.e., cognitively). In Figure 2, the multiple lines between the aspects of lexical complexity at the various levels of construct specification show that there is no straightforward relationship between theoretical, behavioral, and operational constructs. Nevertheless, in order to establish the validity of lexical complexity and its sub-construct measures, a clear statement of what they mean theoretically is "necessary," although such a claim is sometimes subject to debate (Bulté & Housen, 2012, p. 28).

   2.7.1.2.2 Measurements.

   According to Polio (2001), in spite of the fact that the investigation of lexicon as one of many text components appears to be less common than examining complexity and accuracy, the discussions that have already been made are excellent. Crossley, Salsbury, and McNamara (2009) and Crossley, Salsbury, Titak, and McNamara (2014) agreed that in order to increase the awareness of how the L2 learners produce and process the language, an understanding of their lexical acquisition in relation to its cognitive functions has to be provided. One method of undertaking such an understanding is by specifically analyzing their lexical complexity. Wolfe-Quintero et al. (1998) maintained that the lexical complexity analysis focuses on determining how sophisticated or varied the words or word types are in the writing of L2 writers rather than the number of words found (see Younes, 2016). Thus, they emphasized that since all CAF constructs' metrics are correlated with the lexicon in the L2 development, researchers must not utilize other metrics meant to calculate fluency and accuracy when researching lexical complexity. For example, Polio (2001) contended that the Jacob scale, as well as the metrics of the number of lexical errors and the number of lexical errors by the total number of errors, should be used when aiming to determine the overall quality of texts or their lexical errors (i.e., lexical accuracy). Wolfe-Quintero et al. (1998), on the other hand, argued that the metrics of the number of lexical errors per clause/lexical word, the number of verb lexical errors per verb, and lexical quality index and lexical accuracy index (both latter metrics indicate complexity as well) should be considered only for lexical accuracy. Also, the metrics of the number of words and the number of verbs found in texts should be considered for lexical fluency and not complexity, since these frequency metrics capture "quantity" rather than sophistication or variation (p. 101). Bulté and Housen (2015) believed that the number of words is an indicator of lexical complexity.

Yet, many scholars such as Skehan (2009) have classified lexical complexity metrics into two types: production-internal metrics (i.e., the production/sample is enough for the calculation of these metrics), and production-external metrics, "Which require some sort of general reference material, usually based on word frequency" (p. 514). Type-token ratio metrics are the best example of the former type, whereas for the latter, it would be the metrics that rely on frequency lists generated from corpus analysis, which will be discussed afterwards. Nonetheless, such categorization of the construct's measures is still simplistic and general.

As indicated before, many of the cited researchers (e.g., Bulté & Housen, 2012; Bulté et al., 2008; Daller, Milton, Treffers-Daller, 2007; Lu, 2012; Polio, 2001; Read, 2000; Read & Nation, 2002; Wolfe-Quintero et al., 1998) deduced that lexical complexity entails many sub-constructs, each of which utilize different metrics. Lu (2012) believed that this is due to researchers expanding their studies' scope to address the construct from different angles and testing the validity and reliability of the proposed metrics as being able to capture learners' task performance and their language proficiency. The results of these studies were in some cases discrepant, while in some others they were more complemental. Here, the discussion of lexical complexity will be in relation to the problems identified in the metrics of its major sub-constructs: lexical density, lexical sophistication, and lexical diversity.

2.7.1.2.2.1 Lexical density.

The lexical density measures are the ratio of the number of lexical words, as opposed to syntactic ones, to the total number of words (can be with errors), or to the total number of function words in a sample (Bulté & Housen, 2012; Polio, 2001; Wolfe-Quintero et al., 1998). Johansson (2008) contended that although counting the lexical words (i.e., verbs, adjective, adverbs, and nouns) has the strength of being easy to operationalize, the measure's effectiveness not only depends on the mode (oral or written) of investigation, the age of participants, and genre of production (text types), but it is influenced by repetition. A learner, for example, can have high lexical density from writing redundantly lexical words (and vice versa) when writing pronouns and auxiliaries. Polio raised another problem regarding the researchers' lack of consideration of whether the words are used correctly or not, which may—along with the repetition issue—affect the quality of the texts. Lu (2012) maintained that even with the transparent nature of lexical words, each type has received various definitions. This has made him interpret lexical words more inclusively as

> Nouns, adjectives, verbs (excluding modal verbs, auxiliary verbs, 'be,' and 'have'), and adverbs with an adjectival base, including those that can function as both an adjective and adverb (e.g. 'fast') and those formed by attaching the -ly suffix to an adjective root (e.g. 'particularly'). (pp. 3-4)

Still, the application of this definition is manageable both in time and effort if undertaken by incorporating computer software programs (e.g., Lexical Complexity Analyzer) rather than manually. Halliday (2009) stated that there is an indeterminacy problem in identifying the lexical items as "it can be thought of as a 'content word,' and so treated as if always just one word (in English, defined orthographically), with the line between it and the grammatical item (the 'function word') drawn more or less arbitrarily provided" (p. 75). On the other hand, Laufer and Nation (1995) argued that what may influence the validity of the lexical density measure and make it not necessarily able to capture lexis is the fact that it relies on cohesive and syntactic features of writing in the sense that the increases of structural non-lexical features, such as ellipsis, participial phrases, and subordinate clauses, may lead to decreases in function words used for calculating the measure. Thus, Halliday (1994, 2009) suggested comparing the percentage of lexical words to that of non-embedded clauses—if embedded clauses are counted, a clause will be counted twice. Regardless, this measure has not been used much.

2.7.1.2.2.2 Lexical sophistication.

For lexical sophistication or rareness (Daller et al., 2003), Wolfe-Quintero et al. (1998) mentioned five metrics: sophisticated word types per word types, sophisticated lexical words per lexical words, individual lexical words per lexical words, basic word types per word types, and sophisticated verb types per verbs. The issue of these metrics when undertaking the task manually is the lack of consensus on what are basic, sophisticated, and lexical words between researchers. The definitions of the above-mentioned could be dependent on participants' proficiency levels, their amount of instruction, and their countries' different educational objectives, and thus would result in making the metric "unstable" and minimizing the comparisons between studies (Laufer & Nation, 1995, p. 310). One way to address such subjectivity is to use wordlist programs, which are based on the supposition that learners' acquisition of vocabulary exhibits two processes: first internalizing the most frequent English words (basic) and later internalizing the less frequent (sophisticated [Chapelle, 2001]), and which were generated from analyzing written and spoken language corpora (Adolphs & Schmitt, 2003). A plethora of studies proved their effectiveness in determining students'

word knowledge (Daller et al., 2007). Some of these lists are *General Service List* (West, 1953), which includes 2,000 word families, and word lists that are based on *British National Corpus*, which contains 100 million words (Leech, Rayson, & Wilson, 2014). Skehan (2009) stated that some researchers may also use a Poisson Distribution, which is suitable for low frequency level events in that a production will be divided into 10-word chunks, and then the researchers will then compute how many difficult words—based on their threshold frequency—there are in these chunks. There is also one other frequency list that has been constantly used when analyzing lexical sophistication and variation (Alp, Kerge, Pajupuu, 2013): the *Lexical Frequency Profile* (LFP), which was developed by Laufer and Nation (1995).

The LFP categorized the proportion of words at various degrees of frequency: the first 1,000 words, the second 1,000 words, *The University Word List*, and words that are not included in either of these lists. In other words, the LPF allows researchers to reveal the percentage of words writers use at various lexicon frequency layers. These layers—as observed by Cobb and Horst (2004)—rely on word families (i.e., threshold), and the essential criteria are "frequency, regularity and transparency of inflection and affixation" (p. 27). Even with such potentiality of LFP in yielding—for example, a result for the ratio of the total number of sophisticated word types ("'beyond 2000' words," divided by the total number of word types [Lu, 2012, p. 4])—the issue of labeling basic and sophisticated words remained. For Tidball and Treffers-Daller (2008), the level of word frequency by which sophisticated words are specified is only one dimension of word processing, as there is one equally important dimension—overlooked, however, in the wordlist programs—which is the cognate status of words. These authors stated that the cognate items, which are words in two languages that are similar in spelling and sound, are processed faster and easier than non-cognate words. Therefore, "For British learners of French…, many infrequent items are easy, because the French and English translation equivalents are cognates, e.g., French détester 'to detest,' which is infrequent but probably highly transparent to learners" (p. 300). Polio (2001) identified another problem in the LFP, which is that "risk takers who used advanced words incorrectly may actually detract from the quality of an essay but will score higher in" it (p. 100). Lemmouh (2008) stated that the LFP is unable to discriminate between texts that are either relatively weak or strong.

In addition, some of the lexical sophistication metrics have other issues. Specifically, Arnaud (1992), Irimiea (2012), and Doró (2014) did not incorporate the individual lexical words per lexical words metric for lexical sophistication. Instead, they used it for another

sub-construct of lexical complexity: lexical individuality/originality. Wolfe-Quintero et al. (1998) argued that the sophisticated verb types per verbs metric is like any other type of metrics used for lexical variation, which are affected by the text length.

2.7.1.2.2.3 Lexical Variation.

For lexical variation, range, or diversity (Strömqvist, Johansson, Kriz, Ragnarsdóttir, Aisenman, & Ravid, 2002; Gregori-Signes & Clavel-Arroitia, 2015; Malvern et al., 2004), Crossley, Salsbury, McNamara, and Javis (2011) emphasized that there has been a wide range of more sophisticated approaches to capturing the sub-construct. Wolfe-Quintero et al. (1998), for example, demonstrated nine measures, while the total number of measures discussed by both McCarthy (2005) and Lu (2012) reached more than 27 measures. Nevertheless, each exposes validity issues. McCarthy and Jarvis (2010) stated that one of most significant lexical variation measure problems is that they are based on the supposition of textual homogeneity which "is different rhetorical purposes and strategies may necessitate that different parts of a text have different diversity levels" (p. 382). The issue is that words are treated in isolation, without consideration of their context or even the structure, both of which play a significant role in establishing mental representations of the whole text by the readers. The other problem emerged from the measures' dependency on text length (Crossley et al., 2011). For instance, both Lu (2012) and Šišková (2012) maintained that one straightforward measure of lexical variation is achieved by counting the number of different words written in a text (NDW), but Durán, Malvern, Richards, and Chipere (2004) disapproved of the measure because a writer's production of 100 tokens and 35 types is not clearly more diverse than another writer's production of 50 tokens and 25 types. To overcome such a problem, many researchers have proposed different methods.

Lu (2012) contended that when researchers compare texts of different lengths, they could "truncate all samples to a set length that is no longer than the shortest sample" (p. 5). He also reported the discussion of Malvern et al. (2004) in regards to the standardization of two other procedures. The method used in both consists of randomly choosing a set of sub-samples of equal size from the sample and then averaging the sub-sample's NDW to approximately measure the value of the sample's NDW. The sub-sample used in the first method is made up by randomly selecting a uniform number of words. Yet, in the second method, it is a uniform number of consecutive words that are chosen, beginning from a randomly selected beginning point. Unfortunately, this truncation is not only data-wasteful, but it also demonstrates that using different procedures can cause changes in the data

obtained by comparing the NDW of differing samples (Malvern et al., 2004, as cited in Lu, 2012). Again, Durán et al. (2004) contended that the issue is that researchers' agreement on the number of tokens or methods of selecting them is lacking, which will make the comparison between studies difficult. This has also led researchers to consider a ratio measure (TTR), which is dividing the number of tokens by the number of types, as well as the mean word frequency (MWF) which, according to Šišková, is "a reciprocal measures to the TTR" (p. 28). These measures are, again, problematic. Ellis and Barkhuizen (2005) maintained that it is easier for short productions to show greater TTR than long productions. Also, there is a negative correlation between the text size and the type/token ratio, since when the former increases with the recurrence of more words, the latter will decrease (Richards & Malvern, 2000), "Forming a hyperbolic curve…[that is] if the text were long enough, would eventually end at a point of zero. The problem with TTR, therefore, is that the more language in the sample, the lower our score of language diversity becomes" (McCarthy, 2005, p. 23). Vermeer (2004) deduced that as much as the TTR is influenced by text length and the number of topics discussed, it is also dependent on the relation between or the distribution of tokens and types. The major issue "is that this relation varies as proficiency develops, and if a relation between tokens and types is dependent on development, a stable measure cannot be found" (p. 174).

For Wolfe-Quintero et al. (1998), Durán et al. (2004), and Šišková (2012), there are several approaches to eradicating the impact of text length on at least TTR. The first is to specify the word count, specify the time limit to execute the task, and limit the analysis to the shortest sample. Nonetheless, all of these authors questioned the reliability of such approaches. Specifically, Šišková (2012) contended that researchers' intervention during data collection or their tampering with the samples to make them of an equal length may affect the data and skew the findings, and it may lead to comparing one whole text with just the introduction of another text, or perhaps half of another text (i.e., wasting data). Wolfe-Quintero et al. (1998) raised his concern towards instances when two writers under the same time condition produce texts of the same percentage of words types and receive the same score in TTR even if one of the texts is longer than the other. The longer text indicates greater lexical complexity of whoever wrote it. Durán et al. (2004) stated that assigning an allotted time will result in confounding lexical variation with fluency.

Moreover, the second approach to eliminating the influence of text length on TTR is to apply Mean Segmental TTR (MSTTR [Lu, 2012]). This measure is calculated by dividing a text into consecutive segments of a particular length and then counting the average TTR of all

segments. Ellis and Yuan (2004), for example, divided each of the participants' writings into segments of 40 words and calculated the type-token ratio of each segment by dividing the total number of different words by the total number of words in the segment. Although they acknowledged the limitations of this measure and the lack of access involving the software designed by Malvern and Richards (2002), they claimed that "it is a more valid measure than standard type-token ratio" (Ellis & Yuan, 2004, p. 83). However, Covington and McFall (2010) criticized the MSTTR for not being able to extract the overlapping segments of the production. The alternative for these authors is to use Moving-Average Type-Token Ration (MATTR). The MATTR allows researchers to "choose a window length (say 500 words) and then compute the TTR for words 1–500, then for words 2–501, then 3–502, and so on to the end of the text" (p. 96). Also, such a measure is believed to be uninfluenced by accidental interactions between production unit boundaries and segment boundaries, and it can adequately track the changes within productions.

Furthermore, the third suggested approach to fixing TTR's sensitivity to the text length is to use different "mathematical transformations to compensate for the falling TTR curve… either [by] square roots or logarithms to turn the curve back up and create a model in which the number of types slowly grows instead of slowly falling" (Šišková, 2012, p. 29). Examples included Corrected TTR (CTTR), Bilogarithmic TTR (LogTTR or C), R/Root TTR (RTTR), Advanced Root TTR (ARTTR), Advanced TTR (ATTR) and the Uber Index or Uber U (Daller et al., 2003; Durán et al., 2004; Lu, 2012; McCarthy, 2005; Mellor, 2006; Šišková, 2012; Yu, 2009)[3]. For instance, Larsen-Freeman (2006) counted the number of word types divided by the square root of two times the total number of words (CTTR). The results revealed that the scores increased as a consequence of the increase of the number of types and length altogether. Even though Wolfe-Quintero et al. (1998) emphasized that such a measure is partially impacted by the sample length, along with the other previous measures as indicated by Tweedie and Baayen (1998) and Vermeer (2000). Wolfe-Quintero et al. stated that "in comparisons of second language learners when there are time or conceptual limits [such as requiring narrative writing of past episodes to be finished within 15 minutes], [they] feel that a measure that increases as the length and number of types increases is a better measure" (p. 103).

---

[3] McCarthy (2005) discussed various types of other lexical diversity measures such as Summer's (1996) S, Tuldava's (1993) T, Honore's (1979) R, Michea's (1969) M, Sichel's (1975) S, Mass' (1972) a^2, and Yule's (1944) K. However, they were not included in this discussion since they suffer from the same criticisms pointed for the other measures.

Šišková (2012) mentioned the fourth approach to fixing the influence of text size on TTR is utilizing measures that incorporate "specific or all spectrum elements or using parameters of probabilistic models" (p. 29). McCarthy (2005) stated that because of the TTR curve, which shows the distribution of the TTR throughout the text, it was clear that all of the previous measures "looked at single points within the curve or [usually], the very final point of the curve;" consequently, some researchers felt the need for "curve fitting…to produce a formula that offers a theoretical parameter which closely fits the empirical TTR curve. This parameter would then stand as a measure of" lexical variation (p. 47). The best example of such is the D measure. [4] This measure was produced by Richards and Malvern (2000) and was implemented in a software called vocd. It has three advantages: it analyzes all data available (even short ones), it is not a function of the number of words in the text, and it is more informative in telling about how the TTR differs over a range of token size for each writer. Šišková (2012) explained that the software works by using the original TTR formula and taking 100 random samples of 35-50 tokens (without replacement) and then calculating the average D. The final score that results from repeating the whole procedure three times is the overall average. The results of Richards and Malvern's (2000) study in applying the new measure on the L2 French interviews proved the validity of D as a metric of lexical variation and the effectiveness of the software in analyzing language data. They admitted; however, that the reliability of the D measure could be questioned if the text analyzed is less than fifty tokens and that the results would slightly change every time the software is run.

For this reason and others, two other measures were proposed: the hypergeometric distribution function (HD-D)—which is not approximated as the vocd-D measure—by McCarthy and Javis (2007, 2010) and the Measure of Textual Lexical Diversity (MTLD) by McCarthy (2005). The HD-D can be seen as a corrected version of vocd-D; nevertheless, it "calculates, for each lexical type in a text, the probability of encountering any of its tokens in a random sample of 42 words drawn from the text" (Jarvis & McCarthy, 2010, as cited in Treffers-Daller, 2013, p. 82). On the other hand, the MTLD is based on the MSTTR in that it "treats the text as a sequential whole, rather than a randomly ordered concatenation of discrete items" (McCarthy, 2005, p.90). Nonetheless, it addressed the latter measure's "problem by replacing segmented strips of tokens with segments of a given TTR score." This means that the MTLD does not attempt to identify the length of texts in terms of tokens (i.e.,

---

[4] The D measure here is a more valid replacement of the original D measure that was invented by Malvern and Richards (1997) based on a single child and had different equation.

the whole texts TTR), but instead identifies the common TTR decline to all texts (i.e., the number of "times a common TTR can fit into a single text" [p. 94]). In Crossley et al.'s (2009) words, the MTLD separates a sample into segments of a TTR value of .71. Then, for a forward value, the sample total segments are divided by its total number of words. After that, for a reverse value, the whole process is repeated starting from the sample end rather than the beginning. Finally, for the MTLD value of the sample, both the forward and reversed values are summed and divided by two.

McCarthy and Javis (2010) examined the MTLD, vocd-D, HD-D, and Mass, and they concluded that: (a) the HD-D is a vital replacement to the vocd-D standard; (b) the MTLD, vocd-D, and Mass can measure unique lexical information; (c) the MTLD was the only measure to perform well as regards divergent, convergent, incremental, and internal types of validity, and it was the only measure not impacted by the size of texts. Koizumi (2012) investigated TTR, RTTR, vocd-D, and MTLD and found that the last measure was the least likely to be influenced by L2 texts of 50-200 tokens. However, Koizumi suggested that the measure should not be used with texts less than 100 tokens. Finally, McCarthy's (2005) analysis of 23 genres of spoken and written productions revealed that out of the 14 measures of lexical variation, only MTLD did not correlate with text length. In contrast, Treffers-Daller (2013) found that MTLD, HD-D, and vocd-D are all affected by the size of text, highlighting the necessity of equalizing the text lengths if one seeks to attain significant results.

2.7.2 Accuracy.

2.7.2.1 Definitions.

Accuracy/correctness is believed to be the simplest and most internally consistent, straightforward, and transparent—and probably the oldest—construct of the triad (Dahmardeh & Shahmirzadi, 2016; Housen & Kuiken, 2009; Pallotti, 2009). Housen and Kuiken (2009) maintain that "there is not the same amount of (relative) denotative congruence in the applied linguistics community with regard to fluency and complexity as there is with regard to accuracy" (p. 462). The reason for such a statement is that most researchers, regardless of the mode of interest and investigation (oral or writing) that interests them, have agreed on one thing—that this "broad term…has to do with the absence of errors" (Polio, 2001, p. 94).

Exemplifying this, Skehan (1996a) believed the construct "concerns how well language is produced in relation to the rule system of the target language" (p. 16), while elsewhere (1996b), he contended that it "relates to a learner's belief in norms, and to performance which

is native-like through its rule-governed nature" (p. 46). Skehan and Foster (1999) defined accuracy as "the ability to avoid error in performance, possibly reflecting higher levels of control in the language, as well as a conservative orientation, that is, avoidance of challenging structures that might provoke error" (p. 96). Also, Thornbury's (2000) interpretation was "the extent to which the learner's output matches some external standard—traditionally the output of an idealized native speaker" (p. 3), whereas Ellis' (2003) definition was "the extent to which the language produced in performing a task conforms with target language norms" (p. 339). For Wolfe-Quintero et al. (1998), the term referred to "the ability to be free from errors while using language to communicate in either writing or speech" (p. 33). Ellis and Barkhuizen (2005) stated that "learners who prioritize accuracy are seeking control over the elements they have already fully internalized and thus adopt a conservative stance towards L2 use" (p. 139). In all of the definitions above, accuracy refers to "the extent to which an L2 learner's performance (and the L2 system that underlies this performance) deviates from a norm" (Housen et al., 2012, p. 4).

On the other hand, Towell (2012) and Housen et al. (2012) argued that there are some disputes over how the construct should be defined. The reason is that while the term may seem to be "conceptually simple," there could be some problems in establishing a firm definition to it and in applying that in L2 performance analysis (Housen et al., 2012, p. 4). For example, in the pursuit of determining the construct, its scope can differ from one study to another, "And may or may not include word choice, spelling or punctuation errors" (Polio, 2001, p. 94). Housen et al. (2012) asserted that there are other problems related to any errors' or deviations' relative nature—since some of them are often more erroneous or less deviant than others—and to the criteria used to specify the deviated structures. For Housen and Kuiken (2009) and Towell (2012), these criteria can be based on native speakers' norms, nonnative speakers' non-standard but acceptable (i.e., in some communities or some social contexts) ones, or on the same nonnative speakers but at different stages (low and advanced) of learning. Thornbury (2000) maintained that how widespread English is resulted in making it an international language, which has developed into different "Englishes," and it could be hard for researchers to agree on one standard (p. 3). Such problems drove Housen et al. (2012) to suggest broadening the interpretation of accuracy to acceptability and appropriateness, even though measuring all of these terms is still not achieved.

2.7.2.2 Measurements.

Conti (2015) asserted that the problem of using accuracy as a metric of proficiency development is that the level of L2 learners' accuracy in handling some language features is determined by comparing them to the norms of native speakers, let alone the fact that such a process will not yield significant information in regards to these former students' interlanguage development. Wolfe-Quintero et al. (1998) has long before disputed such an argument, maintaining that the goal of accuracy metrics is exactly "the comparison with target-like use," because the metrics that can capture the interlanguage system are those of lexical and syntactic complexity (p. 34).

Errors, which are "word[s, phrases, or sentences]…that [are] not said or written down correctly" (Oxford Advance Learner's Dictionary, 2010, p. 948), have always been one of the elements used in the analysis of second language development in general and accuracy in particular. Therefore, these errors—along with the correct forms of some structural and morphological aspects (e.g., connectors, pronouns, articles, and plurality)—were generally the basis of three types of accuracy measures: indices, frequencies, and ratios. The fundamental suppositions for such inclusion or errors—as highlighted by Wolfe-Quintero et. al. (1998)—are (a) when writers become more proficient, they will produce more accurate texts easily and increasingly; (b) accuracy development (one part of the language overall development) in having few errors coincides with the evolution of complexity (i.e., more grammatical and lexical variations) and with the growth of fluency (i.e., greater length and faster rate). The latter assumption is nevertheless to be proven, as some researchers (e.g., Li, 2000; Jong, 2009) substantiated the belief of Skehan (1996a, 1996b, 1998, 2009), Skehan and Foster (1997, 1999; 2001, 2005, 2012), and Foster and Skehan (1996, 1999)—that CAF constructs do not develop collinearly and synchronously in the sense that when learners try to produce complex language (increase of complexity) that is not automatized and at disposal, they will most likely commit more errors (decrease of accuracy; cf. Sasayama, 2011).

Returning to the three types of measurements, accuracy indices (e.g., intelligibility index, error index, error formula 1 or 2, lexical quality index, and lexical accuracy index) emphasize some formula for the calculation of accuracy. The issues with these indices are that they are either dependent on identifying lexical and rare words types whose definitions varied between researchers (see section 2.7.1.2.2) and errors (will be explained afterwards) or subjectivity, as is the case of intelligibility index in which the researcher is the one who scores the T-unit according to four ranks of intelligibility: 0 = unintelligible, 1 = partly intelligible, 2 = completely intelligible, and 3 = completely accurate. The frequency measures

emphasize the proportions of correct forms in articles, pronouns, and connectors, as well as the number of errors, their degrees, and how error-free some T-units and clauses are. The ratio measures, on the other hand, emphasize the proportions of correct structural and morphological forms in relation to contexts or word, and the number of errors in syntax, semantics, morphology, and lexis and errors degrees, in relation to units of production (words, sentences, T-units, and clauses). In terms of accuracy measures of correctness, Ellis and Barkhuizen (2005) observed two issues in target-like usage of verb morphology and plurals measures. For them, it is questionable to assume that if learners are capable of using one grammatical feature accurately they will be capable of using others as well—since grammatical features are not acquired "concurrently"—some of which can be acquired earlier than others. Also, factually speaking, the easiness and difficulty of a particular grammar feature is dependent on its existence in learners' L1. For "learners whose L1 does not contain plural markers may be penalized by the target-use of plurals measure (p. 151). Ellis and Barkhuizen recommended using such specific measures with more general ones, like error-free clauses. Wolfe-Quintero et al. (1998) admitted that the analyses of studies that adopted target-like measures—which are known to follow a certain sequence of development (plurals, then definite articles, then indefinite articles, then third-person singular)—are theoretically motivating. Not only did these studies partially support these measures' developmental claims, but the distinction between developmental measures and developmental sequences research is also not clear. Consequently, Wolfe-Quintero et al. (1998) call for "more sophisticated ways to measure the language use in writing in developmental terms—not to measure language use directly against target language norms, but against well-established developmental sequences (such as those related to morphemes, negatives, questions, and relative clauses)" (p. 37).

With respect to error-free measures, such measures also suffer from many issues driven from the errors themselves, their criteria, and the language aspects they regard. For example, Ellis and Barkhuizen (2005), Hirano (1989), and Polio (1997, 2001) claimed that there are still some difficulties in identifying what constitutes an error, a T-unit, and a clause—the latter two language aspects have been discussed in the syntactic complexity measurements section. Wolfe-Quintero et al. (1998) added that because of the errors' various types, researchers are forced to make decisions on what to count as an error in order to determine whether the production units are error-free or not. These decisions to exclude or include different types of errors are based on the "learners' level, the discriminative value of the errors within the population, and the researcher's preferences" (p. 35).

For example, in determining whether a unit of production was error-free, Ellis and Yuan (2004), Armstrong (2010), Storch (2005), Storch and Tapper (2009), Al-Humaidi (2008), Piri et al. (2012), Farahani and Meraji (2011), Serrano (2011), Rezazadeh et al. (2011), Ghavamnia et al. (2013), and Tavakoli and Rezazadeh (2014) counted errors in syntax, morphology, and lexical choice. Abdollahzadeh and Kashani (2011) considered errors in syntax and punctuation (i.e., only comma errors related to restrictive/non-restrictive relative clauses) errors. Wigglesworth and Storch (2009) regarded grammatical errors (e.g., omitted plural 's', omitted preposition, omitted articles), as well as capitalization and lexical choice errors only when they impede meaning, whereas Mohammadzadeh Mohammadabadi et al. (2013) counted lexical, grammatical, and spelling errors. Kuiken et al. (2005) and Kuiken and Vedder (2007a, 2008; 2011) considered errors in spelling, meaning, and grammatical form. They elsewhere (i.e., Kuiken & Vedder, 2007b) regarded errors in grammar, lexicon, orthography, appropriateness, and other areas. Jong (2009) counted spelling, grammar errors (e.g. tense/reference and articles, such as omission or overuse), and nonnative-like usage in word choice to an inappropriate context, as well as word order errors, while Dobao (2012) considered mechanical, grammatical, and vocabulary errors. Not all researchers examining error-free units of production indicated what they calculated as an error (e.g., Jafari & Ansari, 2012; Khorasani et al., 2012; Larsen-Freeman, 2006; Nikou & Eskandarsefat, 2012; Ong & Zhang, 2010; Zabihi et al., 2013). Wolfe-Quintero et al. (1998) asserted that "these differences between criteria for determining 'error-free' would certainly affect the results, making it difficult to interpret the meaningfulness of comparisons across studies" (p. 35). They also stated that the incorporation of multiple types of errors in any study could make any metric of general error (e.g., error per T-unit and error-free T-unit ratio) "too coarse," and that the results of analyzing other types of errors (e.g., in prepositions, articles, or/and verb tense) "should be viewed cautiously," as they "have not been tied to finely-tuned information about developmental stages."

However, as mentioned earlier, Housen et al. (2012) argued that errors differ from one another in their level of deviancy and raised a question of "whether [their] criteria should be [dependent on]…prescriptive standard norms (as embodied by an ideal native speaker of the target language) or [on] non-standard and even non-native usages acceptable in some social contexts or in some communities" (Housen & Kuiken, 2009, p. 463; Housen et al., 2012). Other scholars, such as Bardovi-Harlig and Bofman (1989, as cited in Wolfe-Quintero et al., 1998, pp. 35-36), contended that error-free variables are problematic because they do not indicate what types of errors, how many types of errors, or even how these errors were

distributed within the T-unit or any other units in the analysis. As Polio (1997) stated, an error-free T-unit "does not…take into account the severity of the error nor the number of errors within one T-unit" (p. 112). As a result, a T-unit containing multiple errors is treated similarly to a T-unit containing only one error (Wolfe-Quintero et al., 1998). Conti (2015) deduced that what is even more problematic is that without further consideration to the errors types and levels, the accuracy of a learner who committed 10 errors in advance structures would be regarded as equivalent to the accuracy of a learner who made 10 errors in simple structures. He also highlighted the possibility of penalizing the learner who commits more mistakes as a consequence of attempting complex structures, in that the learner would score less than the learner of the same proficiency level who plays it safe in addressing only simple structures.

As much as classifying the gravity of errors into three different hierarchal degrees (i.e., by assigning errors with different weights according to how far they impede communication) would help overcoming Bardovi-Harlig and Bofman's (1989) and Polio's (1997) criticisms (e.g., Kuiken et al., 2005, Kuiken & Vedder, 2007a, 2008, 2011), there are other remaining issues of such classification judgements being subjective—and therefore not comparable— and not dependent on an "explicit model of what constitute…[the] knowledge" needed for each type of these errors (Wolfe-Quintero et al., 1998, p. 36). Pallotti (2009) maintained that "there is [also] a risk…of making [accuracy] a spurious construct covering distinct dimensions" (p. 592). He further explained this by stating that when analyzing two essays of 100 words in which one of them contains 10 errors that do not affect communication while the other includes 10 mistakes that impact the delivery of the intended meaning, the former text is not more "accurate" than the latter text but rather more "'understandable, or 'communicatively effective'" (p. 592). In other words, when two students produce two sentences, one of which has compromised communication (e.g. *colorless green ideas sleep furiously on the justification where phonemes like to plead vessels for diminishing our temperature*"), while the other only violates the L2 grammar rules (e.g. *"me no likes go dance"*), this "proves that [researchers] are dealing with different constructs (accuracy versus comprehensibility), not the relative degree of the same construct." (pp. 592-596).

A similar problem arises in the case of assigning errors with grades based on their developmental sequence level. When judging a text of 100 words that contains 10 errors in "subjunctives and conditionals," that does not make it "more accurate" than a text with the same length and with 10 errors in "articles and pronouns," but instead makes it more "developed" or "advanced." It is possible to find texts that are rarely developed, nevertheless

accurate, "And texts containing many errors but exhibiting several traits of evolution." Therefore, "Development," and "accuracy" are separate norms, and should be examined by using various indices, rather than by using measures that mix the two construct together (Pallotti, 2009, p. 592).

2.7.3 Fluency.

2.7.3.1 Definitions.

According to Abdel Latif (2013), there has been much dispute among researchers and linguists over the term fluency, a claim justified by the existence of a greater number of varying definitions proposed for writing fluency as compared with speaking and reading fluency. This is not to say that there is a consensus on one definition (Kowal, 2014) in any of these skills, but rather that the interpretations of what fluency consists of in composition skills are more numerous and therefore problematic. The fact that some research (e.g., Ishikawa, 2007; Jong, 2009; Ong & Zhang, 2010; Rahimpour & Hosseini, 2010) incorporated two or more metrics to assess writing fluency and how such a construct was conceptualized in the literature evidently shows that there is a "definitional confusion" over the construct (Abdel Latif, 2013, p. 99).

Authors have proposed different approaches to defining the term in composition. First, some researchers have argued that the construct could be interpreted by describing it qualitatively (i.e., characteristics). The definition could be general implicit/explicit, as is the case of the term referring to automaticity of language use (Kowal, 2014), or learners' L2 general knowledge proficiency, specifically as characterized by the ease of impressions, smoothness, and rhetoric of writing (Dahmardeh & Shahmirzadi, 2016; Housen & Kuiken, 2009). For example, Crystal's (1987, p. 421) interpretation was the "smooth, rapid, effortless use of language." Lennon (2000) defined the construct as "the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language" (p. 26) and Housen et al. (2012) viewed it as "the ease with which learners produce the L2" (p. 3). Conversely, some authors suggested that the definition should be narrowed. For instance, Dobao (2012) interpreted fluency as the "length of the text" (p. 47). For Polio (2001), the definition of writing fluency should not be concerned with how fast a writer can write, but rather how close the writing is to that produced by a native speaker, or "how native-like the writing sounds" (p. 105). Oh (2006) contended that the construct can also be explained in terms of the "rating scale criteria" or measures used in investigating it (p. 12), exemplifying the study of Tarone, Downing, Cohen, Gillette, Murie, Dailey (1993) that used a holistic

scale to compare the written samples of Southeast Asian-American immigrant children with those of international students and native undergraduates, and interpreted fluency as standardness, ease of reading, length, idiomaticity, and nativeness. Abdel Latif (2013), nonetheless, argued that writing fluency should be defined operationally as "writers' ability to produce texts in large chunks or spans and is optimally measured through using the length of writers' translating episodes or production units" (p. 104).

Some researchers even recommended adopting the interpretations of fluency from other modes—mainly speaking—or interpreting the fluency of both writing and speaking together. For example, speaking fluency is defined as the ability "to produce (and comprehend) speech at relatively normal rates, approaching (but not necessarily identical to) one's own native-language speech rates. In particular, one would look at features such as rate, pausing, reformulation, hesitation, redundancy, and the use of lexical units" (Skehan, 1996b, p. 48), as "virtually error free, using a wide range of structures and vocabulary, and using them appropriately," or as "to produce speech at the tempo of native speakers, unimpeded by silent pauses and hesitations, filled pauses ('ers' and 'erms'), self-corrections, repetitions, false starts" (Lennon, 1990, p. 390). Even if writing scholars ignore the fact that some of these definitions implicitly make native speakers the baseline of comparison by "imply[ing] some normative reference, usually assumed to correspond to native speakers' behavior" (Pallotti, 2009, p. 591)—a problem said to be true of Polio's (2001) interpretation of the term, as well—there is also the inevitable problem of being based on spoken abilities rather than writing (Abdel Latif, 2013). In addition, Wolfe-Quintero et al. (1998) believed that as much as Fillmore's (1979) way of determining speaking fluency in terms of how fast fluent speakers speak and how appropriate, coherent, complex, and creative their speech is can be used to analyze writing fluency, it is wrong—as Fillmore acknowledged—that such interpretation contained ambiguous characterizations, since they overlap with complexity, accuracy, and other "discourse-related criteria" (p. 13). Also, when defining speaking and writing fluently altogether, Fellner and Apple (2006) criticized Brown's (1994, p. 113) interpretation: "Saying or writing a steady flow of language for a short period of time without any self- or other correction at all." The reason is that it would lead to an inevitable conclusion that the more words that are produced and the longer the flow of language, the more fluent the text is, while the reality is that when learners write using computer programs, increased word counts would be attributed at least partially to the development of their typing speed throughout the duration of the program instead of true writing fluency improvement. Also, the aforementioned definition disregards important variables like text comprehensibility

and lexical complexity, and when ignoring such, students' fluency would be wrongly conceived as developing even if the students kept repeating the same simple sentence over the allotted time. Consequently, Fellner and Apple (2006) preferred interpreting writing fluency "as the number of words produced in a specified time frame, together with lexical frequency, irrespective of spelling and content, provided that the writer's meaning is readily understandable" (p. 19).

Furthermore, some definitions proposed are believed to be process-based (Abdel Latif, 2013), that is, to describe "the extent to which the writer is willing to follow through a range of processes or procedures that lead to the completion of the piece of writing" (Bruton & Kirby, 1987, p. 90). In speaking mode, Skehan and Foster (1999) defined the term as "the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems" (p. 96). Skehan (1996b) asserted that fluency is "the learner's capacity to mobilize an interlanguage system to communicate in real time" (p. 46; Skehan, 1996a). These definitions indicated the pressure performing language users encountered while delivering their messages in speech and writing (Wolfe-Quintero et al., 1998). Therefore, to Wolfe-Quintero et al. (1998), fluency means, "More words and more structures are accessed in a limited time, whereas a lack of fluency means that only a few words or structures are accessed" (p. 14). Nevertheless, all of the interpretations address the writing processes generally, and because of the lack of specification of the processes in writing, researchers turned to the sub-dimensions or processes identified for the speaking mode. In such a mode, Skehan (2003a, 2009), and Tavakoli and Skehan (2005), identified four sub-dimensions or processes for fluency: repair fluency (i.e., repetitions, self-corrections, misformulations, and false starts), speed fluency (i.e., linguistic units, production rate, and density), breakdown fluency (i.e., location, length, and number of pauses), and automatization (i.e., length of the run). Some of these processes were modified for writing—such as changes made to text, writing pausing, and rate—and were defined by Ransdell, Arecco, and Levy (2001, p. 114) as "the number of words word-processed per minute, controlling for typing speed and including those words deleted before the final draft" (p. 114) and Palvianen, Kalaja, and Mäntylä (2012) as "pauses, and revision made to the text being composed" (p. 48). Yet, these processes should not be used when defining fluency, as they are still in need of being empirically validated.

Finally, although many researchers differ in their product- and process-based definitions, it is of significant importance when investigating the construct to specify the sub-component(s) in question and provide proper specialized definition(s), since "once [this] is

established…, it is in principle relatively transparent what is [or are] being measured" (Pallotti, 2009, p. 592). Otherwise, assuming that the same interpretation of fluency is shared by everyone will lead the validity of the measures used and the judgements made by assessors to being questionable (Chambers, 1997).

2.7.3.2 Measurements.

As indicated before, many definitions proposed for writing fluency resemble in one way or another those for speaking fluency. Therefore, it is of no surprise that the two modes share some similarities in measurements. Wolfe-Quintero et al. (1998) maintained that in writing, "Fluency is not a measure of how sophisticated or accurate the words or structures are, but a measure of the sheer number of words or structural units a writer is able to include in their writing within [/without] a particular period of time" (p. 14). Such a perspective, which is shared by many researchers, led to the availability of many frequency and ratio measures.

The frequency measures concern capturing the number of production units (e.g., words, verbs, T-units, clauses, and sentences) and the number of words in error-free production units or just production units (e.g., T-units and clauses). However, Polio (2001) contended that when using the number of T-units and clauses as measures of fluency, reliability will become trivial, since writers will be penalized if they write longer structures, which reduces the number of these language aspects. Abdel Latif (2009) added that counting the aforementioned along with sentences will only reflect the quality features of writers' texts rather than the flow of their writing processes. On the other hand, the ratio measures emphasize tapping into the rate of the production unit (i.e., words per minute, which is used along with the similar syllabus per second measure in speaking) and the length of production units, such as T-units, clauses, error-free T-units/clauses, and complex nominals per T-units/clauses, which are similar to oral pause-bound utterances measure (Wolfe-Quintero et al., 1998).

As posited by Abdel Latif (2013), there are some measures of fluency that were adapted from the oral fluency research (e.g., Ellis & Barkhuizen, 2005; Koponen & Riggenbach, 2000; Skehan, 2003a, 2009; Tavakoli & Skehan, 2005). These metrics can be divided into two different categories: (a) process-based, such as breakdown fluency (e.g., writers' pausing) and length of bursts occurring between pauses (e.g., length of rehearsed sample between pauses, and length of translating episodes composed between pauses); (b) product-based, such as speech rate (e.g., writing rate, text quantity), repair fluency (e.g., changes

made to the writing), and perceptions of listeners towards speakers' fluency (e.g., perceptions of readers towards writers' fluency in linguistic features characterizing rhetorical functions, text cohesion, coherence, and structure, sentence length, and length and number of T-units). Nevertheless, Abdel Latif (2013) argued that all the above metrics, except one, are problematic. The reasons are that for the first category of metrics, the pausing and its length are affected by the writers' types—as amateur writers are likely to have shorter pauses than professional writers—pausing location, and some writing processes (i.e., planning, monitoring, retrieving, and reviewing, but not the physical writing) used. Abdel Latif (2013) maintained that, "When writers pause, they use any composing process other than transcribing…; some of these processes may facilitate the ease of written language production while others may hinder it" (p. 102). In addition, there is the fact that when writers execute a task, they will consume a much more varied amount of time and will therefore have much more varied amounts of pausing time, let alone the fact that the texts they submit are most likely the last copied versions of two or more drafts written in a longer time than the final versions. Abdel Latif (2013) also asserted that the reliability of the length of writers' rehearsed text measure is questionable, since it relies on students' verbal ability, and the fact that in some cases L2 students plan their texts in their L1 and do not transcribe or rehearse much of their planned productions. He also showed that even though some studies highlighted the possibility of capturing writing fluency through the length of the sentence measure, these studies indicated that the length of translating episodes measure has more validity because it "assesses real-time fluent written production and is compatible with the cognitive characteristics of writing performance" (p. 104).

Regarding product-based metrics, Wolfe-Quintero et al. (1998) contended that some depend on the task type. For Abdel Latif (2013), this is one of the justifications why, for example, text quantity, writing rate, and repair aspects should not be regarded as real-time measurements of writing fluency. For him, since writing is the most demanding of all language skills, writers are obliged to incorporate many of the aforementioned processes, and a lot of the time allotted for performing the composition task will be allocated to the four preparatory processes rather than to the writing, which may impact the text quantity and writing rates measures. Also, these measures are influenced by some performance variables, such as writers' negative attitudes towards writing—which will result in varying time spent in executing a given task—their pre-task decision in including a particular number of line or words in the text, and their familiarity with the topic assigned (see also Abdel Latif, 2009). Moreover, Abdel Latif (2013) emphasized the issue of repair aspects metrics as appearing to

be indicators of writing strategies rather than writing fluency. He also maintained that the number and length of T-units, linguistic features characterizing rhetorical functions, and text structure, coherence, and cohesion metrics are only able to assess complexity and/or accuracy aspects in the written production not its flow.

In contrast, Wolfe-Quintero et al. (1998) disagreed with the assumption that length-based metrics are measures of complexity, believing instead that they indicate fluency. The reasons are not only because length measures do not distinguish the various ways of how the length can be achieved, but also because of the exemplified study by Ortega (1995) in which factor analysis showed that words per utterance (length) was strongly correlated to syllables per second (rate). Both of these "were part of one factor presumed to be fluency, but not related to factors associated with lexical complexity of grammatical accuracy" (Wolf-Quintero et al., 1998, p. 14). Also, Wolfe-Quintero et al. (1998) acknowledged that instead of considering the average length of complex nominals in T-units/clauses as measures of complexity, they regarded them as indicators of fluency. Their argument was that these metrics do not gauge language use complexity, nor "how prevalent [complex nominals] are in the writing sample, nor the sources of the complexity[; nonetheless] they merely measure how fluent (how long) the complex nominals are" (p. 15). Additionally, Wolfe-Quintero et al. (1998) stated that they used many error-free variables, such as the average number of words per error-free T-unit/clause (W/EFT and W/EFC) and the total number of words within error-free T-units/clauses (WEFT and WEFC). In spite of the fact that W/EFT is considered a complexity measure—similar to words per T-unit (W/T) —all of the error-free length variables show the level of fluency (but not the level of accuracy) of the error-free clauses and/or T-units. For example, when one of two writers writes one very long, accurate T-unit, the writer will score higher on W/EFT or WEFT than the other writer who composes more and shorter (on average) accurate T-units. Whereas the first writer is more fluent regarding accuracy, as the words per error-free T-unit measure shows, the second writer is more generally accurate, a result not indicated by these measures (Wolfe-Quintero, 1998).

On the other hand, Norris and Ortega (2009) criticized the previous conclusions on length-based measures for many reasons. First, they adduced that in spite of the length-based measures' lack of ability to specify how the lengthening of production unit is accomplished, such general and broad measures of complexity are needed, as they might be able to capture long-term or large-scale variation that would be overlooked by specific and finer-grained measures. Second, Wolfe-Quintero et al.'s (1998) proposal was based on the wrong interpretation of the study of Ortega (1995). Norris and Ortega (2009) emphasized that the

major findings of the previous study were reported in Ortega's (1999) other study; yet, the factor analysis was not included, since it was exploratory and was applied for the purpose of specifying repetitive measures. Thus, only five out of the ten measures used in Ortega (1995) were kept for the major analyses discussed in Ortega's (1999) study. Besides, for Norris and Ortega (2009, p. 568), Ortega (1995) was "cast[ing] doubts on the validity of using the utterance (intonationally and pausally defined…) as the dominator in complexity measures," rather than "concluding that length measures must be considered to tap fluency." When defining utterance as such, complexity and fluency are wrongly combined and such problems increased when measuring the speaking proficiency of low/intermediate level participants' syntactic complexity, because they could be capable of producing more or less complex language (a question of complexity) and simultaneously could or could not be capable of uttering them in one flow of speech, which is a question of fluency (Norris & Ortega, 2009). In fact, considering utterance as the dominator of length-based metrics makes it identical to the mean length of run that was identified by Tavakoli and Skehan (2005) as a promising measure for fluency (Norris & Ortega, 2009).

Another example that substantiates the validity of interpreting length-based measures for complexity is the study of Oh (2006) that was noted by Norris and Ortega (2009). In that study, Oh incorporated six measures of fluency: raw frequency measures (total number of T-units, clauses, and words), a rater measure (words per total minutes spent composing), and length-based measures (mean length of T-unit and clause). After applying Varimax factor analysis, it was clear that the first factor in the data is very highly loaded, having four (raw frequency and rater measures) of the six possible measures, whereas the other two measures, the mean length of T-units and clauses, are highly loaded on the second factor. Norris and Ortega (2009) maintained that "note that the first factor included measures that were computed by raw frequency counts as well as by ratio calculation. Thus, the possibility of a superficial commonality, such as the length-based calculations, explaining the results is clearly inadequate" (p. 569). Instead, they argued that fluency is reflected in the first factor while complexity is reflected by the second. If this interpretation is adequate, then such analysis proved the traditional consideration of length-based measures to be valid for complexity. However, CAF constructs, as mentioned before, suffer from other issues related to their linguistic and psycholinguistic correlations and the factors affecting them.

2.8 The Challenge in Determining CAF's Linguistic and Psycholinguistic Correspondences, Factors, and Correlations.

Although CAF constructs have been acknowledged and accepted to be distinctive dimensions of TL development, performance, and proficiency, the field is still in need of addressing other challenges: identifying correspondences between these constructs, the factors influencing them, and how they are correlated and interconnected. These three challenges are relatively linked to one another, because even if the CAF corresponds to different language processing aspects, this does not necessarily mean that factors may not affect them similarly or that they do not correlate and interact with one another. According to Housen et al. (2012), to specify a simple correspondence between CAF constructs in terms of linguistic and psycholinguistic processes and mechanisms—which underlie both their diachronic development in the duration of L2 acquisition and their synchronic manifestation during task production—is difficult, as each construct is multicomponential and multidimensional. Wolfe-Quintero et al. (1998) in the definitions cited for complexity, accuracy, fluency, claimed that from a linguistic standpoint, these constructs correlate with language in two separate features of language processing, namely language access (termed procedural knowledge, control of or access to representations, performance, or use) and language representation (termed declarative knowledge, linguistic representation, competence, or usage). From this viewpoint, accuracy and complexity indicate the present measure of the second language learner. Complexity demonstrates the range of expansion or restructured knowledge, whereas accuracy reveals how well that second language knowledge complies with the standards of the target language. Fluency, on the other hand, is an indicator of the amount of control the second language learner has in gaining access to that information. This control increases the more the learner automatizes the access-gaining activity.

For such speculative conjecture, Housen et al. (2012), provided a scenario of how the CAF's different language processing correspondences would interconnect in tandem during L2 development rather than its use. They stated that the acquisition of novel and more complex L2 structures will result in "more complex IL systems (i.e., greater complexity), followed by the" acquired structures modification "(leading to greater accuracy)." Moreover, this results, finally, in "the development of performance control over and consolidation of the acquired structures (resulting in more robust IL systems and more fluent L2 performance)" (p. 7). In spite of these authors acknowledging the intuitive plausibility of such a developmental interconnection sequence, they cautioned that it was both simplistic and speculative. Their

reasons for this caveat are that many studies that traced CAF development and/or incorporated internal linguistic factors—which need more verification and determination, such as specific formal and functional properties of a particular linguistic phenomenon and features (e.g., rules, patterns, constructions, items)—revealed that the language development aspects appear to be non-linear and do not correlate with one another in a fixed way. For example, Larsen-Freeman (2006) conducted a longitudinal research to examine the written and oral proficiency development of five high-intermediate level Chinese EFL learners regarding CAF. The qualitative and quantitative analyses showed that all students as one group became more complex, accurate, and fluent over time; nevertheless, each one of them had a different path in regards to the type of construct and the rate of improvement.

Gunnarsson (2012) examined the longitudinal development of CAF in the writings of five L2 French intermediate learners. The findings revealed that there were major individual differences between participants and that CAF followed separate developmental routes. The analysis of morphosyntactic characteristics (i.e., past tense, negation, subject-verb agreement, and clitic object pronouns) indicated that there was no correlation between syntactic complexity and fluency, though there was a correlation between accuracy and fluency that could vary depending on the structural complexity of these characteristics.

Koizumi and In'nami (2014) analyzed CAF spoken productions of 224 Japanese EFL junior and senior high school students whose proficiency levels were elementary and low intermediate. The results showed that despite correlating to varying degrees, CAF were independent constructs. It is true that learners were inclined to repair their speaking more and produced greater number of language units and clauses (sentences) and more accurate utterances when they tried to produce more words per minute. Nonetheless, the improvements of syntactic complexity, accuracy, and fluency were observed to be occurring gradually but not synchronously for learners progressing from beginning to low-intermediate levels. Furthermore, the findings revealed that there was a weak correlation between speed fluency and accuracy, whereas there were strong or moderate positive relationships between syntactic complexity and accuracy and syntactic complexity and speed fluency. Such varying degrees of correlation made the researchers suggest considering CAF constructs of speaking mode separately.

Instead, a lot of researchers followed the recommendation of Larsen-Freeman (2009) in investigating all CAF constructs to determine their interaction and the way they change with time. Most of these researchers, however, incorporated other types of factors (i.e., external [Housen et al., 2012; see Table 2 on page 66 for a list]), such as language task. Task-based

scholars were able to provide another complementary view of the correlation between learners' cognitive processing mechanisms, such as working memory, reasoning, automatization, and attention, and their CAF performances. From a psycholinguistic perspective, Towell (2012) contended that "the degree of success in attaining accurate, complex and fluent performances would be a product of successful interaction and integration between the growth of linguistic competence, the development of learned linguistic knowledge and the development of linguistic processing ability" (p. 66). Yet, a plethora of studies—which will be detailed afterwards— have indicated that not only is the interaction between CAF constructs complicated in the sense that it is sometimes competitive and sometimes mutually supportive, but that CAF constructs do not develop collinearly (Housen et al., 2012). The justifications provided by Ellis (1994, as cited in Housen et al., 2012, p. 7) are that L2 development is influenced in many ways by explicit and implicit knowledge, the fact that knowledge analysis and automatization experience differential development in L2 acquisition, and finally that "the psycholinguistic processes involved in using L2 knowledge are distinct from acquiring new knowledge."

Again, with the task-based researchers commenting on CAF construct developments and correlations, Sasayama (2015, 2016) maintained that there are still some problems regarding the fact that the proposed models for cognitive task complexity designing and sequencing were based on different and unique theoretical orientations, and as a result, presented different predictions for the role of cognitive complexity in language acquisition and task productions. The testing of such models, as emphasized by Housen et al. (2012), has been difficult partially because the operational and conceptual clarity of CAF is lacking, and even the studies that did so showed inconsistent findings (Sasayama, 2016).

In other words, one of the major issues encountered by authors concerned with capturing the effect of task features and conditions on L2 learners' CAF productions is exploring how to precisely identify the level of task complexity (Elder, Iwashita, & McNamara, 2002). Many attempts have been made by task-based pedagogy researchers to address the former. Interestingly, two influential models have come to surface. The first model, the Limited Attentional Capacity Model, was developed by Skehan and his associates (e.g., Skehan, 1996a, 1996b, 1998, 2003a, 2003b, 2009, 2013; Skehan & Foster, 1997, 1999; 2001, 2005, 2012; Foster & Skehan, 1996, 1999). On the other hand, the second model, the Triadic Componential Framework, was proposed by Robinson (2001a, 2001b, 2003, 2005, 2007, 2011a, 2011b, 2015; see also Baralt, Gilabert, & Robinson, 2014; Robinson & Gilabert, 2007). Although all of these scholars agreed on the idea that (a) when transacting tasks, L2

learners make choices that are not neutral but dependent on the types of the tasks (Skehan & Foster, 2001); and (b) their models hypothesize how to determine task complexity factor, they still offer competing views on how the manipulation and sequencing of tasks' cognitive characteristics influence L2 participants' CAF and how their attention is deployed during executing their performance.

2.8.1 Skehan's limited attentional capacity model.

Skehan and Foster (2001) defined task complexity as "the amount of attention the task demands from the participants. Difficult tasks require more attention than easy tasks" (p. 196). Skehan (2013) argued that in order to understand how a task might be performed by learners and in order to allow them to develop all CAF constructs simultaneously, it is very crucial to assess and sequence its complexity level. Such assessment and sequencing, according to Skehan (1996b, 1998), can be undertaken through three principal areas: language, cognition, and performance conditions. Under each of these areas, there is one complexity dimension: code complexity, cognitive complexity, or communicative stress, with factors that specify its level. Skehan's Limited Attentional Capacity Model is presented in Table 3.

Table 3

Skehan's (1998, p. 99) Limited Attentional Capacity Model

Code complexity	Cognitive complexity	Communicative stress
• Linguistic complexity and variety • Vocabulary load and variety • Redundancy and density	Cognitive familiarity • Familiarity of topic and its predictability • Familiarity of discourse genre • Familiarity of task Cognitive processing • Information organization • Amount of 'computation' • Clarity of sufficiency of information given • Information type	• Time limits and time pressure • Speed of presentation • Number of participants • Length of texts used • Type of response • Opportunities to control interaction

In the table, the first area (language: code complexity) is related to the linguistic and vocabulary demands of the task. The task will be regarded as more complex if it requires learners to use greater densities or a more advanced repertoire of structures and/or lexicons. The second area is cognition (cognitive complexity), and it distinguishes between cognitive familiarity and cognitive processing. The cognitive familiarity refers to the extent to which

the task takers can rely on their previous experiences when performing such tasks and similar ones. Therefore, if the task is novel to the learners, then it is labelled as more complex. The cognitive processing is concerned with whether a task requires the task taker to think to perform it or not. The more steps needed to complete a task, or the more the learners need to organize the information, the more taxing it will be. The third area, performance conditions, deals with the communicative stress of the task and comprises some factors. Time pressure is related to the amount of the allotted time in which the learner has to perform the task; the shorter it is, the more complex the task will be. The scale factor refers to the number of participants and relationships in the task, as increasing the number of such will contribute to making the task more complex. Modality factor concerns the mode of the task: speaking, listening, reading, or writing. Speaking is believed to exert more pressure on the speaker than writing, and so is listening in comparison to reading. Stake factor deals with the significance of executing the task and of executing it well; the greater the stakes factor, the more demanding and therefore more complex the task is. Lastly, control factor deals with the extent to which the participants of the task can affect the execution of the task. A task in which students are not permitted to ask clarification questions are more taxing than those in which they have such a privilege.

However, Skehan (1996b, 1998, 2003a, 2003b, 2009, 2013) emphasized his basic assumption stating that since L2 learners are not like native speakers in terms of having multiple attentional and working memory capacities, the various sources of task complexity—if not sequenced carefully when designing—will result in an individual or combined effect between CAF constructs. Foster and Skehan (1999) stated that "learners have limited capacities, which they deploy selectively, reflecting whatever performance priorities that they have or that the tasks and task conditions support" (p. 221). Skehan (2003b) maintained that "on some occasions, task characteristics and task conditions can prioritize new language, and risk taking; on the other occasions they can predispose conservatism and error avoidance; and on others push learners to gain fluent control over aspects of target language" (p. 394). In other words, depending on the task features and/or conditions, a trade-off influence on CAF constructs will take place, preventing them from all developing simultaneously. The trade-off effect can occur (a) between content/meaning and language linguistic aspects, which leads learners to shift their focus on fluency (increase) at the expense of their focus on complexity and accuracy (decrease); or (b) between meaning and only one linguistic aspect of language, resulting in increases in fluency and accuracy or fluency and complexity. Skehan (2009, 2013) pointed out the possibility of another trade-off impact occurring between the language

aspects themselves, as they could be competing for attentional resources: the raised level of prioritization on accuracy will deplete prioritization on complexity and vice versa. Yet, according to Skehan and Foster (2012), both accuracy and complexity can address the limitations of the trade-off and increase at the same time only if the task is modified to certain features and conditions. Robinson also predicted such an increase in both constructs but provided different justification: "For Robinson, task complexity is the driver. For Skehan, it is the combination of task characteristics and task conditions" (Skehan & Foster, 2012; p. 202; Skehan, 2009). The later researcher believed that altering only task complexity demands will cause all CAF constructs to degrade simultaneously, which is again different from what Robinson hypothesized.

2.8.2 Robinson's triadic componential framework.

Robinson (2007) maintained that the factors of code complexity, cognitive complexity, and communicative stress, which were proposed to highlight the task features that influence "the 'difficulty' of task," despite being "intuitive, [and] often insightful," they "lack of cultural knowledge, confidence and motivation…; and the transition from easier information gap, to reasoning gap, to more difficult opinion gap activities" (p. 14). Thus, he proposed the Triadic Componential Framework (see Table 4).

Table 4

Robinson's (2007, pp. 15-16) Triadic Componential Framework

Task complexity	Task condition	Task difficulty
(Cognitive factors)	(Interactive factors)	(Learner factors)
(Classification criteria: cognitive demands)	(Classification criteria: interactional demands)	(Classification criteria: ability requirements)
(Classification procedure: information-theoretic analyses)	(Classification procedure: behavior descriptive analyses)	(Classification procedure: ability assessment analyses)
Sub-categories: (a) Resource-directing variables making cognitive / conceptual demands	Sub-categories: (a) Participation variables making interactional demands	Sub-categories: (a) Ability variables and task relevant resource differentials
+/− here and now	+/− 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 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
+/− prior knowledge	+/− same gender	h/l control of emotion
+/− single task	+/− familiar	h/l task motivation
+/− task structure	+/− shared content knowledge	l/h processing anxiety
+/− few steps	+/− equal status and role	h/l willingness to communicate
+/− independency of steps	+/− shared cultural knowledge	h/l self-efficacy

In the framework, not only did Robinson (2007) distinguish between three broad dimensions—task complexity, task conditions, and task difficulty—but each of the three dimensions could be further classified into subcategories "having a systematic hierarchical relation to each other;" cognitive criteria, interactional criteria, and ability-determinant criteria (p. 14). The reasons are that the impact of these dimensions on task performance and learning is different in kind (Robinson, 2001a), and researchers would be able to study them separately or all together in a manageable way, as they would have complex interactions with one another (Robinson, 2005; Robinson & Gilabert, 2007). That is, all of these dimensions could be used to complement one another because each of them has a specific purpose.

In terms of task difficulty, Robinson (2001a) contended that "it concerns learners' perceptions of the demands of the task" (p. 295). He (2001b) stressed that learners' factors that contribute to making a task more or less difficult (as opposed to complex) must be differentiated from cognitive factors (task complexity). In the learners' factors, there are two types of variables: ability and affective. The ability variables include working memory, reasoning, task-switching, aptitude, filed independence, and mind-reading. Robinson (2007) maintained that these variables are strongly correlated with the learners' perceived difficulty of performance on task manipulated along cognitive demands (task complexity). Also, Robinson (2001b) believed that such variables can be predicted in advance of applying a task, and that over a course of instruction, they are more stable, permanent, and fixed determinants of resource pools than effective variables. The latter variables, which contain openness, control of emotion, task motivation, processing anxiety, willingness to communicate, and self-efficacy, appeared to be related to the students' perceived difficulty of performance on a task increased along interactional task demands (task conditions [Robinson, 2007]). Robinson (2001a) stated that to diagnose task difficulty based on affective variables before learners' actual engagement with the task is impossible, the fact that these variables "can sometimes be unpredictably influenced by participants variables" (i.e., those in task conditions [p. 295]). Robinson (2001b) described affective variables as being susceptible to change, and on a temporary basis they may impact the size of resource pool availability. Two learners undertaking the same task may perform differently as a consequence of having different levels of ability variables (e.g., aptitude) or affective variables (e.g., processing anxiety). The learner with high aptitude and low processing anxiety would perceive the task as simple, whereas the learner with low aptitude and high processing anxiety would perceive it as difficult. Therefore, Robinson (2001a) emphasized that the task difficulty should aid explaining the diversity in task performance between any two learners undertaking the same task. Yet, he suggested that it should not take any role in the task sequencing decisions (i.e., from simple to difficult), even though he (2001b) acknowledged that there is still no clear research evidence that substantiates the interactions between task complexity, performance, and learners factors.

Regarding task conditions, Robinson (2003) said that they "concern…the interactive demands of task performance" (p. 56). The demands can be divided dependent on participation factors and participant factors. Robinson (2007) clarified that the participation factors include (a) whether the solution to the task is open (optional) or closed (fixed); (b) whether information exchange goes one way (from A to B) or two-way (reciprocal); (c) whether agreement is convergent (required) or divergent (opposite); (d) whether the

interaction has few or many participants; (e) whether one, few, or all of the participants can contribute to it; (f) whether the contribution to the interaction demands no, little, or extensive negotiation. On the other hand, the participant factors contain (a) whether learners have the same or different gender and proficiency levels; (b) whether they are familiar or unfamiliar with each other; (c) whether they share or do not share knowledge of the domain or relevant cultural knowledge about how the interactions are conducted in the L2; (d) whether they have the same role in a task regarding position in the workplace, seniority, status, and social/institutional standing. Robinson (2001a) argued that the previous factors "are unlikely to be a useful basis for a priori sequencing decisions, since they will largely have been specified on the basis of the needs analysis, and fidelity to the target task performance the pedagogic tasks are aiming to facilitate" (pp. 295-296).

Most importantly, Robinson (2011a) defined task complexity as "the intrinsic cognitive complexity of tasks" (p. 14), while elsewhere (2001b) his explicit interpretation was "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). For Robinson (2001a), task complexity should be used for sequencing tasks from simple to complex because they would help explain the performance diversity in the simple and complex tasks within a learner. Within the framework, the Multiple Attentional Resources Model emerged. In the model, Robinson developed what he called the Cognition Hypothesis, in which distinctions between the task demands that differentiate learners' performance and those which stimulate their development were made. Robinson (2005) clearly stated that the model

> Distinguishes between dimensions of task complexity which can be manipulated to increase the conceptual and linguistic demands tasks make on communication, so creating the conditions for L2 development, and the dimensions of task complexity which can be manipulated to increase the demands made on accessing a current interlanguage repertoire during real-time L2 performance. (p. 3)

In other words, Robinson (2003) provided two subcategories for task complexity— resource-directing and resource-dispersing—each of which would influence students' production differently. Robinson (2011b) claimed that

> Increasing complexity along resource-directing dimensions promotes greater analysis, and representational redescription of L2 conceptual-linguistic knowledge, and form-function mappings, while increasing complexity along resource-dispersing dimensions promotes greater control over, and faster access to, existing interlanguage systems of knowledge. (p. 17; see Robinson, 2015).

In the former subcategory, task complexity can be altered along "cognitive/conceptual" demands (Robinson, 2007, p. 17). The task that asks students to (a) refer to an event occurring now (Here-and-Now); (b) take just one first-person perspective on an event; (c) address few easily distinguished elements, easily identifiable spatial locations, or simple information transmissions is easier than that which requires learners to (a) refer to events happening in the past (There-and-Then); (b) take multiple second and third person perspectives on an event; (c) address many elements or a novel location; (d) provide causal or intentional reasoning (Robinson & Gilabert, 2007). In the latter resource-dispersing subcategory, task complexity can be increased along "performative/procedural" demands (Robinson, 2007, p. 18). Here, the simple task would either (a) allow planning time; (b) provide background knowledge about the task or a clear structure to help on deciding which steps are needed to complete it; (c) require only one thing to be done, one or few steps to complete it, or no necessary sequence of steps to be followed. The difficult task, on the contrary, would not allow planning time and not provide background knowledge about the task or a clear structure by which students can decide which steps are needed for task completion. It might also require dual or multiple things to be done simultaneously, multiple steps to be completed, or follow a chain of steps in which one step must be undertaken before another.

Robinson (2011a) stated that elevating task complexity on resource-directing dimensions will impact the allocation of cognitive resources to particular features of L2 code. For instance, increasing task complexity in terms of intentional-reasoning demands would direct the student's attention to linguistic references (i.e., psychological state terms such as believe, think, and wonder) that are used to describe the mental states of others. These terms may be unknown or known, but nonetheless not well controlled by students, and when they attempt to complete the task, these terms may become "salient and 'noticeable'" (p. 15; Baralt et al., 2014). As such, both complexity and accuracy will increase, whereas fluency will decrease. Conversely, instead of directing learners' attention to particular features of L2 code, altering task complexity along resource-dispersing dimensions (e.g., planning) would disperse their attention and memory resources over many linguistic and non-linguistic features (Robinson, 2011a). The lack of students' practice on removing such a process demand (Baralt et al., 2014) would cause these dimensions to "create…problems for learners attempting to access their current repertoire of L2 knowledge," and therefore, result in deteriorations in complexity, accuracy, and fluency (Robinson, 2005 p. 7). Finally, in spite of the fact that the predictions of Skehan and Foster and those of Robinson have been investigated in many

studies, there is other research that addressed text type factor, all of which are closely related and cited below.

2.8.3 CAF studies on writing tasks and task complexity factor.

Task complexity and text type features are both linked since the latter factor—always with specific aspects—has a definite level of intricacy. Therefore, a difference in L2 production will be caused by various tasks (Rahimpour, 2007). According to Kuiken and Vedder (2008) the explanation for this is that a study by Hamp-Lyons and Mathias (2008) examining professional evaluations of how task difficulty related to ESL composition test grades showed that although the information derived supported the soundness of the expected connection between task type and production, the relationship was, in fact, the opposite of what was predicted. Contrary to the usual presumption that expository and personal prompts would be easily done, they were actually shown to be associated with the lowest writing grades. On the other hand, argumentative and public prompts produced the highest grades. The authors speculated that one cause for this was when a more cognitively difficult composition task is assigned, students are motivated to access more writing ability than they would had a less cognitively complex task been assigned.

However, some researchers seemed to treat the two factors differently in their investigations. In some studies, researchers were explicit about the writing tasks they used. By examining such factors, these researchers hope to, as argued by Sweller (1994) and Franken and Haslett (2002), be capable of describing exactly how learners can automatize specific features (CAF) of the writing tasks and deal with the additional load to process these features.

For instance, Bonzo (2008) conducted a study on the effect of teachers' assigning topics and students' free choice of topics on the fluency and complexity of 81 nonnative learners of German written production. These participants were in four classes. In the first four sessions, two groups were required to write on four specified topics, whereas the other two groups chose theirs. During the last four classes, the order of the topic treatment changed in the sense that the first two mentioned groups' participants were asked to self-select their writing topics, while the instructors of the other two groups assigned topics. The statistical analysis showed that only fluency was significantly higher in the students' self-selected topics than in the teacher-assigned topics. The partial replication—as complexity was not investigated— of this study by Sponseller and Wilkins (2015) revealed that the 75 EFL Japanese undergraduate had similar results regarding fluency.

Ferreira (2013) examined the influence of self-selected and assigned topics on the written production fluency of 47 ESL Japanese undergraduates. The participants were not randomly assigned to two groups but were recruited from two first-semester standard introductory English classes in writing and reading that were taught by the researcher. In four sessions, each group was asked to either choose one topic or to do the topic that had already been chosen by the teacher. The findings indicated that the writing fluency increased more significantly in self-selected topics than in those assigned by the teacher. Such results were substantiated by Dickinson's (2014) research of 46 Japanese EFL learners.

Pourdana and Behbahani (2011) researched the impact of text types (topic writing, picture description, and text reconstruction) on the CAF of 65 undergraduate Iranian EFL learners' writing production. Each of these subjects was required to undertake all of these tasks. The findings showed that in terms of the topic writing task, participants gained significantly in accuracy and syntactic complexity (but not fluency) in comparison to the other types of writing tasks. Fluency was only significantly high when students performed the picture description task. These results proved that different types of composition tasks would lead to different effects on CAF, a finding that was also confirmed by the findings of Pourdana, Behbahani, & Safdari's (2011) study that recruited 80 Iranian EFL undergraduates.

Rezazadeh et al. (2011) studied the role of task types (i.e. instruction task and argumentative task) in foreign language CAF written production. Voluntarily, 168 intermediate level EFL students participated in the study, all of whom were registered in two writing courses at two Iranian universities. These participants were divided into two groups: 74 students in the instruction writing task condition and 94 students in the argumentative writing task condition. In the instruction condition, learners were required to write a message to a friend who was going to visit them in their city, describing how to reach the university from the train station. Nonetheless, participants in the argumentative condition were asked to state their opinions on the harmful influences of examinations on education. The results revealed that while subjects in the instruction-task group performed significantly better in accuracy and fluency than the argumentative-task group, the latter group produced more complex language than the former group.

In a doctoral dissertation, two parts of Jong's (2009) investigations were (a) the correlation between writing task types and 13 Korean's EFL primary school task performances and outcomes, and (b) these participants' perception towards these task types. The writing tasks incorporated were a picture-describing task of a family member, furniture, and an animal; a story-creating task of school events; and an opinion-expressing task on

particular issues addressed in a news article. All participants with self-selected partners did these three tasks. After that, they all attended an interview in pairs. The analysis of these interviews and writings indicated that learners considered the opinion-expressing task to be the most challenging, whereas the picture-describing task was considered to be the least challenging task. Also, there was evidence of a trade-off between fluency, accuracy, and lexical complexity in the sense that students were the most fluent in the picture-describing task and the least so in the opinion-expressing task, while they were the most accurate, lexically varied, and dense in the opinion-expressing tasks and the least so in the story-creating task.

Li (2000) aimed at capturing the impact of incorporating task-based email activities in a writing course of 22 ESL learners. For the completion of the writing course, all participants were required to write four email writing tasks (narrative, cultural comparison, persuasive, and reaction), each in one week. These tasks are relevant to university-level academic essays, but they differed from each other in terms of purpose, audience, interaction, and task structure. For example, cultural comparison and persuasive essays would involve interaction (feedback) between the writers and their peer audience, whereas in narrative and reaction essays, the audience (readers) would be passive. Also, in narrative and persuasive essays, students were guided with questions, while in cultural comparison and reaction essays, they were provided with only general comments on their instructor's expectations. In the latter two tasks, learners had the freedom to write whatever topics and ideas they want, while in the former two tasks, the topics were assigned by the teacher. The findings showed that there were significant differences between the four various tasks (i.e., regarding purpose and features) in syntactic complexity, lexical variation and density, and grammatical accuracy. Participants produced more complex sentence structures with a greater number of grammatical errors in persuasive text than in narrative text, which proved the existence of a trade-off effect between complexity and accuracy. The results furthermore revealed that there were greater levels of lexical variation, density, and errors in writing tasks with audience interaction than in writing tasks with no audience interaction. Finally, the students' productions were more lexically varied and dense, and they displayed longer and more complex sentences in the non-structured tasks than in the structured tasks. Again, there was a trade-off effect between accuracy and linguistic complexity.

On the other hand, following in the footsteps of scholars such as Skehan and Foster and Robinson, researchers in other studies were more concerned with how enhancing task

complexity (along the resources-directing and/or the resource-dispersing dimension) can affect the learners' CAF production.

For example, Kuiken et al. (2005) tested the influence of task complexity on various aspects of L2 written linguistic performance of low and high proficiency level Dutch learners of Italian. Participants were 62 students attending the University of Amsterdam. The researchers adopted three models for this investigation: Robinson's Cognition Hypothesis, Skehan and Foster's Limited Attentional Capacity Model, and Cummins' Threshold Hypothesis (Cummins 1979). In the two experimental tasks, students were given questions in their native language and were asked to answer using Italian. Task complexity of the prompts were in Dutch and were manipulated by assigning different numbers of elements. The task was about choosing one destination choice out of five and then writing a letter to a friend recommending this as a specific place to spend the holiday together. Participants were asked to base their choices on the number of different criteria assigned to each destination. These criteria differed according to the type of task (i.e., six in a complex task, and three in a simple task). The findings indicated that in the more demanding task, written samples were more accurate with significantly lower error ratios per T-unit than those in the less taxing task. This gave partial support to the Cognition Hypothesis. However, no impact by task complexity manipulation on lexical variation and syntactic complexity were found. The results, in addition, showed that there was significant correlation between the manipulation of task complexity and learners' proficiency levels. The higher the task complexity and proficiency level, the greater the writing accuracy. This could be interpreted as supporting the Threshold Hypothesis, as learners with low proficiency could not progress in the more demanding task due to insufficient command of their L2 (Italian).

Likewise, Kuiken and Vedder (2007a) examined the effect of task complexity on different elements of L2 written production at various levels of proficiency. The 76 Dutch subjects were first and third year learners of L2 French. Yet, their proficiency was determined by a cloze test. The findings revealed no support to Skehan and Foster's model, since participants' performance in the less cognitively demanding task was not significantly better than in the more cognitively demanding task. Nevertheless, Robinson's model was partially supported, as texts in the complex task included fewer errors (more accuracy) and greater lexical variation than those in the simple one. Because (WT/W), but not (WT/√2W), measure indicated such progression in lexical variation, the researchers decided not to "attribute too much importance to that" (p. 130). Finally, there was no relationship between task complexity and proficiency level. Thus, Cummin's Threshold Hypothesis was rejected, as such a result

contradicted the researchers' earlier results (i.e., Kuiken et al., 2005). The justifications provided for such a discrepancy were that the selection of participants in Kuiken et al. (2005) was biased because the proficiency level (low and high) differences were larger than in this study. Moreover, it is a fact that the Italian students had very poor command of the language, which was the opposite of the French students who had been exposed to the language for several years prior to their university enrollment.

Moreover, with more participants and specific, rather than general, measures, Kuiken and Vedder (2007b) reanalyzed the second language writing skills proficiency of 75 Dutch undergraduate students of French and 84 of Italian. The analysis of data collected showed that students of French had fewer orthography errors and significantly more appropriateness and other errors in the complex task, while no differences were found for the students of Italian. The authors maintained that it was not clear why the former group had more appropriateness errors, knowing that it was more proficient than the latter group (Italian). Regardless, both of these groups had fewer lexical errors in the difficult task, which means that the increase of accuracy resulted from the decrease of such errors. Regarding lexical variation, the students of French had more infrequent words in the complex task, whereas the students of Italian had more high frequency words in the above task and vice versa in simple task. These findings supported both the Cognition Hypothesis in the case of the French students and the Limited Attentional Capacity Model in the case of Italian subjects. Finally, there was no correlation influence between task complexity and proficiency level.

Similarly, Kuiken and Vedder's (2008) studied the correlation between L2 written linguistic performance and cognitive task complexity by embracing two models mentioned before. The participants of the experiment were 91 Dutch university student of Italian and 76 students of French, all of whom had been divided into two groups according to their L2 proficiency levels—either high or low. The results were in line with the Cognition Hypothesis in the sense that the compositions of participants in the more taxing task appeared to have more accuracy (i.e., significantly lower error ratio per t-unit) but not more lexical variation or syntactic complexity than those in the less taxing task. Finally, the findings revealed no correlation between task types and proficiency levels. Such an outcome was contrary to the researchers' earlier finding (in 2005) where high-proficient learners had stronger accuracy when the task was complex; nonetheless, this was in line with their research results (2007a, 2007b).

Kuiken and Vedder (2011) continued their long-lasting interest in researching the impact of the Cognition Hypothesis and the Limited Attentional Capacity Model on L2

learners' production. However, in this study, two types of language modes were used in the investigation. In other words, the researchers compared written texts executed by 91 Italian L2 students with oral texts produced by 44 Italian L2 students. In each mode, participants were divided into two groups based on proficiency levels. All of them were required to undertake the same tasks. The findings indicated that in speaking and writing modes, task complexity affected only accuracy (less lexical errors) and had no effect on lexical variation or syntactic complexity, which partially support of Robinson's Cognition Hypothesis. The results also proved that there was no influence of task complexity on the linguistic performance of students with different proficiency levels, neither in speaking or writing mode.

Abdollahzadeh and Kashani (2011) studied the impact of task complexity on Persian EFL learners' narrative writing CAF. The task complexity was manipulated along the complexity dimensions (i.e., simple: here-and-now and complex: there-and-then). According to the analysis of a TOEFL test administered to the 107 participants, four groups were generated. Two groups of 16 students served as high proficient level (HPL), while the other two groups of 18 learners were regarded as low proficient level (LPL). All participants in the above groups were shown a set of nine frames forming one story for five minutes. Then, a here-and-now simple task prompt was distributed among one group of high proficiency level and one group of the low proficiency level. The question was to produce a narrative writing using the present tense. Nonetheless, the other two groups in HPL and LPL were assigned a there-and-then complex task to compose a story using the past tense. The findings revealed a significant effect of task complexity and language proficiency in terms of complexity and accuracy. Higher-proficiency learners had more accuracy and complexity in the complex task. Finally, regarding task complexity, language proficiency, and fluency, no significant influence was found.

Ishikawa (2006) investigated the impact of language proficiency and task complexity on L2 narrative written productions of 52 Japanese high school students. In this study, participants were divided equally into four low- and high-proficiency groups. For all groups, complex and non-complex task conditions were assigned. The results indicated that in the difficult task, the low-proficiency level groups experienced a greater increase in accuracy, syntactic complexity, lexical variation, and fluency than in the simple task. The same findings were observed for the high-proficiency level groups, except for lexical variation, which decreased in the complex task. The results also showed that the low-proficiency learners had greater gains when manipulating task complexity from simple to difficult, because their

aspects (CAF) of production in the complex task were not inferior to those of high-proficiency students in the non-complex task. Specifically, the low-proficiency participants gained about double the benefits in using the target-like English articles (accuracy) in comparison with high proficient subjects (a growth of 12% vs. 6%). All of these findings revealed that the correlation between task complexity and proficiency appeared to be independent in fluency and syntactic complexity and dependent in accuracy and lexical variation.

Ishikawa elsewhere (2007) aimed at specifying the effect of increasing task complexity on the narrative writings of 54 Japanese third-year high school students. These participants underwent an English placement test and were then divided randomly into two task conditions—Here-and-Now and There-and-Then—with each condition containing an almost equal number of low- high-proficiency students. The results indicated that the There-and-Then group produced more accurate, syntactically complex, lexically varied, and fluent texts than the Here-and-Now group.

Nikou and Eskandarsefat (2012) tested the influence of task complexity (simple and complex) and psycholinguistic types of tasks (decision-making, and information-gap) on the written performance (i.e., CAF) of 60 EFL intermediate learners. During the five classes in which the study was conducted, participants were given a simple decision-making task and were asked to write depending on the fire-chief tasks that were distributed to them. Every two students were then randomly placed in groups, each of which received a simple, different information-gap task that had to be performed without seeing each other's pictures. One learner asked some questions of his or her partner, according to his or her own picture and wrote the answers. Depending on these answers, he or she was required to provide a piece of writing. The same thing was done by the other partner. Two weeks later, the same process and application took place, but for the complex task. The findings showed that in decision-making, there was positive impact of task complexity on EFL learners' writing accuracy and fluency, but not on their syntactic complexity, which was believed to be better in simple tasks. Regarding the information-gap task, task complexity had a significant effect on the accuracy, fluency, and complexity of the subjects' writings. Also, the results revealed that in simple and complex decision-making tasks, subjects were more fluent than in the simple and complex information-gap tasks. Nevertheless, no significant difference was found towards these different tasks in terms of accuracy or syntactic complexity.

Sasayama (2011) analyzed the influence of cognitive task difficulty on 10 Asian (Korean and Japanese) ESL learners' spoken and writing production. As exchange students

from their home universities, these learners were at the University of Hawai'i. In each mode, there were simple and difficult tasks. Although in all of these story-telling tasks a set of pictures addressing different topics were involved, the levels of task complexity varied depending on these tasks requirements. For example, in the simple speaking task there was one character that students have to address, while in the more difficult task there were multiple characters. Furthermore, in the simple writing task students had to explain foreground information (main points), whereas in the difficult task they had to also link this foreground information with background events (i.e., comment, exemplify, and assist). The findings indicated that there was partial disconfirmation and partial support of the Cognition Hypothesis, as the complex speaking task did not elicit more syntactic and accurate performance, while complex writing task elicited more syntactically complex, yet not more accurate, texts. In fact, accuracy not only remained at the same level in all four tasks, but it also did not deteriorate as a consequence of participants' elevated syntactic complexity in the difficult writing task only.

Rahimpour and Hosseini (2010) examined the impact of altering task complexity on 52 Iranian EFL learners' narrative writings in terms of fluency, accuracy, and complexity. All of these participants performed two composition tasks that were based on a picture story but differed in their levels of complexity. The complexity was increased according to the context and the tense used. In other words, in the context-supported simple task students were given a picture with a prompt written in the present tense and were asked to rely on the picture while writing using the same tense of the question. In the context-unsupported difficult task, learners were given another picture strip to observe and analyze for five minutes with a prompt written in the past tense. Then, the picture story was collected and students were required to compose using the same tense of the question. The results showed that the more demanding task led to more fluency, but it did not lead to greater accuracy or syntactic complexity.

Jiaxin (2015) conducted a study on the effect of manipulating task complexity on 31 non-English major Chinese undergraduates' argumentative writings. These learners were asked to complete two writing tasks that differed in their levels of complexity within two weeks. The findings revealed that when increasing the task complexity, participants' fluency, syntactic complexity, and accuracy decreased. However, their lexical variation developed.

Salimi and Dadashpour (2012) researched the influence of enhancing task complexity on the compositions of 29 Turkish students majoring in Translation and TEFL. Based on a picture, these participants were asked to first perform a simple decision-making fire task.

Then, after two weeks, they were required to do a more difficult decision-making fire task. The reason for such a delay between the two tasks was to lessen the memory and task repetition impact on the participants' production. The results indicated that there was no significant effect of task complexity on accuracy. Nevertheless, the manipulation of the task complexity resulted in more fluent and complex language.

Frear and Bitchener (2015) studied the influence of cognitive task complexity on the written lexical and syntactic complexities of 34 intermediate proficiency level nonnative speakers of English. All participants completed three letter-writing tasks that varied in their levels of complexity (i.e., low, medium, and high). The interval time between performing the first task and the second task was three to five days, whereas it was only five minutes between executing the second task and the third task. The analysis of the collected data showed that "a significant effect for task complexity on decreases in syntactic complexity using a ratio of dependent clauses to T-units measure where independent clauses were measured separately. In contrast, significant findings were found for increases in lexical complexity [i.e., variation], analyzed as a mean segmental type-token ratio" (p. 45).

Ismail, Abd. Samad, Eng, and Noordin (2012) investigated the impact of altering task complexity along resource-directing (+/- reasoning demands) and resource-dispersing (dyadic and individual) on the grammatical accuracy and syntactic complexity of 76 EFL high school students in Selangor. These participants were divided into four groups and were asked to perform dictogloss tasks (simple) and opinion-gap tasks (difficult) in pairs and individually. The results revealed that learners' accuracy scores were the highest when they performed the simple task in dyad (M = .70) and individually (M = .61), than when they performed the difficult task in dyad (M = .47) and individually (M = .40). In both tasks, dealing with the task in pairs resulted in high accuracy scores. Conversely, the syntactic complexity scores were the highest when participants performed the difficult task in dyads (M = 1.84) than individually (M = 1.43), whereas the construct scores were higher (but not significantly) when the participants performed the simple task individually (M = 1.45) than in dyads (M = 1.43).

Masrom, Alwi, and Daud (2015) tested the effect of increasing task complexity on the syntactic and lexical productions of 88 Malaysians EFL undergraduates. Participants were divided into four groups, each of which served one writing condition. The writing task (miscommunication at workplace) was manipulated according to causal reasoning demand (+/- CRD) condition and task instruction (+/- TS) condition. The findings indicated that the highest increase of syntactic complexity was in (+ CRD / - TS), then (- CRD / - TS) and (- CRD / + TS), whereas the lowest was in (+ CRD / + TS). The biggest development in lexical

variation was in (– CRD / –TS), followed by (+ CRD / – TS) and (- CRD / + TS), while the lowest development was in (+ CRD / + TS). For lexical sophistication, the highest improvement was in (+ CRD / + TS), followed by (+ CRD / – TS) and (- CRD / - TS), but the lowest improvement was in (- CRD / + TS).

Mohammadzadeh Mohammadabadi et al. (2013) researched the influence "of simultaneous use of pre-planning along +/- here-and-now dimension of fluency, complexity, and accuracy of [30] Iranian EFL learners' written performance" (p. 49). All subjects were put into four experimental conditions. This means that two types of complexity tasks (simple vs. complex) with two kinds of planning (no-planning vs. pre-task planning) were applied to students. The application took place in two sessions using four wordless-picture stories of topics familiar to the participants' culture. The results showed that planning in both complex and non-complex conditions enhanced accuracy more than non-planned here-and-now and there-and-then. In the four various tasks, participants' performances were similar to one another in fluency and complexity.

2.9 Summary

This chapter discussed in detail how and why the CAF constructs were originated and used in tracing the development of L2 learners. It also included a thorough elaboration of the challenges that researchers encounter when incorporating such constructs. These challenges emerged from the constructs' proposed various definitions and measurements, their linguistic and psycholinguistic correlations and interconnections, as well as the factors influencing them such as task complexity. The type of and the justification for task complexity effect on CAF constructs appeared to be different between task-based scholars as shown in the cited studies.

Chapter Three
Methodology

3.0 Introduction

This chapter explores the methodology applied in the study to investigate the effect on CAF in the written production of Saudi EFL university students in three learning environments, and to determine which of these environments will be the most/least effective in developing such constructs. It contains the statement of the problem, research design, setting, participants, courses, teaching approach, and teaching and learning contexts. It also includes details of the materials, instruments, and procedures for data collection, ethical considerations, and data analysis.

3.1 Statement of the Problem

The purpose of the present research is to determine the difference between traditional, blended, and online learning contexts in the case of CAF development in written language among Saudi university students. We will also explore the impact of four different factors: proficiency levels, learning contexts, text types, and times on such constructs. The research questions investigated in the study were:

1. How are the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates in the three learning environments: TLC, BLC, and OLC, affected longitudinally across two writing tasks (classification and argumentative) that differed in their level of complexity? And when?

2. Which of the three learning environments: TLC, BLC, and OLC, lead to the most/least increase or decrease in the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates across two writing tasks: classification and argumentative? And when?

3.2 Research Design

This study adopts a combination of exploratory and longitudinal approaches. The purpose of choosing an exploratory design is that it is beneficial for "'unpack[ing]' issues or topics, identify[ing] a problem, clarify[ing] the nature of it and defin[ing] its scope" (McGivern, 2009, p. 73). A longitudinal approach in general allows researchers to "monitor things—attitudes, behavior, experiences, [and] perspectives—over a period of time…from the same sample" (p. 84). It also, as maintained by Phakiti and Paltridge (2015, p. 12), helps investigators to observe stability or change in learning, abilities, behaviors, and/or other

social/cognitive development, and "establish sequences of events or changes." But, specifically, the prospective method helps to gather information of changes almost once they occur and at micro levels (Dörnyei, 2007). In other words, in order to provide a deep understanding of the current study's goals and objectives, answer its questions accurately, and determine the feasibility of conducting research on this area and context, the adoption of such an approach is believed to be adequate.

However, like any other research designs, the exploratory method suffers from one main disadvantage related to its inclination to incorporate a qualitative design, especially if the study is in its early stages (Schilderman, 2012), that is, exploring a topic in a novel context, which would normally require two phases of (qualitative, and then quantitative) data collection (Creswell & Clark, 2007). Instead, in this research, a quantitative approach was used since recruiting a large number of Saudi EFL participants was possible. Further, the theories of and the instruments to capture CAF development in writing, along with the factors affecting these dimensions, have already been established (Schilderman, 2012). There are also other reasons for adopting this design driven from its many advantages. For example, Dörnyei (2007) contended that the inquiry of quantitative method is focused, compactly controlled, systematic, and rigorous, "Involving precise measurement and producing reliable and replicable data that is generalizable to other contexts" (p. 34). He also stated that some in-built indices and checks, which enable readers to vouch for the validity of quantitative results, are provided by the approach, as well as the fact that it has a refined and a far-reaching statistical analytical apparatus. McGivern (2009) maintained that the method is useful for testing, validating, measuring, and quantifying theories and hypothesis, and for providing a sparse and a nomothetic description of a large population. Sukamolson (2012) highlighted other strengths of the quantitative design, some of which are its flexibility in condensing the findings to statistics, in drawing a statistical comparison between different groups, and lastly, its definiteness, precision, and standardization. With all of these positive sides of the approach, this study can generalize its findings, provide insightful remarks and comments, as well as broaden the understanding—by allowing different detailed interpretations/descriptions—, and the scope of the topic in question since a Saudi context has rarely been investigated for such matter. The method will also help build or at least examine the veracity of the already built sophisticated/narrow theories in the field of writing skills (i.e., task complexity), and specify which of the CAF indices can tap into the development of Saudi EFL learners longitudinally.

Finally, the standardized and structured data collection, despite giving reliability in return, may lead to the production of superficial understanding and description of the topic

investigated (McGivern, 2009). This is not to say that the latter is the only weakness of the quantitative approach as there are others. For instance, Dörnyei's (2007) asserted that the quantitative approach's general exploratory capacity is narrow because participants' responses are averaged out. As a result, the subjective variety of an individual life would impossibly be captured. There is another issue which is the fact that "similar scores can result from quite different underlying processes, and [such approaches] are generally not very sensitive in uncovering the reasons for particular observations or the dynamics underlying the examined situation or phenomenon" (p. 35). Nevertheless, there are many reasons why these few articulated downsides of the quantitative method are not applicable in the situation of this research. First, it goes without saying that these disadvantages were observed in studies that collected their data by means of a questionnaire rather than the actual production of learners/respondents. Second is that the major aim of this research is to investigate the impact of using different teaching and learning environments on low-and-high proficient Saudi EFL learners' writing development in various text types. Thus, by examining such, together with controlling other variables such as applying the same teaching approach (i.e., process genre), supplementary materials, textbooks, and assignments on all subjects, and allowing students to write with no restrictions on the time of completion or the length of texts, there would a space for individual differences to surface. The standardized metrics used, which were developed by well-known writing skill researchers, in analyzing these quantitative written data will hopefully be efficiently sensitive to uncover in detail these differences, as well as determining to what extent the above-mentioned factors influence the writing development of participants; even if they obtain similar results in CAF constructs.

3.3 Setting

This research was carried out at one university in Saudi Arabia in 2013. The duration of the application of the research lasted for 14 weeks (= 3 months and 2 weeks). The research participants were recruited, taught, examined, and their written productions were analyzed at the same location.

3.4 Participants

At a college of English language, six groups were generated from level one and level six EFL undergraduate male students. Demographically, in both levels, the subjects' ages varied, from 18 to 22 years, and their length of exposure to English ranged between 10 to 14 years. The English proficiency levels of these participants were determined by their levels at the college rather than their results in any English language proficiency tests. For example, level

one students served as low proficient participants, while level six learners served as high proficient subjects. Wolfe-Quintero et al. (1998) emphasized that "of all the types of proficiency measures, program level may be the most valid developmentally" (p. 9). The abandonment of all English language proficiency tests were for many reasons. Davies, Hamp-Lyons, and Kemp (2003), for instance, maintained that not only are most of these examinations designed based on native speakers instead of nonnative speakers, but it is also most likely that the criteria of corrections and giving scores are dependent on the former as well. Kong, Powers, Starr, and Williams (2012, p. 2) contended that the scores assigned to learners in the proficiency test(s) "may not reflect their true knowledge and skills if language barriers prevent them from correctly answering the questions." Other researchers emphasized other factors that could influence the obtained marks such as students' anxiety of and/or their beforehand preparation for these tests (Mirici, 2003), and the fact that there are various types of examinations, each of which would adopt different questions to capture the proficiency of the basic language skills: reading, writing, listening, and speaking (e.g., Jones & Spolsky, 1975). Even if any of the proficiency examinations had been used, reassigning participants to groups of proficiency levels based on their test scores would have been impossible as the levels of the writing courses these subjects were taking had been already assigned and could not be altered.

In addition, these research groups fell into three categories based on the imposed teaching/learning environments used in the study: traditional learning context, blended learning context, and online learning context. To each of these environments, an equal number of random participants was assigned. For example, in level one, forty-five students were distributed to the three groups, each of which had 15 students, whereas in level six, 30 students were distributed to the three groups with 10 students in each.

3.5 EFL Courses

This study was conducted during the second semester of the academic year 2013. During this period, level one students were taking a writing course entitled (Eng.111) for three hours a week. The designated course book for that course was *'Great Writing 2: Great Paragraphs'* by Folse, Muchmore-Vokoun, and Solomon (2010). Students were required to study only the first six chapters of the book. These chapters were about explaining the construction of a paragraph (i.e., how to write a topic, supporting, concluding sentences, and definition paragraphs). Nonetheless, some supplementary materials on the basic and fundamental characteristics of two text types: argumentative and classification, were assigned

and explained to such students. Level six students, on the other hand, were studying their writing course (Eng.316) for two hours per week. The textbook assigned was *'The Researched Essay: Effective Academic Writing'* by Liss and Davis (2012), and students were taught the last three chapters of the textbook on argumentative, classification, and reaction essays. Extra lessons on writing were also introduced and taught to the students. Despite the difference in the total amount of taught hours between level one students (42 hours), and level six students (28 hours), the comparison allows insight into the learners' CAF development at different proficiency levels.

Also, these courses are mandatory by the department for all of its students and are the prerequisites for other writing courses and other courses. For example, level one students cannot register for level 2 writing course without first finishing the level one writing course. Level six students, on the other side, cannot register for research methodology without finishing all writing courses from level one to level six.

3.6 Teaching Methodology, Teaching/Learning Contexts, and Participants

Teaching and learning writing skills in general are considered to be a very time-consuming experience for both teachers and students (Spelkova & Hurst, 2008). The difficulty of teaching and even learning these skills escalates when applying a specific teaching method such as the process genre approach in different teaching contexts/environments, and on a large number of participants who are divided into small groups. In other words, because the researcher divided the research participants randomly into six groups, with three pairs of groups for the purpose of comparison, there was a need to recruit a fellow colleague to teach two of these groups. With such help, the research objectives would hopefully be met, and its questions would be undertaken and answered accurately by being capable of assessing and measuring the outcomes of learners in the three learning contexts; a point emphasized by Thompson and Wrigglesworth (2013). Since one of any research aims is to eliminate extra variables which would jeopardize its findings, the researcher worked as a teacher for four of the groups and as a coordinator with the teacher of the other two groups. This was very important to make sure that the other teacher and the researcher covered the same textbook chapters, used the same teaching method—process genre approach—, the same supplementary materials, and required students to undertake the same writing tests, weekly assignments, and activities. The teaching and learning contexts and participants are outlined in the following:

3.6.1 Traditional learning context groups for level one and six students (TLC1/TLC6).

The writing skills traditional learning context in this study means a form of education which was based on face-to-face classroom methods. In other words, all students were required to attend their writing skills class on-campus. The teacher of these groups was a professor of English writing skills in the college. The class was equipped with white and smart boards, a projector, and a PC. The teacher was asked to use the equipment when teaching (i.e., TLC1 and TLC6 groups) the writing skills. The number of participants was 15 students at level one, and 10 students at level six. Student assignments such as writing paragraphs or essays, revising their writings, correcting their fellow students' writings, answering some of their writing skills textbooks' exercises, and receiving feedback from the teacher were undertaken in class and at home. Based on the teacher's and fellow students' corrections and comments, students were asked to do multiple drafts for every single topic and submit them to the instructor by hand during their following classes.

In addition, in this type of class, students had supplementary materials and exercises (e.g., lessons and sample texts found on professional writing skills websites/books), to help them master the composition skills of argumentative and classification paragraphs/essays. Participants were told to analyze these sample texts. The analysis was to underline the words, phrases, expressions, or even sentences that students thought were worth incorporating in their forthcoming paragraphs/essays. The explanations of these words, expressions, and sentences, how to use them correctly and accurately, and other lessons were taught by the instructor in the classes.

3.6.2 Blended learning context group for level one and six students (BLC1/BLC6).

The writing skills blended class in this study means a form of education which combined face-to-face classroom methods with computer-mediated activities. In other words, the participants in this category were exposed to two different learning environments (i.e., traditional and online learning contexts). Every other week and until the end of the semester, participants (i.e., BLC1 and BLC6 groups) were taught by the researcher in either their traditional class, and all of their assignments, and activities were undertaken in class or at home, or in a cyber-environment in which their classes were held online, and all of their assignments with their requirements, and activities were undertaken online on designated blogs. A detailed description of how the online learning context works will be described

below. The total number of students in this category was (level one: 15 students, and level six: 10 students).

3.6.3 Online learning context groups for level one and six students (OLC1/OLC6).

The writing skills online learning context in this study means a form of education which took place with computer-mediated technologies and on a website on the Internet. In other words, all students were required to follow their writing skills class online on a specialized website, which the researcher had designed with the help of a professional website-designing company. The website is entitled 'English Online Collaborative Learning:' (Eocl.net/#). It consists of three forums. One is *'Forums,'* which consists of many sub-forums/blogs. For example, for both OLC1 and OLC6 groups, they have separate sub-forums i.e., *'English Writing Skills Level One (OLC1),'* and *'English Writing Skills Level Six (OLC6),'* each of which contains five child forums: course syllabus, general discussions, general questions, online classes discussions/questions, and assignments. All participants in the above groups were required to use these child forums accurately. For example, if they wanted to know the syllabi of their courses, they could browse the threads specified for them where the researcher had put copies of them in .pdf files. Furthermore, if participants wanted to discuss general issues regarding improving English language skills –especially writing-, raise some questions about them, or watch and listen again to their recorded online classes, they can do so in the designated forums.

Moreover, students were asked to do their assignments in composing paragraphs or essays, revising their writings, correcting their fellow students' writings, and answering some of their writing skills textbook exercises on the *assignment* forum. They were also required to follow and take into consideration the teacher's and fellow students' corrections and comments before posting their new drafts on that specific blog. Finally, in that blog students had similar extra supplementary materials and exercises to those of the traditional group students. However, the only two differences between the four groups (TLC1, and TLC6, and OLC1, and OLC6) were that the analyses undertaken by participants in the latter two groups were posted on the designated blogs, and their explanation together with the other incorporated lessons by the instructor took place in online lectures. These lectures had been recorded, and the explanations of the analyses (i.e., on how to use the underlined words, phrases, expressions, and sentences accurately) had been put verbatim in the blog.

The second forum is *'Chatting.'* In this section, participants have two types of synchronous computer conferencing, all of which "allow…students and instructor…to

interact in real time using personal computers [laptops, or smartphones] to deliver a variety of text, voice, visuals, shared applications, and videos" (Moore and Kearsley, 2012, p. 80). The first type—known to be the oldest and the simplest form—is *'Website Chatting,'* which permits participants to interact with each other by typing text messages, and recently, posting icons of self-expressions and emotions (i.e., emoji; a Japanese word for a picture letter), recording and sending voice(s), and sending files. Yet, the highly text-form conversation within this system is "multiperson[s]," because everyone in the class can see it. The strength of this kind of chatting is that it "allow[s] question-and-answer sessions and a chance for participants to complement the more useful asynchronous communications on forums and bulletin boards with the experience of exchanging ideas spontaneously." Nevertheless, it is not considered to be a "powerful tool for online classes" (Moore and Kearsley, 2012, p. 80). The reasons are that not only does it lack the basic elements of an effective class such as a whiteboard, but there is also an issue of learners being prevented from engaging in the negotiation of meaning as a consequence of not being able to see each other or their instructor (Schenker, 2015). Therefore, it was obligated to use other types of synchronous computer conferencing.

The second type of computer conferencing is *'Online sessions'* in which participants can attend their online/virtual classes. In other words, because participants were required to attend their writing skills classes online, the researcher had integrated a program called Adobe Connect within the website. Adobe Connect is simply a software designed and developed by Adobe Systems Incorporated which works on the Flash Player base, and which is used for online training/teaching materials, learning modules, and web conferencing. The software has many capabilities such as multiple meeting rooms per user, polling, file sharing, audio and video conferencing, meeting recording, screen sharing, chat, whiteboards, and notes (Adobe Systems Incorporated, 2016). Inside the online sessions' section, participants have links with their groups' names for their online class. The strength of this type of computer conferencing is that it will ignite "collaborative learning [between participants] that goes beyond information exchange and necessitate moderated critical discourse to realize new and worthwhile learning" (Kaye, 1992, as cited in Garrison & Akyol, 2013, p. 104).

The third forum is *'Videos and Audios'* in which many free videos and audios made by professional English native writing skills teachers had been uploaded, and which explain the types of lessons students were required to study. These lessons were about developing topics, supporting, and concluding sentences, and how to construct argumentative and classification paragraphs/essays. Following some of the suggestions of Naidu (2013), the researcher showed

"a great deal of care and consideration...[when]…selecting [only] the [appropriate] media that [would] enable the achievement of the intended learning outcomes" (p. 270). OLC1 and OLC6 group students were asked by the instructor to view these videos and audios, and to post their comments, suggestions, and questions of these clips on their designated forum(s). These students were also given permission to share only free writing-skills-related videos/audios by uploading them on the website.

3.7 Materials and Procedures for Data Collection

3.7.1 Materials: writing paragraphs/essays.

In this research, participants from all six groups undertook three different writing skills tests (i.e., pre-test, mid-term, and post-test). Each of these exams consisted of two questions on composing essay/paragraphs. These questions were about writing argumentative and classification text types which would encourage subjects to use the present tense. The topics of these prompts were taken and modified from English proficiency exams such as IELTS and TOEFL, and had to be of a transparent nature to allow students write as much as possible. The justifications of having such writing tasks were: 1. to determine how the CAF constructs and sub-constructs would be influenced in such genres that differ in their level of complexity; 2. both of which are among the basic and most frequently taught writing genres to Saudi EFL undergraduates who would most likely do them in other courses such as literature, translation, and linguistics; and 3. to be able to compare the yielded results to those of other studies which incorporated other text types such as narrative, opinion-expression, topic-writing, picture-description, information-gap, decision-making, instruction, text-reconstruction, and story-creation. In addition, in the argumentative task, students are required to articulate two/more opposing points of view about a specific controversial topic with examples and evidence; vouching for the veracity of one of these with reasonable justifications, while in the classification task, participants are asked to categorize people or items into different classifications with explanations, examples, and justifications. Although both of these composition tasks require reasoning demands and the use of the same temporal reference (i.e., present tense), and according to Robinson's Triadic Componential Framework, they would be labelled as difficult rather than simple tasks, he explained that the plus and minus of the tasks' features could be regarded as continuum. There is also the fact that manipulating these tasks in terms of the reasoning demands and the tense is difficult. Nonetheless, the argumentative task is considered to be more taxing for the Saudi EFL participants than the classification task. The reasons are that the characteristics of the argumentation in English is different from

Arabic and the fact that Saudi EFL learners find them difficult. Shukri (2014) emphasized that "the Western mode of argumentation is based on syllogistic model of proof, while Arabic argumentation is characterized by repetition, which is deeply rooted in the language" (p. 196). Al-Abed Al-Haq and Ahmed (1994) found that the Saudi EFL learners, even with a different range of proficiency levels, had major problems in organization, supporting, developing and increasing the persuasiveness of the argument, completing and balancing it and its counter-argument, and clarifying, qualifying, and directing the thesis to the readers. Finally, for a copy of the pre-test, mid-term test, and post-test, see appendices A, B, and C respectively. The prompts used in the three tests are presented in Table 5.

Table 5

Prompts of Pre-test, Mid-Term Test, and Post-test

Writing Tasks	Tests	Questions
Argumentative Task	**Pre-Test**	1. Are you for or against smoking in general and smoking in public places in particular? Discuss people's different opinions about this.
	Mid-Term Test	2. Are you for or against bodily punishment at schools? Discuss people's different opinions about this issue.
	Post-test	3. Mobile phones have brought more good than harm to our life. Do you agree or disagree?
Classification Task	**Pre-Test**	1. The different brands of cars in Saudi Arabia. Discuss the advantages and disadvantages of each brand.
	Mid-Term Test	2. Nowadays, people in Saudi Arabia and the Arab world can watch a variety of TV channels. Discuss the advantages and disadvantages of each type of these channels.
	Post-test	3. Saudi university students' different hobbies. Discuss the advantages and disadvantages of each hobby type.

3.7.2 Procedures.

One month prior to the application of this study, the researcher liaised with the professor of writing skills in the department on a regular basis, to discuss the dates for applying the pre-test, mid-term test, post-test, the course syllabi, the textbook contents, the supplementary materials, the teaching approach, and the learning contexts. This was significant to control any other external variables that might affect the results of the study. In the first week of the semester, all participants from the six groups met with their writing teachers, were introduced to the course textbooks, materials and requirements, and learning contexts. Participants were told that regardless of whether their learning contexts were traditional, blended, or online, they had to undertake their three tests on campus and on specific dates provided to them in the course syllabi.

In the second week and under the supervision of the composition skills teachers, all subjects from level one and level six were put into one large class to undertake the pre-test. The pre-test (see appendix A) was distributed, and students were required to write about two different topics with no constraints put on the time of submission, or on the length of their paragraphs/essays. Participants were given sufficient time to write their outlines about these topics. Furthermore, the two teachers of the composition skills photocopied each of their group participants' pre-test twice. Using standard correction criteria, one photocopy of each student's work was corrected and commented on thoroughly by the instructors, while the other photocopies were left for the participants' fellow students to correct and comment on. In the following classes, the instructors distributed the uncorrected photocopies between students, and they were asked to give them to their fellow students to correct them. At the end of the class, the teachers' corrected photocopies of the pre-test were distributed among students and they were requested to take the suggestions made by teachers and colleagues on board when undertaking second drafts. For the traditional and blended environment groups, they had to submit their second drafts by hand to the teachers, whereas the online context groups had to post the edited versions of the pre-test online, on their designated forums, that is, *assignments*.

Likewise, in the middle of the semester (weak 7), all group participants completed the second test, following the same procedures as the pre-test. Participants were asked to write about two topics for this mid-term test (see appendix B). After receiving all of the mid-term tests, the instructor of the traditional learning context did the same procedure of photocopying, correcting, and commenting as the aforementioned, and his participants were required to complete second drafts based on his comments and corrections as well as their fellow students'. The submission of the second drafts took place in class. However, the teacher of the other groups: online learning context groups (OLC1, and OLC6), and blended learning context groups (BLC1, and BLC6), did something different. He made individual .pdf files of each of the participants' mid-term tests, uploaded them on the designated forums, and asked the students to log in and try to correct their friends' writings. Following the students' corrections, the teacher posted the corrected copies online. All participants were asked to take all of the comments and corrections made by their teacher and colleagues into consideration when doing the second drafts. Again, all of these had been posted online on the specified forums.

Finally, at the end of the semester (week 14), all participants were required to do their post-test (see appendix C) by writing about two topics.

3.8 Ethical Considerations

Many authors are concerned with highlighting various types of ethical considerations which researchers must take into account when conducting their studies. However, the ethical issues that were considered to be relevant to the nature and application of the current research were emphasized and addressed. These issues were about consent of application, orientation to use the website (Howell, Saba, Lindsay, & Williams, 2004), attendance at classes, appropriateness of the materials (Thompson & Wrigglesworth, 2013), technical support (Thompson & Hills, 2005), and authorships/copyrights (Lipinski, 2013). As regards the consent, almost three months prior to conducting the study, an official letter was sent to the College of English, requesting permission to collect the data required. After that, while the research was conducted, at the beginning of the semester the importance of the research in the field of EFL writing skills, and the procedures of collecting the data were explained and clarified to all participants in (TLC1/6, BLC1/6, and OLC1/6) groups. Subjects were requested to write their own names when undertaking the different tests (pre-test, mid-term test, and post-test), and confidentiality was assured.

After assigning participants to their specified learning contexts, participants in OLC1, OLC6, BLC1, and BLC6 groups were invited to attend two tutorial classes with the researcher in the computer lab located in the college. In the tutorial lectures, the instructor gave an introduction to the website i.e., the English Online Collaborative Learning, explained how each group should use it accurately, and how to register and participate in it. At the end of these lectures, all students were successful in accessing all of the websites' different forums, either for online lectures, website chat, or even their own forums. All of these participants were provided with technical support by the website design company and the researcher. Moreover, in order not to have any schedule difficulties (Steele & Thurmond, 2009) between participants' writing online lectures and their other classes, they were asked to agree on a specific time in which they would have their lectures. Based on the attendance of subjects, the specified time for their online classes had been confirmed to be suitable for everyone.

Finally, the authorship of both of the supplementary materials used in this research and the videos uploaded in the researcher's website, was not an issue as they were provided by free websites and/or uploaded by their owners in public websites such as YouTube.

3.9 Data Analysis

After collecting all the data (i.e., 450 essays/paragraphs) in hardcopy forms from the six groups of participants, the researcher decided first to transfer them to electronic versions, that

is by using Office WORD document and later to transfer them to plain texts. This was deemed essential and necessary as part of the written texts analyses would take place using some analyzing software programs. Then, all of the data were analyzed manually and/or electronically, based on different types of measures designated for each of CAF constructs. For example, there were 45 measures used in the research divided into: syntactic complexity (12 measures), lexical complexity: lexical density (one measure), lexical sophistication (five measures), and lexical variation (19 measures), accuracy (four measures), and fluency (four measures). All of them were incorporated in the analysis since previous research found them to be the best measures of writing development and the fact that Kuiken and Vedder (2007b) concluded that "the use of more global and more specific measures may complement each other" (p. 276). Also, because of the longitudinal nature of this research and using a number of participants, proficiency levels, text types, and measures, there was a need to use a method to present the results yielded in a readable and consistent matter. The method was to only cite the significant results and disregard the non-significant ones. However, before discussing some important issues regarding these measures and the software programs used to analyze them, and presenting the significant findings, it is worth mentioning that a professional statistician was recruited to help with the entry of the results and the decision on which statistical tests the researcher should use to find the answers to his research questions. The agreement was reached on two types of tests. First is a T-test, which would help determine whether each of the six groups' CAF measures remained unchanged, increased, and/or decreased significantly over time in the two text types. Second is an ANOVA test, which would aid in specifying which of these groups outperformed the others in terms of the CAF's indices in the two text types. In all of these tests, two kind of comparisons were undertaken to see if the findings of each measure were significant for each of the three CAF constructs in the short term and/or the long term.

In other words, in order to determine adequately whether there is a significant result for each of the CAF constructs in the two writing tasks, and the time at which it occurs, the comparisons were first drawn between the results of all measures in all of the participants' written texts (i.e., classification and argumentative) in the pre-test with the results of the same types of texts in the mid-term test, and then between the pre-test and the post-test. Since these comparisons in the t-test took place within each learning context group—whereas in the ANOVA test, they were across each three learning context groups—the yielded CAF findings were classified dependent on their kinds of effect (i.e., uninfluenced, increasing, decreasing,

or both) and the number of measures indicating such effect, and independent of the types of these measures.

3.9.1 Complexity.

3.9.1.1 Syntactic complexity analysis.

For syntactic complexity, there is a software program called *L2 Syntactic Complexity Analyzer* (L2SCA), designed by Xiaofei Lu. The L2SCA is used to analyze written texts produced by L2 learners by identifying and counting first the occurrences of a number of syntactic structures and production units, and then producing the results of fourteen metrics of syntactic complexity proposed in the literature on L2 development (Lu, 2010). Because of the program's analytical effectiveness, it has been incorporated in many studies (e.g., Kim, 2014; Long & Tabuki, 2014; QI, 2014; Wind, 2013). The reliability of it has also been substantiated. For example, one of Yoon and Polio's (2014) study's aims was to compare the results of a large number of L2 written texts obtained through hand coding and through the L2SCA. These researchers concluded that the software produced reliable findings. Crossley and McNamara (2014) emphasized that contrary to human raters whose analyses are subjective and require training, time to score, and monitoring, computational tools to assess syntactic complexity provide reliability, flexibility, and speed. Yet, there are two major issues that have to be mentioned with respect to the L2SCA. First, since the software requires an in depth knowledge of python, which is a high level of programming language, as well as in entering the commands manually, a batch mode of the software developed by Haiyang Ai, (available online at http://aihaiyang.com/software/l2sca/batch/), was used instead. This batch mode allows analysis of compressed plain texts of thirty files each time it is used. There is also a single mode option of the program, which allows analysis of texts individually, and to compare two texts for selected metrics of syntactic complexity (Ai, 2016a).

In all of the above, nine units of production (i.e., words, sentences, verb phrases, clauses, T-units, dependent clauses, complex T-units, coordinate phrases, and complex nominal) have been used, each of which requires a definition by which the software can perform. Lu (2010) emphasized that his choice of the production unit interpretations was based on one of two factors: acceptability or operationalization. In other words, if there are competing definitions for one specific syntactic structure, only the more widely accepted interpretation would be used, while if these definitions were used in the literature equivalently; with no preference of one over the other(s), the choice would be the one that

could be operationalized and implemented in the program appropriately and accurately.

Finally, some of the definitions used in the software are presented in Table 6.

Table 6

Definitions of Terms

Sentences	"A sentence is a group of words delimited with one of the following punctuation marks that signal the end of a sentence: period, question mark, exclamation mark, quotation mark, or ellipsis (Hunt, 1965; Tapia, 1993)."
Clauses	"A clause is defined as a structure with a subject and a finite verb (Hunt, 1965; Polio, 1997), and includes independent clauses, adjective clauses, adverbial clauses, and nominal clauses."
Dependent clauses	"In line with the definition of clause, a dependent clause is defined as a finite adjective, adverbial, or nominal clause (Cooper, 1976; Hunt, 1965; Kameen, 1979)."
T-units	"A T-unit is 'one main clause plus any subordinate clause or nonclausal structure that is attached to or embedded in it' (Hunt 1970b, p. 4)[5]."
Complex T-units	"A complex T-unit is one that contains a dependent clause (Casanave, 1994)."
Coordinate phrases	"Only adjective, adverb, noun, and verb phrases are counted in coordinate phrases (Cooper, 1976)."
Complex nominals	"Complex nominals comprise (i) nouns plus adjective, possessive, prepositional phrase, relative clause, participle, or appositive, (ii) nominal clauses, and (iii) gerunds and infinitives in subject position (Cooper, 1976)."
Verb phrases	"Verb phrases comprise both finite and non-finite verb phrases."

(Adapted from Lu, 2010, pp. 9-13)

The After calculating all of the above syntactic structures as well as words, the results of fourteen measures of grammatical complexity are produced. This bring us to the second issue which is the fact that two of these measures (i.e., T-unit length [W/T, or MLT], and clause length [W/C, or MLC]) had been transferred to fluency. The reason for such decision will be explained when discussing fluency measures. Moreover, in the literature on syntactic complexity, there have been some disputes on the best indices that can tap into such construct (see section 2.7.1.1). However, the decision to use the software which adopted such a wide range of measures was that Lu (2010) contended that eventually it would not only raise the reliability and validity of syntactic complexity findings, but it would also help L2 writing instructors and researchers to easily assess, track, understand, and see the development of their students' writing abilities from different angles. He maintained that it would also provide an aid to, almost effortlessly, measure the effectiveness of pedagogical

---

[5] The reference provided by Lu (2010) for T-unit definition was
Hunt, K. W. (1970b). Do sentences in the second language grow like those in the first?
TESOL Quarterly, *4*, 195-202. doi:10.2307/3585720
The correct reference is:
Hunt, K. W. (1970a). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development (Serial No. 135)*, 35(1), iii-vi+1-67.

interventions/teaching approaches/contexts applied on learners. There is also the possibility to capture how such measures would react/evolve or not to different text types produced by L2 learners. Finally, since the production unit findings were accessible in the program, it was easy for the researcher to revise and recheck the application of adopted definitions in the actual writings. The software indicated that findings for both syntactic structures and measures were valid and reliable as the manual analysis and calculation yielded similar results. The syntactic complexity measures, formula and analysis and coding are presented in Table 7.

Table 7

Syntactic Complexity Measures, Formula and Analysis and Coding

Construct	Measures	Basic Formula	Analyzed and Coded by
Syntactic Complexity	The mean Length of sentence (MLS)	# of words / # of Sentences	Web-based L2 Syntactic Complexity Analyzer (Batch Mode [Ai, 2016a]), and the researcher)
	Sentence complexity ratio (C/S)	# of clauses / # of sentences	
	T-unit complexity ratio (C/T)	# of clauses / # of T-units	
	Complex T-unit ratio (CT/T)	# of complex T-units / # of T-units	
	Dependent clause ratio (DC/C)	# of dependent clauses / # of clauses	
	Dependent clauses per T-unit (DC/T)	# of dependent clauses / # of T-units	
	Coordinate phrases per clause (CP/C)	# of coordinate phrases / # of clauses	
	Coordinate phrases per T-unit (CP/T)	# of coordinate phrases / # of T-units	
	Sentence coordination ratio (T/S)	# of T-units / # of sentences	
	Complex nominals per clause (CN/C)	# of complex nominals / # of clauses	
	Complex nominals per T-unit (CN/T)	# of complex nominals / # of T-units	
	Verb phrases per T-unit (VP/T)	# of verb phrases / # of T-units	

(Adapted from Lu, 2010, p. 6)

3.9.1.2 Lexical complexity analysis.

For lexical complexity, Lu has also designed another computational software called *Lexical Complexity Analyzer* (LCA), which has been incorporated in many studies (e.g., Fazilatfar, Fallah, Hamavandi, & Rostamian, 2014; Lorenzo & Rodríguez 2014; Mazgutova & Kormos, 2015; Tsai, 2013; Yang, 2014) for its analytical efficiency. According to Lu (2012) the LCA provides an automated analysis of English texts in terms of 25 metrics of lexical complexity across three dimensions: lexical sophistication (five metrics), variation (19

metrics), and density (one metric). The LCA is designed to treat, and therefore, analyze texts according to two options of preference in spelling; either American English or British English. When choosing one of them, the program will automatically bring the built in wordlist designated for the choice either American National Corpus, or British National Corpus (Ai, 2016b). In the case of this research, American English was chosen. Then, the results of the metrics are undertaken after calculating the number of types of: words, sophisticated words, lexical words, sophisticated lexical words, adjective, nouns, adverbs, verbs, and sophisticated verbs, and the tokens of words, lexical words, sophisticated lexical words, and verbs (Lu, 2012). The software has a batch mode developed by Ai (2016b, [available online at http://aihaiyang.com/software/lca/batch/]). This batch mode enables the analysis of up to one hundred compressed plain texts each time it is used, with the possibility to transfer the yielded results from EXCEL to any other statistical programs. Crossley and McNamara (2009) maintained that the automatic analysis software, as the LCA, affords "a more accessible and theoretically sound approach for the quantitative evaluation of texts," and reduces the subjective nature of intuitive judgements and fallibility of human raters (p. 121).

In terms of choosing lexical complexity measures, Wolfe-Quintero et al. (1998) believed that of the nine measures they examined, only one lexical variation measure (i.e., word type variation: WT/√2W, or CTTR), and one lexical sophistication measure (i.e., sophisticated word type ratio: SWT/WT, or LS2), seemed to be a promising indicator of L2 development. Yet, they acknowledged implicitly that their conclusion was somehow not conclusive as the studies addressing the construct were limited, and inclining to compare L2 learners with native speakers instead of using developmental and proficiency measures. There is also the fact that the significant findings for SWT/WT measure was based on the wordlist used (i.e., university wordlist, and 2000-word list). However, these issues and others have led to the design of the program. According to Lu (2012), the software can empirically eliminate many problems that L2 development and proficiency levels suffer from. That is, with the easy application of the program, not only will the researchers have the tool to expand the number of measures used, and the amount of texts (from different language tasks) analysed, but also the comparison between their studies will be readily accessible. By doing so, we can attain the most reliable and valid predictors of L2 proficiency levels and development, and understand how such indices interact and correlate with one another. Lu (2012) stated that the program would also provide a valid method to correctly assess, and therefore, link lexical complexity to the quality of productions instead of using the rating scales (i.e., excellent, good, poor) endorsed by the various language proficiency exams. Bulté and Housen's (2015) study results

made them "stressed the importance of calculating a sufficiently wide range of [syntactic and lexical] complexity measures in order to obtain a comprehensive picture of L2 development" (p. 42).

Moreover, after analysing the research data, the researcher revised and rechecked the accuracy of some of the yielded results. This was possible as the software provided some of the findings of the word types and tokens along with the findings of the measures. For example, the software revealed the results of word types, sophisticated word types, lexical types, sophisticated lexical types, word tokens, sophisticated word tokens, lexical tokens, and sophisticated lexical tokens. By calculating these, the results of measures such as LD, LS1, LS2, NDW, TTR, CTTR, RTTR, LogTTR, Uber, and LV, are produced. The manual analysis showed reliability of all outcomes. Finally, all of lexical complexity measures, formula and analysis and coding are shown in Table 8.

Table 8

Lexical Complexity Three Dimensions Measures, Formula and Analysis and Coding

Construct	Measures	Basic Formula	Analyzed and Coded by
Lexical Density	Lexical density (LD)	# of tokens of lexical words / # of tokens of words	Web-based Lexical Complexity Analyzer (Batch Mode [Ai, 2016b]), and the researcher
Lexical Sophistication	Lexical sophistication-I (LS1)	# of tokens of sophisticated lexical words / # of tokens of lexical words	
	Lexical sophistication-II (LS2)	# of types of sophisticated words / # of types of words	
	Verb sophistication-I (VS1)	# of types of sophisticated verbs / # of tokens of verbs	
	Corrected VS1 (CVS1)	# of types of sophisticated verbs / square root of two times the # of tokens of verbs	
	Verb sophistication-II (VS2)	# of two times the types of sophisticated verbs / # of tokens of verbs	
Lexical Variation	Number of different words (NDW)	# of types of words	
	NDW (first 50 words) (NDWZ)	# of types of words T in the first 50 words of sample	
	NDW (expected random 50) (NDWERZ)	Mean *T* of 10 random 50-word samples	
	NDW (expected sequence 50) (NDWESZ)	Mean *T* of 10 random 50-word sequences	
	Type/Token ratio (TTR)	# of types of words / # of tokens of words	

(continued)

Construct	Measures	Basic Formula	Analyzed and Coded by
Lexical Variation	Mean Segmental TTR (50) (MSTTR)	Mean TTR of all 50-word segments	Web-based Lexical Complexity Analyzer (Batch Mode [Ai, 2016b]), and the researcher
	Corrected TTR (CTTR)	# of types of words / square root of two times the # of tokens of words	
	Root TTR (RTTR)	# of types of words / square root of the # of tokens of words	
	Bilogarithmic TTR (LogTTR)	log # of types of words / log # of word tokens	
	Uber Index (Uber)	log of two times # of word types / log (# of word tokens / # of word types)	
	Lexical word variation (LV)	# of types of lexical words / # of tokens of lexical words	
	Verb variation-I (VV1)	# of types of verbs / # of tokens of verbs	
	Squared VV1 (SVV1)	# of two times the types of verbs / # of tokens of verbs	
	Corrected VV1 (CVV1)	# of types of verbs / square root of two times the # of tokens of verbs	
	Verb variation-II (VV2)	# of types of verbs / # of tokens of lexical words	
	Noun variation (NV)	# of types of nouns / # of tokens of lexical words	
	Adjective variation (AdjV)	# of types of adjectives / # of tokens of lexical words	
	Adverb variation (AdvV)	# of types of adverbs / # of tokens of lexical words	
	Modifier variation (ModV)	# of types of adjectives plus # of types of adverbs / # of tokens of lexical words	

(Adapted from Lu, 2012, pp. 3-7)

3.9.2 Accuracy analysis.

In the case of accuracy, there were four measures used. The measures were one frequency measure (i.e., error-free T-unit [EFT]), and three ratio measures (i.e., errors per T-unit [E/T], error-free T-unit to T-unit [EFT/T], and errors per word [E/W]), all of which were analyzed by the researcher manually. The types of errors/mistakes counted in this study were grammatical, morphological, and spelling, and any other types such as capitalization, punctuation, or lexical choice were neglected. The reason for choosing these measures was based on Wolfe-Quintero et al.'s (1998) suggestions and findings. For example, they stated that the EFT, despite being "far less useful for comparisons across studies" (p. 119), the E/T, and the EFT/T, all seemed to be correlated with holistic ratings and short-term change across different levels and within interact classes, but not across school levels or programs. This means that the rise of proficiency would result in a decrease of E/T, and a growth of both EFT and EFT/T. Also, these authors maintained that, based on the two studies that incorporated such a measure, E/W was a developmental measure that was either highly/moderately related to proficiency, or showed "an overall effect for proficiency…together with a significant

difference between three or more adjacent proficiency levels," or "for two or more proficiency levels" (p. 44). Lastly, all of accuracy measures, formula, and analysis and coding are presented in Table 9.

Table 9

Accuracy Measures, Formula and Analysis and Coding

Construct	Measures	Basic Formula	Analyzed and Coded by
Accuracy	Error-free T-units (EFT)	#of error-free T-units	Researcher
	Error-free T-unit ratio (EFT/T)	# of error-free T-units / # of T-units	
	Errors per T-unit (E/T)	# or errors / # of T-units	
	Errors in text length (E/W)	# of errors / # of words	

3.9.3 Fluency analysis.

In fluency, there were also four metrics used. They were three ratio metrics: T-unit length (W/T), clause length (W/C), and error-free T-unit length (W/EFT), and one frequency metric (i.e., text length [W]). As mentioned before, both of W/T, and W/C were analyzed automatically in syntactic complexity software, but also used in fluency. The transferring of these metrics was justified by Wolfe-Quintero et al (1998) for many reasons (discussed in detail along with their counter-argument in section 2.7.3), one of which is that they, and any other length metrics, lack the potential to discriminate between how different text lengths are achieved. Furthermore, the reason to use W/T, W/EFT, and W/C ratio metrics is that, as maintained by Wolfe-Quintero et al (1998), they "consistently increased in a linear relationship to proficiency level across studies [examined], regardless of task, target language, significance of the results, or how proficiency was defined" (p. 29). Length of text (W), on the other hand, had mixed results, and therefore, would not be considered a reliable metric for development (Wolfe-Quintero et al., 1998). But, the inclusion of it here was due to using different proficiency levels, and text types (Kim, 2014), and to see how participants' different writings would increase/decrease longitudinally. Moreover, all of the construct metrics were analyzed or reanalyzed manually by the researcher who for the purpose of maintaining the

consistency of all the results, adopted some of the definitions used in the L2SCA program such as those of the T-unit and the clause in the analysis, and counted the words in each text. The reason for undertaking the manual calculation of vocabulary was that in the most frequently asked questions in Lu's website, he explained that what might be a discrepancy between the word count of compositions undertaken by the program and by other writing tools, to name but a few, Office Word, is driven by the fact that the former regard contracted forms such as "wasn't, can't, I'm, and he's" as two tokens rather than one as in the case of the latter. The fluency metrics, formula, and analyses and coding are shown in Table 10.

Table 10

Fluency Measures, Formula and Analysis and Coding

Construct	Measures	Basic Formula	Analyzed and Coded by
Fluency	Mean length of T-unit (W/T)	# of words / # of T-units	Web-based L2 Syntactic Complexity Analyzer (Batch Mode [Ai, 2016a]), and the researcher
	Mean length of clause (W/C)	# of words / # of clauses	
	Mean length of error-free T-unit (W/EFT)	# of words / # of error-free T-units	Researcher
	Text length (W)	# of words	

Finally, the organization and the presentation of the significant findings for all CAF constructs are shown in Figure 3.

Figure 3. The Organization of Presentation of All Participants' Significant Results
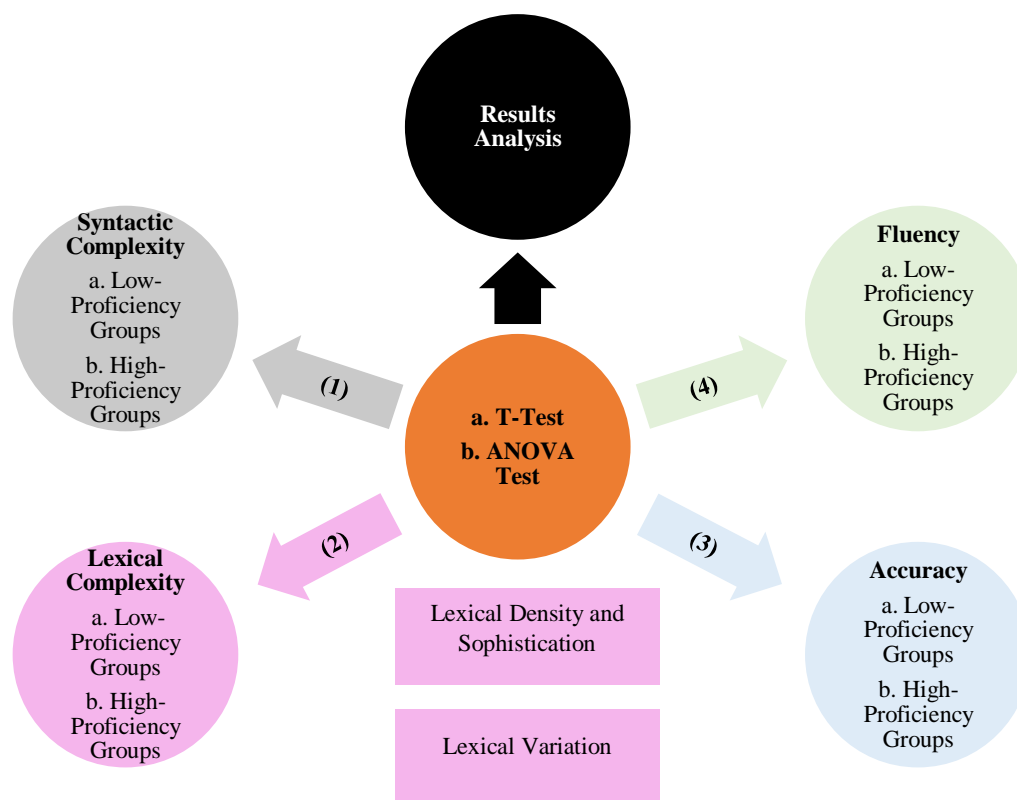
3.10 Conclusion

This chapter presented a detailed description of the adopted methodology. It started with a statement of the problem. Then, it was followed by an explanation of the research design, setting, participants, courses, teaching approach, and teaching and learning environments. It also provided a discussion on the material, instrument, procedures used for collecting the data and ethical considerations. Finally, the chapter ended with a thorough account of the data analyses, and how they would be presented in the following section of results.

Chapter Four
Results

4.0 Introduction

This chapter presents the significant results of the tests conducted on the CAF constructs. The test results are organized as per the structure outlined in the methodology section (see Figure 3), starting with those of syntactic complexity, then lexical density and sophistication, lexical variation, accuracy, and lastly, fluency.

4.1 The CAF Constructs

The findings reported in this section are drawn from analyzing the data related to four factors: the proficiency levels (low- and high-proficiency participants), the writing tasks (classification and argumentative), the timescales (short term and long term), and the learning environments (TLC, BLC, and OLC). To this end, two kinds of statistical tests were conducted. First, a t-test was used to identify the CAF findings of participants instructed in three different learning contexts at two proficiency levels. The t-test helps to determine if the CAF of participants in each of these six groups, representing the three learning environments, would be affected similarly or differently in the classification and argumentative tasks. Second, the ANOVA test was used to specify which of the three learning contexts, at the two proficiency levels, led to the most increase and decrease in CAF constructs among the participants. The CAF measures incorporated into this research include syntactic complexity (12 measures), lexical complexity (25 measures), accuracy (four measures), and fluency (four measures). The participants' answers data were analyzed using the above measures, in two timescales: the short term (from the pre-test to the mid-term test), and the long term (from the pre-test to the post-test). This means that all CAF measures were applied to each type of task (classification and argumentative) in three types of tests (pre-test, mid-term test, and post-test). In the short term, the analysis compared the data for each task in the pre-test with the same task type in the mid-term test. In the long term, the analysis compared the data for each task in the pre-test with the same task type in the post-test.

As mentioned earlier, in the t-test, the CAF results of these comparisons within each group were categorized with regard to their kinds of effect (i.e., unaffected, increasing, decreasing, or both) and the number of measures indicating such an effect, and without regard to the types of these measures. The number of measures—but not their types—was used to specify if there is equal increase and decrease (e.g., one measure of fluency indicates an

increase, while one other measure indicates a decrease), more increases than decreases (e.g., four measures of lexical variation indicate increases, whereas one measure indicates a decrease), or more decreases than increases (e.g., six measures of lexical variation indicate decreases, but only one measure indicates an increase). In the comparisons' results of the ANOVA test, on the other hand, the learning context (TLC, BLC, or OLC) whose group increased the most, increased in more measures than those of the other learning context groups, or decreased positively (i.e., in accuracy E/W and E/T measures) the most, would be regarded as the most effective. The same procedure had been applied toward designating the learning context that could be the least effective, that is, when its group increased the least or negatively (i.e., increased in accuracy E/W and E/T measures), or decreased the most or in more measures than those of the other groups.

4.1.1 Syntactic complexity.

4.1.1.1 Comparative changes in syntactic complexity within the low- and high-proficiency groups.

Table 11 displays the statistically significant t-test results relating to the syntactic complexity for all six groups (TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6) in the two text types.

4.1.1.1.1 Low-proficiency groups.

At the low-proficiency level, TLC1's findings showed that, in the short term, of the 12 metrics used, only two metrics (T/S and CN/C) revealed statistically significant scores (development or deterioration) for the classification task: T/C (M = -0.114, $p < .007$) and CN/C (M = 0.284, $p < .016$), while none of these metrics indicated any statistically significant scores for the argumentative task. All of these results mean that the syntactic complexity of TLC1 equally improved and declined in the simple task, whereas it remained unchanged in the difficult task. However, in the long term, there were no statistically significant scores for both composition tasks which show that whether the task was non-complex or complex, TLC1's syntactic complexity remained uninfluenced.

For BLC1, in the short term, only the C/T measure revealed a significant score in decreasing (M = -0.452, $p < .038$) in the classification task, but there were no statistically significant scores in the argumentative task. All of these findings indicate that the syntactic complexity of BLC1 declined in the simple task, while it remained unaffected in the difficult task. Nevertheless, in the long term, there were some statistically significant scores in both writing tasks, as shown by four measures. In the classification task, the significant

deteriorations were in: MLS (M = -0.5.492, $p < .044$), VP/T (M = -0.323, $p < .029$), C/T (M = -0.567, $p < .007$) , and CN/T (M = -1.012, $p < .009$) , whereas in the argumentative task, they were in: C/S (M = -0.609, $p < .018$), C/T (M = -0.468, $p < .015$), DC/T (M = -0.326, $p < .021$), and CT/T (M = -0.160, $p < .029$). All of these results reveal that regardless of the task being non-complex or complex, BLC1's syntactic complexity decreased equally in both task types, though across various types of measures.

In the case of OLC1, in the classification task, only three metrics (DC/C, CP/C, and CN/C) in the short term indicated statistically significant deterioration (i.e., DC/C: M = -0.148, $p < .047$) and increases (i.e., CP/C: M = 0.181, $p < .010$, and CN/C: M = 0.366, $p < .039$). Yet, there were no statistically significant scores in the argumentative task. All of these findings show that the syntactic complexity of OLC1 developed more than it declined in the simple task, but it remained unchanged in the difficult task. In the long term, there were no statistically significant scores for both text types, which implies that OLC1's syntactic complexity remained uninfluenced in the non-complex and complex tasks.

To summarize, the syntactic complexity of the low-proficiency Saudi EFL participants in the three learning environments was affected by the task complexity factor in one timescale. In only the short term, this construct, in the simple task, equally improved and decreased for TLC1, deteriorated for BLC1, and increased more than it declined for OLC1. In the difficult task, each group's syntactic complexity remained unaffected. Conversely, no impact of level of task complexity on syntactic complexity was found in the long term for all groups because such construct remained unchanged (e.g., TLC1 and OLC1) or equally decreased (e.g., BLC1) in both composition tasks.

4.1.1.1.2 High-proficiency groups.

At the high-proficiency level, TLC6's results revealed that in the short term, there were no statistically significant scores for both the classification and argumentative tasks. This means that whether the task was non-complex or complex, the syntactic complexity of TLC6 remained unchanged. Moreover, in the long term, there were no statistically significant scores in the classification task, while both C/S and C/T measures indicated significant developments (M = 0.849 and 0.535, respectively; and $p < .010$ and .017, respectively) in the argumentative task. All of these findings show that TLC6's syntactic complexity remained uninfluenced in the simple task, whereas it improved in the difficult task.

For BLC6, in the classification task, both CP/T and CP/C metrics deteriorated significantly (M = -0.194 and -0.145, respectively; and $p < .017$ and .019, respectively) in the

short term, but in the argumentative task, there were no statistically significant scores. This reveals that the syntactic complexity of BLC6 declined in the non-complex task, while it remained unaffected in the complex task. Furthermore, in the long term, there were no statistically significant scores in both writing tasks, which indicates that regardless of the task being simple or difficult, BLC6's syntactic complexity remained unchanged.

OLC6's results, on the other hand, showed that in the classification task both CP/C (M = -0.127, $p < .025$) and CN/C (M = -0.286, $p < .027$) decreased significantly in the short term, and in the argumentative task, the CN/C measure deteriorated significantly (M = -0.211, $p < .041$). All of these findings mean that the syntactic complexity of OLC6 declined in the non-complex task more than in the complex task. Nonetheless, in the long term, there were no statistically significant scores in the classification task, whereas the CN/T measure decreased significantly (M = -0.195, $p < .018$) in the argumentative task. These results reveal that OLC6's syntactic complexity remained unchanged in the simple task, but it deteriorated in the difficult task.

Briefly, the syntactic complexity among the high-proficiency Saudi EFL undergraduates in the three learning environments was influenced by the task complexity factor across different timescales. Specifically, in the long term, TLC6's syntactic complexity remained uninfluenced in the non-complex task, while it increased in the complex task. The syntactic complexity of BLC6 in the short term, declined in the simple task, whereas it remained unaffected in the difficult task. OLC6's syntactic complexity decreased in the non-complex task more than in the complex task, in the short term. In addition, in the long term, this group's construct remained unchanged in the simple task, but it deteriorated in the difficult task. Finally, in the short term for TLC6 and the long term for BLC6, there was no effect of level of task complexity on their syntactic complexity since it remained uninfluenced in both text types.

Table 11

TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6 Syntactic Complexity Metrics: Significant Results of the T-Test

Learning Contexts			Pre-Test to Mid-Term Test						Pre-Test to Post-Test					
			Classification Task			Argumentative Task			Classification Task			Argumentative Task		
			M	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*
Proficiency Levels — **Low-Proficiency Groups**	**Traditional Learning Context (TLC1)**	**(T/S)**												
			-0.114	0.192	.007									
		(CN/C)												
			0.284	0.550	.016									
	Blended Learning Context (BLC1)	**(C/T)**							**(MLS)**			**(C/S)**		
			-0.452	0.588	.038				-5.492	7.400	.044	-0.609	0.670	.018
									(VP/T)			**(C/T)**		
									-0.323	0.396	.029	-0.468	0.496	.015
									(C/T)			**(DC/T)**		
									-0.567	0.516	.007	-0.326	0.367	.021
									(CN/T)			**(CT/T)**		
									-1.012	0.974	.009	-0.160	0.195	.029
	Online Learning Context (OLC1)	**(DC/C)**												
			-0.148	0.203	.047									
		(CP/C)												
			0.181	0.174	.010									
		(CN/C)												
			0.366	0.478	.039									

	Learning Contexts	Pre-Test to Mid-Term Test						Pre-Test to Post-Test					
		Classification Task			Argumentative Task			Classification Task			Argumentative Task		
		M	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*
High-Proficiency Groups	**Traditional Learning Context (TLC6)**										**(C/S)**		
											0.849	0.826	.010
											(C/T)		
											0.535	0.576	.017
	Blended Learning Context (BLC6)	**(CP/T)**											
		-0.194	0.210	.017									
		(CP/C)											
		-0.145	0.161	.019									
	Online Learning Context (OLC6)	**(CP/C)**			**(CN/C)**						**(CN/T)**		
		-0.127	0.148	.025	-0.211	0.280	.041				-0.195	0.214	.018
		(CN/C)											
		-0.286	0.344	.027									

Note. The measures included in the table are only the statistically significant ones

4.1.1.2 Comparison changes in syntactic complexity between low- and high-proficiency groups.

4.1.1.2.1 Low-proficiency groups.

Table 12 presents two statistically significant findings of the ANOVA test, indicating which of the three low-proficiency groups' (TLC1, BLC1, and OLC1) syntactic complexity developed or declined the most in the classification and argumentative tasks. As presented in the table, in the classification task, one metric (DC/C) showed that while the syntactic complexity of BLC1 (M = -0.114) decreased the most in the short term, followed by OLC1 (M = -0.050), that of TLC1 improved (M = 0.096); the p-value was < .029. This reveals that in the short term the TLC was the most beneficial, whereas the BLC was the least beneficial, in increasing syntactic complexity in the classification task.

On the other hand, in the argumentative task, one measure (CP/C) indicated that the syntactic complexity of TLC1 (M = -0.054)—followed closely by OLC1 (M = -0.051)—deteriorated the most in the long term, but that of BLC1 was the only group that developed (M = 0.103); the p-value was < .022. This means that in the long term, the BLC was the most successful, while the TLC was the least successful, in improving syntactic complexity in the argumentative task. All the above-mentioned results are graphically presented in Figure 4.

Table 12

TLC1, BLC1, and OLC1 Syntactic Complexity Metrics: Significant Results of the ANOVA Test

Writing Tasks	Classification Task				Argumentative Task			
Test Comparisons	Pre to Mid				Pre to Post			
Metrics	**(DC/C)**				**(CP/C)**			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	0.096	-0.114*	-0.050	-0.023	-0.054*	0.103	-0.051	-0.001
SD	0.179	0.234	0.219	0.225	0.211	0.169	0.118	0.182
P-Value	0.029				0.022			
(*) indicates the most increase or decrease								

Figure 4. TLC1, BLC1, and OLC1 Syntactic Complexity Metrics: Significant Results of the ANOVA Test
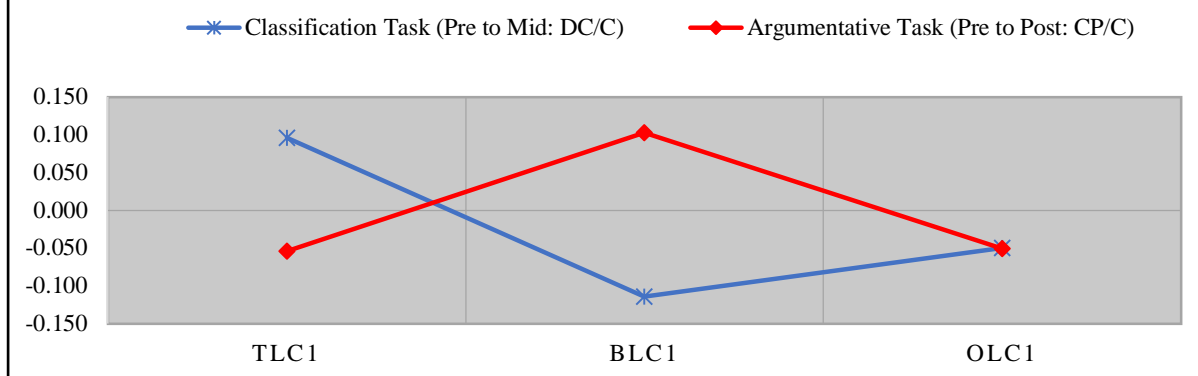
In sum, in enhancing the syntactic complexity of the low-proficiency Saudi EFL participants in the classification task, the TLC was the most effective, whereas the BLC was the least effective in the short term. In contrast, in promoting such learners' syntactic complexity in the argumentative task, the BLC was the most beneficial, but the TLC was the least beneficial in the long term.

4.1.1.2.2 High-proficiency groups.

Table 13 shows two statistically significant findings of the ANOVA test, revealing which of the three high-proficiency groups' (TLC6, BLC6, and OLC6) syntactic complexity increased or declined the most in only the classification task. As indicated in the table, in this writing task, the DC/C metric for the short term showed that TLC6 (M = 0.065), followed closely by BLC6 (M = 0.058), developed the most, while OLC6 decreased (M = -0.097); the p-value was < .016. Besides, despite the significant improvement of the syntactic complexity in OLC6 (M = 0.286), and BLC6 (M = 0.132) in the CN/C metric, TLC6 deteriorated (M = -0.181); the p-value was < .036. These results mean that each of the TLC6 and OLC6 participants increased the most in one of the syntactic complexity metrics compared to the BLC6 participants. However, in the short term, the BLC was regarded as the most successful in developing syntactic complexity in the classification task, since its participants improved in two metrics, conversely to the TLC and OLC, where the participants' scores declined in one metric. Nevertheless, in the argumentative task, there were no statistically significant differences in syntactic complexity scores between the three groups in the short term and long term, revealing that level of effectiveness/ineffectiveness of the three learning contexts in

increasing this construct was the same. [6] All of these findings are graphically shown in Figure 5.

Table 13								
TLC6, BLC6, and OLC6 Syntactic Complexity Metrics: Significant Results of the ANOVA Test								
Writing Tasks	Classification Task							
Test Comparisons	Pre to Mid							
Metrics	**(DC/C)**				**(CN/C)**			
Groups	TLC6	BLC6	OLC6	Total	TLC6	BLC6	OLC6	Total
Mean	0.065*	0.058	-0.097	0.009	-0.181	0.132	0.286*	0.079
SD	0.135	0.108	0.147	0.147	0.443	0.370	0.344	0.423
P-Value	.016				.036			
(*) indicates the most increase or decrease								
Note. There were no statistically significant differences between the three high-proficiency learning context groups in syntactic complexity in the argumentative task.								

Figure 5. TLC6, BLC6, and OLC6 Syntactic Complexity Metrics: Significant Results of the ANOVA Test
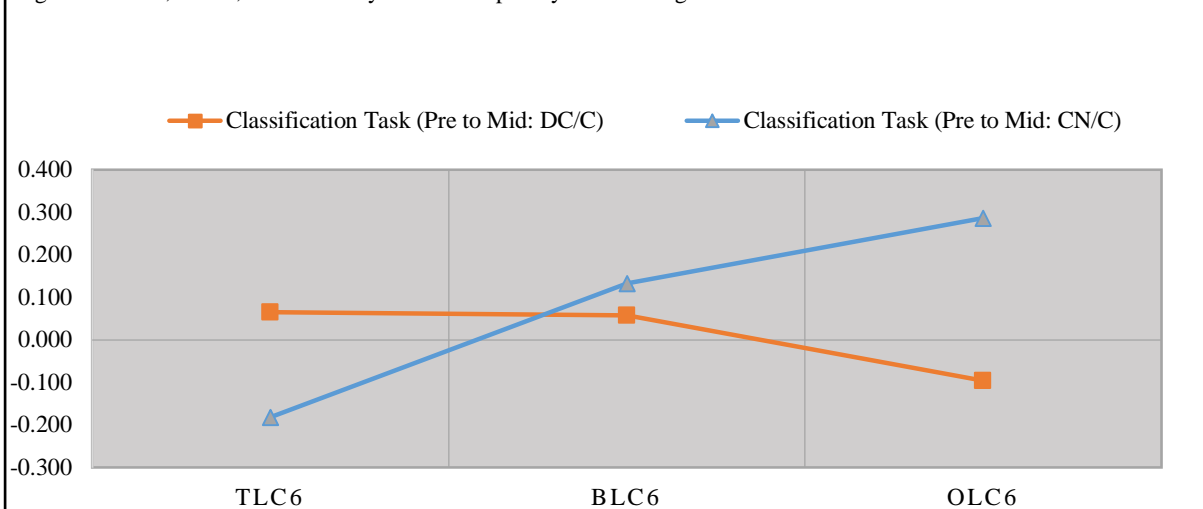
To conclude, the BLC stood out as the most beneficial in enhancing the syntactic complexity of the high-proficiency Saudi EFL undergraduates in the classification task in the short term, whereas both the TLC and OLC were the least beneficial. The reason for this is, although the syntactic complexity of the TLC and OLC groups (TLC6 and OLC6) developed in one measure more than that of the BLC group (BLC6), it also decreased for both groups in one other measure. The syntactic complexity of the BLC group, on the other hand, improved in two measures. Yet, all of the TLC, BLC, and OLC had the same level of success/unsuccess in promoting the syntactic complexity of their learners in the argumentative task in both the short term and long term.

---

[6] Whenever there were no statistically significant differences among the three learning environments' scores in any constructs or sub-constructs, it was labelled as a similarity. However, we use the term "effectiveness/ineffectiveness," or "success/unsuccess." The reason for using such antonyms is that it was difficult to employ only a positive or a negative term as all of these learning context groups either decreased in one measure of a construct, while increasing in its other measure(s), or one/two groups decreased, whereas the other group increased in one measure.

4.1.2 Lexical complexity.

4.1.2.1 Lexical density and sophistication.

4.1.2.1.1 Comparative changes in lexical density and sophistication within the low- and high-proficiency groups.

Table 14 indicates the statistically significant t-test results regarding the lexical density and sophistication for all six groups (TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6) in the two composition tasks.

4.1.2.1.1.1 Low-proficiency groups.

At the low-proficiency level, TLC1's findings did not reveal any statistically significant scores in lexical density across the two writing tasks in the short term and long term. This means that whether the task was non-complex or complex, the lexical density of TLC1 remained unchanged. Nonetheless, in the short term, of the five metrics used for lexical sophistication, only LS1 and LS2 showed statistically significant scores for the classification task. These metrics deteriorated significantly with means and p-values: LS1 (M = -0.089, $p <$ .016) and LS2 (M = -0.070, $p <$ .034). Likewise, in the argumentative task, both LS1 (M = -0.062, $p <$ .002) and LS2 (M = -0.071, $p <$ .001) declined significantly. All of these results indicate that regardless of the task being simple or difficult, TLC1's lexical sophistication decreased equally. In contrast, in the long term, only LS2 (M = -0.057, $p <$ .041) deteriorated significantly in the classification task, but in the argumentative task both LS1 and LS2 (M = -0.044, -0.053, and $p <$ .010, and $<$ .007, respectively) declined significantly. These findings reveal that the lexical sophistication of TLC1 decreased in the non-complex task less than in the complex task.

For BLC1, there was one statistically significant score, in the short term, for lexical density in the classification task: LD measure increased significantly (M = 0.035, $p <$ .015), while in the argumentative task, there was no statistically significant score. This shows that BLC1's lexical density increased in the simple task, whereas it remained unaffected in the difficult task. However, in the long term, there were no statistically significant scores for lexical density in both text types, which indicates that this sub-construct remained unchanged in the non-complex and complex tasks for BLC1. With regard to lexical sophistication, four measures (LS1, LS2, VS2, and CVS1) in the classification task, and two measures (LS1 and LS2) in the argumentative task, revealed statistically significant scores in the short term. All of these measures deteriorated significantly: classification task (LS1 [M = -0.138, $p <$ .001], LS2 [M = -0.122, $p <$ .000], VS2 [M = -0.489, $p <$ .021], and CVS1 [M = -0.235, $p <$ .033]),

and argumentative task (LS1 [M = -0.063, p < .006], and LS2 [M = -0.050, p < .007]). All of these results show that the lexical sophistication of BLC1 declined more in the simple task than in the difficult task. Nevertheless, in the long term, LS2 decreased significantly (M = -0.089, p < .002) in the classification task, and LS1 decreased significantly (M = -0.072, p < .010) in the argumentative task. Such findings mean that whether the task was non-complex or complex, BLC1's lexical sophistication deteriorated equally.

In the case of OLC1, there were no statistically significant scores with regard to lexical density in the classification and argumentative tasks in the short term and long term, which indicates that regardless of the task being simple or difficult, this sub-construct remained uninfluenced for OLC1. For lexical sophistication, two metrics declined significantly (LS1: M = -0.075, p < .027, and LS2: M = -0.055, p < .049) in the classification task in the short term, but in the argumentative task, there were no statistically significant scores. These results reveal that the lexical sophistication of OLC1 decreased in the non-complex task, while it remained unaffected in the complex task. Conversely, in the long term, two metrics (VS2 and CVS1) in the classification task showed statistically significant scores; each of which developed significantly: VS2 (M = 0.630, p < .046) and CVS1 (M = 0.282, p < .017). In the argumentative task, four metrics deteriorated significantly: LS1 (M = -0.073, p < .042), VS1 (M = -0.083, p < .041), VS2 (M = -0.617, p < .023), CVS1 (M = -0.269, p < .021). Such findings mean that OLC1's lexical sophistication improved in the simple task, whereas it declined in the difficult task.

All things considered, BLC1, in comparison to TLC1 and OLC1, was the only group in which lexical density was impacted by the task complexity factor. To clarify, the lexical density of both TLC1 and OLC1 remained unchanged across the two composition tasks in both the short term and long term, and so did BLC1's lexical density in the long term. Yet, in the short term, this sub-construct increased for BLC1 in the non-complex task, but it remained uninfluenced in the complex task. In contrast, the lexical sophistication of the low-proficiency Saudi EFL participants in the three learning environments was influenced by level of task complexity across different timescales. That is, TLC1's lexical sophistication decreased in the simple task less than in the difficult task, in the long term. The lexical sophistication of BLC1 in the short term, decreased in the non-complex task more than in the complex task. OLC1's lexical sophistication, in the short term, deteriorated in the simple task, while it remained unaffected in the difficult task. In the long term, this sub-construct developed for OLC1 in the non-complex task, whereas it declined in the complex task. Lastly, no effect of the task

complexity factor on the lexical sophistication of TLC1 in the short term and of BLC1 in the long term was found as it decreased equally in both writing tasks.

4.1.2.1.1.2 High-proficiency groups.

At the high-proficiency level, TLC6's results indicated that, in the short term of lexical density, there was no statistically significant score in the classification task, but one measure (LD) deteriorated significantly (M = -0.035, $p < .005$) in the argumentative task. This reveals that the lexical density of TLC6 remained unchanged in the non-complex task, while it declined in the complex task. Nonetheless, in the long term, there were no statistically significant scores in both composition tasks. This shows that this sub-construct remained uninfluenced in the simple and difficult tasks for TLC6. With regard to lexical sophistication, on the other hand, there were no statistically significant scores for the classification task, whereas in the argumentative task, two measures (LS1 and LS2) exhibited significant decreases in the short term: (M = -0.059, $p < .020$, and M = -0.050, $p < .011$, respectively) and in the long term: (M = -0.127, $p < .002$, and M = -0.113, $p < .002$, respectively). These findings indicate that TLC6's lexical sophistication remained unaffected in the non-complex task, but it deteriorated in the complex task.

For BLC6, there were no statistically significant scores for lexical density in both the classification and argumentative tasks in the short term and long term. This means that whether the task was simple or difficult, such a sub-construct remained unchanged for BLC6. In terms of lexical sophistication, two metrics revealed statistically significant scores (decline) in either the short term (LS1: M = -0.065, $p < .008$) or the long term (LS2: M = -0.066, $p < .007$) for the classification task. In the argumentative task, the above-mentioned metrics showed significant decreases in both the short term and long term: LS1 (M = -0.052, $p < .016$, and M = -0.093, $p < .009$, respectively) and LS2 (M = -0.069, $p < .001$, and M = -0.075, $p < .001$, respectively). These results indicate that the BLC6's lexical sophistication deteriorated less in the non-complex task than in the complex task.

As regards the lexical density of OLC6, there were no statistically significant scores for both the classification and argumentative tasks in the short term and long term. This reveals that regardless of the task being simple or difficult, this sub-construct remained uninfluenced for OLC6. For lexical sophistication, one measure (LS1) declined significantly (M = -0.072, and $p < .011$) in the classification task, in the short term, while there were no statistically significant scores for the argumentative task. This shows that the lexical sophistication of OLC6 decreased in the non-complex task, whereas it remained unaffected in the complex

task. However, in the long term, two measures (LS1 and LS2) indicated statistically significant scores (deterioration) for the classification task: LS1 (M = -0.087, p < .028) and LS2 (M = -0.151, p < .000), as well as for the argumentative task: LS1 (M = -0.106, p < .001) and LS2 (M = -0.108, p < .000). All of these findings reveal that the OLC6's lexical sophistication equally declined in the simple and difficult tasks.

In brief, TLC6, in comparison to BLC6 and OLC6, was the only group in which lexical density was impacted by level of task complexity in one timescale. To elucidate, the lexical density of TLC1 remained unchanged in the non-complex task, but it decreased in the complex task. This sub-construct remained uninfluenced in both writing tasks for TLC6 in the long term, and for BLC6 and OLC6 in both the short term and long term. Conversely, the lexical sophistication among the high-proficiency Saudi EFL undergraduates in the three learning contexts was influenced by the task complexity factor across various timescales. TLC6's lexical sophistication, in the short term and long term, remained unaffected in the simple task, while it deteriorated in the difficult task. The lexical sophistication of BLC6 in the short term and long term declined less in the non-complex task than in the complex task. Again, the OLC6's lexical sophistication decreased in the simple task in the short term, whereas it remained unchanged in the difficult task. Nevertheless, no effect of level of task complexity on such a sub-construct was found in OLC6 in the long term because it deteriorated equally in both text types.

Table 14

TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6 Lexical Density and Sophistication Metrics: Significant Results of the T-Test

Learning Contexts			Pre-Test to Mid-Term Test						Pre-Test to Post-Test					
			Classification Task			Argumentative Task			Classification Task			Argumentative Task		
			M	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*
Proficiency Levels — Low-Proficiency Groups	Traditional Learning Context (TLC1)		(LS1)			(LS1)			(LS2)			(LS1)		
			-0.089	0.127	.016	-0.062	0.064	.002	-0.057	0.098	.041	-0.044	0.058	.010
			(LS2)			(LS2)						(LS2)		
			-0.070	0.115	.034	-0.071	0.065	.001				-0.053	0.066	.007
	Blended Learning Context (BLC1)		(LD)			(LS1)			(LS2)			(LS1)		
			0.035	0.049	.015	-0.063	0.075	.006	-0.089	0.091	.002	-0.072	0.094	.010
			(LS1)			(LS2)								
			-0.138	0.127	.001	-0.050	0.062	.007						
			(LS2)											
			-0.122	0.092	.000									
			(VS2)											
			-0.489	0.728	.021									
			(CVS1)											
			-0.235	0.386	.033									
	Online Learning Context (OLC1)		(LS1)						(VS2)			(LS1)		
			-0.075	0.117	.027				0.630	1.114	.046	-0.073	0.126	.042
			(LS2)						(CVS1)			(VS1)		
			-0.055	0.098	.049				0.282	0.404	.017	-0.083	0.142	.041
												(VS2)		
												-0.617	0.935	.023
												(CVS1)		
												-0.269	0.401	.021

163

Learning Contexts		Pre-Test to Mid-Term Test						Pre-Test to Post-Test						
		Classification Task			Argumentative Task			Classification Task			Argumentative Task		
		M	SD	P	M	SD	P	M	SD	P	M	SD	P
High-Proficiency Groups	Traditional Learning Context (TLC6)				**(LD)**						**(LS1)**		
					-0.035	0.030	.005				-0.127	0.089	.002
					(LS1)						**(LS2)**		
					-0.059	0.066	.020				-0.113	0.082	.002
					(LS2)								
					-0.050	0.050	.011						
	Blended Learning Context (BLC6)	**(LS1)**			**(LS1)**			**(LS2)**			**(LS1)**		
		-.065	0.060	.008	-0.052	0.056	.016	-0.066	0.059	.007	-0.093	0.088	.009
					(LS2)						**(LS2)**		
					-0.069	0.043	.001				-0.075	0.049	.001
	Online Learning Context (OLC6)	**(LS1)**						**(LS1)**			**(LS1)**		
		-.072	0.071	.011				-.087	0.105	.028	-0.106	0.069	.001
								(LS2)			**(LS2)**		
								-0.151	0.090	.000	-0.108	0.054	.000

Note. The measures included in the table are only the statistically significant ones

4.1.2.1.2 Comparative changes in lexical density and sophistication between the low- and high-proficiency groups.

4.1.2.1.2.1 Low-proficiency groups.

Table 15 displays two statistically significant results of the ANOVA test, showing which of the three low-proficiency groups' lexical sophistication improved or declined the most in the classification task. As presented in the table, across the two timescales, there were no statistically significant differences between the three groups' lexical density scores in the two composition tasks, nor were there any differences between these groups' lexical sophistication scores in the argumentative task. This means that the three learning environments' level of effectiveness/ineffectiveness in increasing lexical density in the classification and argumentative tasks and in increasing lexical sophistication in the argumentative task, in the short term and long term, was the same. Yet, in the classification task, two metrics (VS2 and CVS1) of lexical sophistication indicated statistically significant scores in the long term: VS2 developed the most in OLC1 (M = 0.630), followed by TLC1 (M = 0.051), but it decreased in BLC1 (M = -0.447); the p-value was < .022. Similarly, OLC1 (M = 0.282) increased the most in CVS1, followed by TLC1 (M = 0.089), while it deteriorated for BLC1 (M = -0.151); the p-value was < .020. These findings mean that in the long term, the OLC was the most beneficial, whereas the BLC was the least beneficial, in improving lexical sophistication in the classification task. The aforementioned results are graphically presented in Figure 6.

Table 15

TLC1, BLC1, and OLC1 Lexical Density and Sophistication Metrics: Significant Results of the ANOVA Test

Writing Tasks	Classification Task							
Test Comparisons	Pre to Post							
Metrics	(VS2)				(CVS1)			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	0.051	-0.447	0.630*	0.078	0.089	-0.151	0.282*	0.073
SD	1.111	0.813	1.114	1.094	0.373	0.438	0.404	0.435
P-Value	.022				.020			

(*) indicates the most increase or decrease
Note. There were no statistically significant differences between the three low-proficiency learning context groups in lexical density across the two writing tasks, nor in lexical sophistication in the argumentative task.

Figure 6. TLC1, BLC1, and OLC1 Lexical Density and Sophistication Metrics: Significant Results of the ANOVA Test
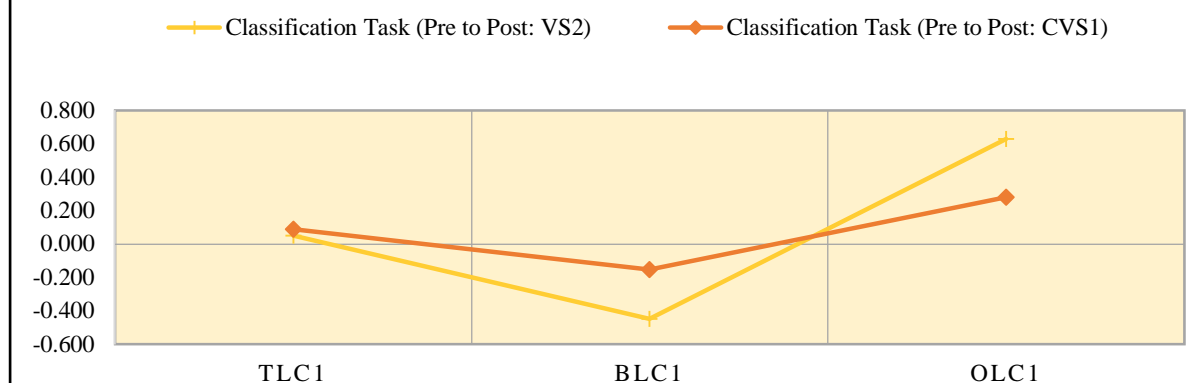
On the whole, the TLC, BLC, and OLC, had the same level of success/unsuccess in enhancing the lexical density of the low-proficiency participants in the classification and argumentative tasks, as well as their lexical sophistication in the argumentative task in the short term and long term. In contrast, in promoting the lexical sophistication of such learners in the classification task in the long term, the OLC was the most effective, but the BLC was the least effective.

4.1.2.1.2.1 High-proficiency groups.

Table 16 presents one statistically significant finding of the ANOVA test, revealing which of the high-proficiency groups' lexical sophistication increased or declined the most in the classification task. As shown in the table, in the two timescales, there were no statistically significant differences between the three groups' lexical density scores across the two writing tasks, nor were there any differences between these groups' lexical sophistication scores in the argumentative task. This implies that the three learning contexts had the same level of effectiveness/ineffectiveness in developing lexical density in the classification and argumentative task and in developing lexical sophistication in the argumentative task, in the short term and long term. Nonetheless, in the classification task, only one measure of lexical sophistication indicated a statistically significant score in the long term: LS2 deteriorated the most in OLC6 (M = -0.151), followed by BLC6 (M = -0.066) and, finally, TLC6 (M = -0.045); the p-value was < .008. This means that all of three learning environments were not successful in improving lexical sophistication in the classification task in the long term with the OLC being the least successful, followed by the BLC, and lastly, the TLC. Figure 7 graphically shows all of these results.

Table 16

TLC6, BLC6, and OLC6 Lexical Density and Sophistication Metrics: Significant Results of the ANOVA Test

Writing Tasks	Classification Task			
Test Comparisons	Pre to Post			
Metrics	(LS2)			
Groups	TLC6	BLC6	OLC6	Total
Mean	-0.045	-0.066	-0.151*	-0.087
SD	0.069	0.059	0.090	0.085
P-Value	.008			

(*) indicates the most increase or decrease
Note. There were no statistically significant differences between the three high-proficiency learning context groups in lexical density across the two writing tasks, nor in lexical sophistication in the argumentative task.

Figure 7. TLC6, BLC6, and OLC6 Lexical Density and Sophistication Metrics: Significant Results of the ANOVA Test
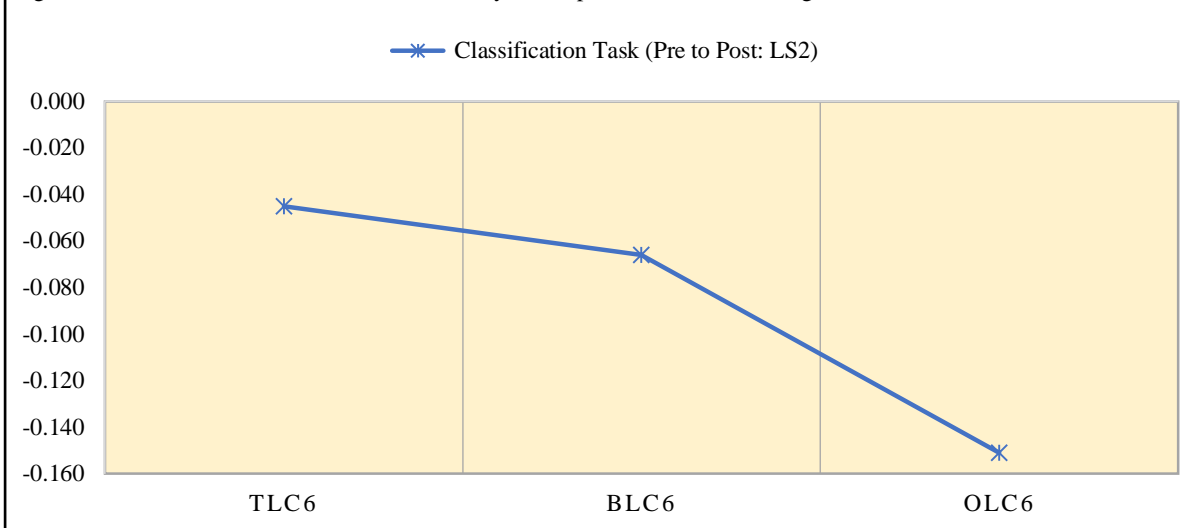
In short, the TLC, BLC, and OLC's level of success/unsuccess in enhancing the lexical density of the high-proficiency Saudi EFL undergraduates in the classification and argumentative task, and their lexical sophistication in the argumentative task, in the short term and long term, was the same. However, none of these learning contexts was beneficial in promoting the lexical sophistication of such learners in the classification task in the long term with the OLC being the least beneficial, followed by the BLC, and finally, the TLC.

4.1.2.2 Lexical variation.

4.1.2.2.1 Comparative changes in lexical variation within the low- and high-proficiency groups.

Table 17 reveals the statistically significant t-test findings on the lexical variation for all six groups (TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6) in the two text types.

4.1.2.2.1.1 Low-proficiency groups.

At the low-proficiency level, TLC1's results indicated that, in the classification task, of the 19 metrics, only two metrics (VV1, and AdjV) showed statistically significant scores

(decrease) in the short term: (M = -0.091 and -0.039, respectively; *p* < .013 and .012, respectively). In the argumentative task, five metrics (NDW, TTR, VV1, AdjV, and ModV) revealed statistically significant scores: while NDW deteriorated significantly (M = -15.533, *p* < .039), TTR (M = 0.043, *p* < .016), VV1 (M = 0.104, *p* < .020), AdjV (M = 0.034, *p* < .016), and ModV (M = 0.042, *p* < .013) increased significantly. Similarly, in the long term, LV declined significantly (M = -0.051, *p* < .043) in the classification task, whereas NDW (M = -16.333, *p* < .039), VV1 (M = 0.117, *p* < .008), and ModV (M = 0.029, *p* < .020) either decreased or developed significantly in the argumentative task. In all of these findings, the lexical variation of TLC1 deteriorated in the non-complex task, but it improved more than it declined in the complex task.

For BLC1, in the classification task, seven measures indicated significant scores in the short term, all of which developed significantly: NDWZ (M = 3.600, *p* < .026), NDWESZ (M = 2.420, *p* < .017), TTR (M = 0.065, *p* < .007), MSTTR (M = 0.045, *p* < .011), LogTTR (M = 0.023, *p* < .011), Uber (M = 2.235, *p* < .034), and VV1 (M = 0.133, *p* < .031). In the argumentative task, there were no statistically significant scores. Such results mean that BLC1's lexical variation increased in the simple task, while it remained uninfluenced in the difficult task. In contrast, in the long term, there were no statistically significant scores for the classification task, whereas one measure (NDW) showed significant development (M = 15.133, *p* < .016) for the argumentative task. This implies that the lexical variation of BLC1 remained unaffected in the non-complex task, but it improved in the complex task.

In the case of OLC1, five metrics (NDWERZ, NDWESZ, MSTTR, AdjV, and ModV) revealed statistically significant scores in the classification task in the short term. Some of these metrics increased significantly, such as NDWERZ (M = 3.207, *p* < .001), NDWESZ (M = 2.673, *p* < .011), and MSTTR (M = 0.049, *p* < .010), while others decreased significantly, such as AdjV (M = -0.032, *p* < .047) and ModV (M = -0.034, *p* < .040). In the argumentative task, there were no statistically significant scores. All of these findings indicate that OLC1's lexical variation developed more than it deteriorated in the simple task, whereas it remained unchanged in the difficult task. Conversely, in the long term, two metrics (NDWERZ and VV2) in the classification task and two other (SVV1 and CVV1) in the argumentative task showed statistically significant scores: both NDWERZ (M = 1.340, *p* < .025) and VV2 (M = 0.028, *p* < .047) improved significantly, but both SVV1 (M = -2.065, *p* < .022) and CVV1 (M = -0.226, *p* < .018) declined significantly. Such results reveal that the lexical variation of OLC1 increased in the non-complex task, while it decreased in the complex task.

To sum up, the lexical variation of the low-proficiency Saudi EFL participants in the three learning environments was impacted by the task complexity factor differently in the two timescales. For example, in the short term, TLC1's lexical variation deteriorated in the simple task, whereas it developed more than it declined in the difficult task. The lexical variation of BLC1 improved in the non-complex task, but it remained uninfluenced in the complex task. OLC1's lexical variation increased more than it decreased in the simple task, while it remained unaffected in the difficult task. Moreover, in the long term, the lexical variation of TLC1 deteriorated in the non-complex task, whereas it developed more than it declined in the complex task. BLC1's lexical variation remained unchanged in the simple task, but it improved in the difficult task. Lastly, the lexical variation of OLC1 increased in the non-complex task, while it decreased in the complex task.

4.1.2.2.1.2 High-proficiency groups.

At the high-proficiency level, TLC6's findings revealed that, in the classification task, two measures (MSTTR and VV2) developed significantly (M = 0.046 and 0.041, respectively; $p < .012$ and .044, respectively) in the short term, whereas there were no statistically significant scores for the argumentative task. Such results mean that TLC6's lexical variation improved in the simple task, but it remained uninfluenced in the difficult task. Nevertheless, in the long term, seven measures indicated statistically significant scores of an increase (VV2: M = 0.023, $p < .043$), and decreases (CTTR: M = -0.549, $p < .004$, RTTR: M = -0.778, $p < .004$, LogTTR: M = -0.022, $p < .021$, Uber: M = -2.812, $p < .020$, LV: M = -0.080, $p < .031$, and VV1: M = -0.090, $p < .027$) for the classification task. In the argumentative task, on the other hand, five measures (NDW, CTTR, RTTR, SVV1, and CVV1) showed statistically significant scores, all of which deteriorated significantly: NDW (M = -18.000, $p < .001$), CTTR (M = -0.426, $p < .033$), RTTR (M = -0.605, $p < .033$), SVV1 (M = -2.982, $p < .033$), and CVV1 (M = -0.305, $p < .032$). All of these findings imply that the lexical variation of TLC6 declined more than it developed in the non-complex task, while it decreased in the complex task.

For BLC6, there were no statistically significant scores in the short term for the classification task, whereas in the argumentative task, seven metrics deteriorated significantly: NDWERZ (M = -1.450, $p < .039$), CTTR (M = -0.392, $p < .019$), RTTR (M = -0.552, $p < .020$), Uber (M = -1.914, $p < .030$), NV (M = -0.114, $p < .050$), AdjV (M = -0.021, $p < .023$), and ModV (M = -0.038, $p < .014$). Such results mean that BLC6's lexical variation remained unaffected in the simple task, but it declined in the difficult task. Yet, in the long term, two

metrics improved significantly (NDWERZ: M = 1.230, $p < .032$, and VV2: M = 0.048, $p <$ .002) in the classification task, while in the argumentative task, VV1 decreased significantly (M = -0.120; $p < .049$). All of these findings reveal that the lexical variation of BLC6 increased in the non-complex task, whereas it deteriorated in the complex task.

With regard to OLC6, in the classification task, six measures (NDWZ, TTR, CTTR, RTTR, LogTTR, and LV) indicated significant scores in the short term. All of these measures developed significantly: NDWZ (M = 5.100, $p < .004$), TTR (M = 0.060, $p < .017$), CTTR (M = 0.426, $p < .034$), RTTR (M = 0.603, $p < .034$), LogTTR (M = 0.022, $p < .006$), and LV (M = 0.063, $p < .041$). In the argumentative task, one measure (NV) declined significantly (M = -0.078, $p < .043$). Likewise, in the long term, both NDWZ and VV2 improved significantly (M = 3.100 and 0.026, respectively; $p < .041$ and .038, respectively) for the classification task, but NV decreased significantly (M = -0.060, and $p < .041$) for the argumentative task. In all of these results, OLC6's lexical variation increased in the simple task, while it deteriorated in the difficult task.

Overall, the lexical variation of the high-proficiency Saudi EFL undergraduates in the three learning contexts was influenced by level of task complexity almost differently in the two timescales. For instance, in the short term, TLC6's lexical variation developed in the non-complex task, whereas it remained unchanged in the complex task. The lexical variation of BLC6 remained uninfluenced in the simple task, but it declined in the difficult task. OLC6's lexical variation improved in non-complex task, while it decreased in the complex task. Furthermore, in the long term, the lexical variation of TLC6 deteriorated more than it increased in the simple task, whereas it declined in the difficult task. Finally, both BLC6 and OLC6's lexical variation developed in the non-complex task, but it decreased in the complex task.

Table 17

TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6 Lexical Variation Metrics: Significant Results of the T-Test

Learning Contexts			Pre-Test to Mid-Term Test						Pre-Test to Post-Test					
			Classification Task			Argumentative Task			Classification Task			Argumentative Task		
			M	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*	*M*	*SD*	*P*
Proficiency Levels	Low-Proficiency Groups	Traditional Learning Context (TLC1)	(VV1)			(NWD)			(LV)			(NWD)		
			-0.091	0.123	.013	-15.533	26.446	.039	-0.051	0.088	.043	-16.333	27.699	.039
			(AdjV)			(TTR)						(VV1)		
			-0.039	0.052	.012	0.043	0.060	.016				0.117	0.145	.008
						(VV1)						(ModV)		
						0.104	0.153	.020				0.029	0.043	.020
						(AdjV)								
						0.034	0.048	.016						
						(ModV)								
						0.042	0.057	.013						
		Blended Learning Context (BLC1)	(NDWZ)									(NDW)		
			3.600	5.604	.026							15.133	21.387	.016
			(NDWESZ)											
			2.420	3.473	.017									
			(TTR)											
			0.065	0.081	.007									
			(MSTTR)											
			0.045	0.060	.011									
			(LogTTR)											
			0.023	0.031	.011									
			(Uber)											
			2.235	3.682	.034									
			(VV1)											
			0.133	0.215	.031									
		Online Learning Context (OLC1)	(NDWERZ)						(NDWERZ)			(SVV1)		
			3.207	3.027	.001				1.340	2.066	.025	-2.065	3.119	.022
			(NDWESZ)						(VV2)			(CVV1)		
			2.673	3.548	.011				0.028	0.050	.047	-0.226	0.325	.018
			(MSTTR)											
			0.049	0.064	.010									
			(AdjV)											
			-0.032	0.057	.047									
			(ModV)											
			-0.034	0.058	.040									

| Learning Contexts | Pre-Test to Mid-Term Test | | | | | | Pre-Test to Post-Test | | | | | |
| | Classification Task | | | Argumentative Task | | | Classification Task | | | Argumentative Task | | |
	M	SD	P	M	SD	P	M	SD	P	M	SD	P
Traditional Learning Context (TLC6)	(MSTTR)						(CTTR)			(NDW)		
	0.046	0.046	.012				-0.549	0.448	.004	-18.000	11.926	.001
	(VV2)						(RTTR)			(CTTR)		
	0.041	0.055	.044				-0.778	0.633	.004	-0.426	0.536	.033
							(LogTTR)			(RTTR)		
							-0.022	0.025	.021	-0.605	0.761	.033
							(Uber)			(SVV1)		
							-2.812	3.148	.020	-2.982	3.760	.033
							(LV)			(CVV1)		
							-0.080	0.099	.031	-0.305	0.382	.032
							(VV1)					
							-0.090	0.108	.027			
							(VV2)					
							0.023	0.031	.043			
Blended Learning Context (BLC6)				(NDWERZ)			(NDWERZ)			(VV1)		
				-1.450	1.895	.039	1.230	1.531	.032	-0.120	0.167	.049
				(CTTR)			(VV2)					
				-0.392	0.436	.019	0.048	0.034	.002			
				(RTTR)								
				-0.552	0.618	.020						
				(Uber)								
				-1.914	2.344	.030						
				(NV)								
				-0.114	0.159	.050						
				(AdjV)								
				-0.021	0.024	.023						
				(ModV)								
				-0.038	0.040	.014						
Online Learning Context (OLC6)	(NDWZ)			(NV)			(NDWZ)			(NV)		
	5.100	4.175	.004	-0.078	0.105	.043	3.100	4.122	.041	-0.060	0.080	.041
	(TTR)						(VV2)					
	0.060	0.065	.017				0.026	0.034	.038			
	(CTTR)											
	0.426	0.540	.034									
	(RTTR)											
	0.603	0.763	.034									
	(LogTTR)											
	0.022	0.019	.006									

Left margin label: High-Proficiency Groups

			(LV)									
			0.063	0.084	.041							

Note. The measures included in the table are only the statistically significant ones

4.1.2.2.2 Comparative changes in lexical variation between the low- and high-proficiency groups.

4.1.2.2.2.1 Low-proficiency groups.

Table 18 shows ten statistically significant findings of the ANOVA test, revealing which of the low-proficiency groups' lexical variation improved or deteriorated the most in the two composition tasks. As presented in the table, in the classification task, six metrics (NDWERZ, NDWESZ, MSTTR, Uber, VV1, and ModV) indicated statistically significant scores in the short term: NDWERZ increased for all three groups with OLC1 (M = 3.207) increasing the most, followed by BLC1 (M = 1.213) and, lastly, TLC1 (M = 0.160). The p-value was < .031. NDWESZ developed the most for OLC1 (M = 2.673), followed by BLC1 (M = 2.420), while it declined for TLC1 (M = -0.220); the p-value was < .039. In addition, MSTTR improved for OLC1 (M = 0.049), followed closely by BLC1 (M = 0.045), whereas it decreased for TLC1 (M = -0.003). The p-value was < .028. Uber increased the most for BLC1 (M = 2.235), followed by OLC1 (M = 2.176), but it deteriorated for TLC1 (M = -1.136). The p-value was < .049. VV1 developed for only BLC1 (M = 0.133); in fact, it declined in both TLC1 (M = -0.091) and OLC1 (M = -0.002). The p-value was < .009. Finally, while ModV decreased in both OLC1 (M = -0.034) and, with an insignificant difference, TLC1 (M = -0.033), it improved for BLC1 (M = 0.013). The p-value was < .049. The above-mentioned results show that BLC1's lexical variation increased in six metrics (three of which were the most), whereas the sub-construct of TLC1 deteriorated in five metrics. These findings reveal that in the short term the BLC was the most successful, but the TLC was the least successful, in developing lexical variation in the classification task.

In the argumentative task, four measures (NDW, VV1, SVV1, and CVV1) indicated statistically significant scores in the long term: NDW showed that while TLC1 (M = -16.333) declined the most, followed by OLC1 (M = -4.400), BLC1 improved (M = 15.133); the p-value was < .002. VV1 decreased the most for OLC1 (M =-0.065), followed by BLC1 (M = -0.006), whereas it increased for TLC1 (M = 0.117); the p-value was < .015. The SVV1 measure deteriorated the most for OLC1 (M = -2.065), followed by TLC1 (M = -1.353), but it developed for BLC1 (M =2.338); the p-value was < .020. Similarly, CVV1 declined the most for OLC1 (M = -0.226), followed by TLC1 (M = -0.142), while it improved for BLC1 (M = 0.272); the p-value was < .017. Thus, of four lexical variation measures, BLC1 was the only group to increase in three of them, whereas OLC1 solely decreased in all of them. This means that in the argumentative task, the BLC was the most effective, but the OLC was the least

effective, in developing lexical variation in the long term. All of these results are graphically presented in Figure 8.

Table 18

TLC1, BLC1, and OLC1 Lexical Variation Metrics: Significant Results of the ANOVA Test

Writing Tasks	Classification Task							
Test Comparisons	Pre to Mid							
Metrics	(NDWERZ)				(NDWESZ)			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	0.160	1.213	3.207*	1.527	-0.220	2.420	2.673*	1.624
SD	3.105	3.099	3.027	3.267	2.872	3.473	3.548	3.495
P-Value	0.031				0.039			
Metrics	(MSTTR)				(Uber)			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	-0.003	0.045	0.049*	0.030	-1.136	2.235*	2.176	1.092
SD	0.047	0.060	0.064	0.061	4.560	3.682	4.150	4.353
P-Value	0.028				0.049			
Metrics	(VV1)				(ModV)			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	-0.091*	0.133	-0.002	0.014	-0.033	0.013	-0.034*	-0.018
SD	0.123	0.215	0.214	0.207	0.066	0.047	0.058	0.060
P-Value	0.009				0.049			
Writing Tasks	Argumentative Task							
Test Comparisons	Pre to Post							
Metrics	(NDW)				(VV1)			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	-16.333*	15.133	-4.400	-1.867	0.117	-0.006	-0.065*	0.015
SD	27.699	21.387	19.29	26.079	0.145	0.195	0.152	0.179
P-Value	0.002				0.015			
Metrics	(SVV1)				(CVV1)			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	-1.353	2.338	-2.065*	-0.360	-0.142	0.272	-0.226*	-0.032
SD	5.337	4.524	3.119	4.741	0.584	0.517	0.325	0.525
P-Value	0.020				0.017			
(*) indicates the most increase or decrease								

Figure 8. TLC1, BLC1, and OLC1 Lexical Variation Metrics: Significant Results of the ANOVA Test
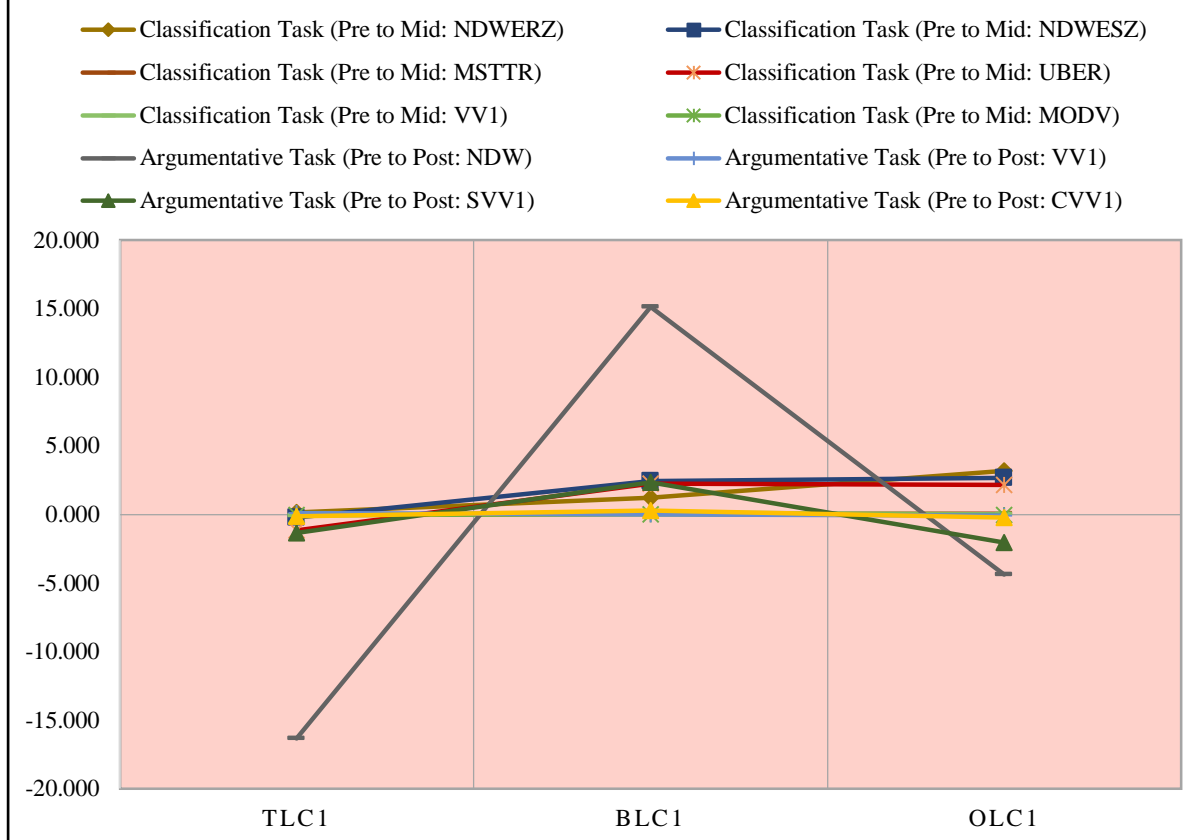
To summarize, in the classification task, the BLC was the most beneficial in enhancing the lexical variation among the low-proficiency Saudi EFL participants in the short term, while the TLC was the least beneficial. The reason is that the lexical variation of BLC learners improved in six metrics, whereas it increased in one metric and deteriorated in five others for the TLC learners. Besides, in the argumentative task, the former learning environment (BLC) was again the most successful in promoting its participants' lexical variation in the long term, but the OLC was the least successful. The justification is that the BLC group developed in three of four lexical variation metrics, while the OLC group declined in all of them.

4.1.2.2.2.2 High-proficiency groups.

Table 19 displays nine statistically significant findings of the ANOVA test, revealing which of the three high-proficiency group's lexical variation improved or decreased the most in the two writing tasks. As indicated in the table, in the classification task, five measures (TTR, CTTR, RTTR, LogTTR, and Uber) showed statistically significant scores in the long term: TTR measure deteriorated the most for TLC6 (M = -0.059), then BLC6 (M = -0.005), whereas it increased for OLC6 (M = 0.034). The p-value was < .030. In terms of CTTR,

TLC6 was the only group to decline (M = -0.549); BLC6 (M = 0.081) developed the most, followed by OLC6 group (M = 0.065). The p-value was < .009. In both RTTR and LogTTR, although BLC6 improved the most (M = 0.115 and 1.110, respectively), followed by OLC6 (M = 0.093 and 0.010, respectively), TLC6 decreased (M = -0.778 and -0.022, respectively). The p-values were < .009 and .011, respectively. Uber also deteriorated for TLC6 (M = -2.812); nonetheless, it increased the most for OLC6 (M = 0.792), followed by BLC6 (M = 0.171). The p-value was < .004. In all of these results, despite BLC6 developing the most in three of five measures and declining in one measure, OLC6 improved, but TLC6 decreased, in all measures. This reveals that the OLC was the most effective, while the TLC was the least effective, in increasing lexical variation in the classification task in the long term.

In the argumentative task, four metrics (ModV, NDW, SVV1, and CVV1) indicated statistically significant scores in the short term or the long term. In the short term, ModV deteriorated the most for BLC6 (M = -0.038), followed by OLC6 (M = -0.012), and developed for TLC6 (M = 0.012). The p-value was < .041. This shows that the TLC was the most beneficial, whereas the BLC was the least beneficial, in improving lexical variation in the argumentative task in the short term. In contrast, in the long term, NDW declined the most for TLC6 (M = -18.000), followed by BLC6 (M = -3.700), but it increased for OLC6 (M = 4.200). The p-value was < .005. Similarly, SVV1 developed for OLC6 (M = 1.406), while it decreased the most for TLC6 (M = -2.982), followed by BLC6 (M = -2.870). The p-value was < .041. Moreover, OLC6 was the only group to improve in CVV1 (M = 0.167), whereas TLC6 (M = -0.305) and BLC6 (M = -0.296) deteriorated the most. The p-value was < .036. In all of the above findings, OLC6 was the only group to increase in three metrics of lexical variation, but both TLC6 and BLC6 declined in all of such metrics. The decreases of TLC6 were the most. This means that the OLC was the most successful, while the TLC was the least successful, in developing lexical variation in the argumentative tasks in the long term. The above-mentioned results are graphically shown in Figure 9.

Table 19

TLC6, BLC6, and OLC6 Lexical Variation Metrics: Significant Results of the ANOVA Test

Writing Tasks	Classification Task							
Test Comparisons	Pre to Post							
Metrics	(TTR)				(CTTR)			
Groups	TLC6	BLC6	OLC6	Total	TLC6	BLC6	OLC6	Total
Mean	-0.059[*]	-0.005	0.034	-0.010	-0.549	0.081[*]	0.065	-0.134
SD	0.087	0.080	0.048	0.081	0.448	0.508	0.475	0.549
P-Value	0.030				0.009			

(continued)

Writing Tasks	Classification Task							
Test Comparison	Pre to Post							
Metrics	(RTTR)				(LogTTR)			
Groups	TLC6	BLC6	OLC6	Total	TLC6	BLC6	OLC6	Total
Mean	-0.778	0.115*	0.093	-0.190	-0.022	1.110*	0.010	-0.004
SD	0.633	0.719	0.671	0.777	0.025	0.023	0.019	0.026
P-Value	0.009				0.011			
Writing Tasks	Classification Task				Argumentative Task			
Test Comparisons	Pre to Post				Pre to Mid			
Metrics	(Uber)				(ModV)			
Groups	TLC6	BLC6	OLC6	Total	TLC6	BLC6	OLC6	Total
Mean	-2.812	0.171	0.792*	-0.616	0.012	-0.038*	-0.012	-0.013
SD	3.148	2.017	1.638	2.781	0.051	0.040	0.032	0.045
P-Value	0.004				0.041			
Writing Tasks	Argumentative Task							
Test Comparisons	Pre to Post							
Metrics	(NDW)				(SVV1)			
Groups	TLC6	BLC6	OLC6	Total	TLC6	BLC6	OLC6	Total
Mean	-18.000*	-3.700	4.200	-5.833	-2.982*	-2.870	1.406	-1.482
SD	11.926	16.111	13.105	16.289	3.760	4.444	4.290	4.532
P-Value	0.005				0.041			
Metrics	(CVV1)							
Groups	TLC6	BLC6	OLC6	Total				
Mean	-0.305*	-0.296	0.167	-0.145				
SD	0.382	0.447	0.487	0.481				
P-Value	0.036							
(*) indicates the most increase or decrease								

Figure 9. TLC6, BLC6, and OLC6 Lexical Variation Metrics: Significant Results of the ANOVA Test
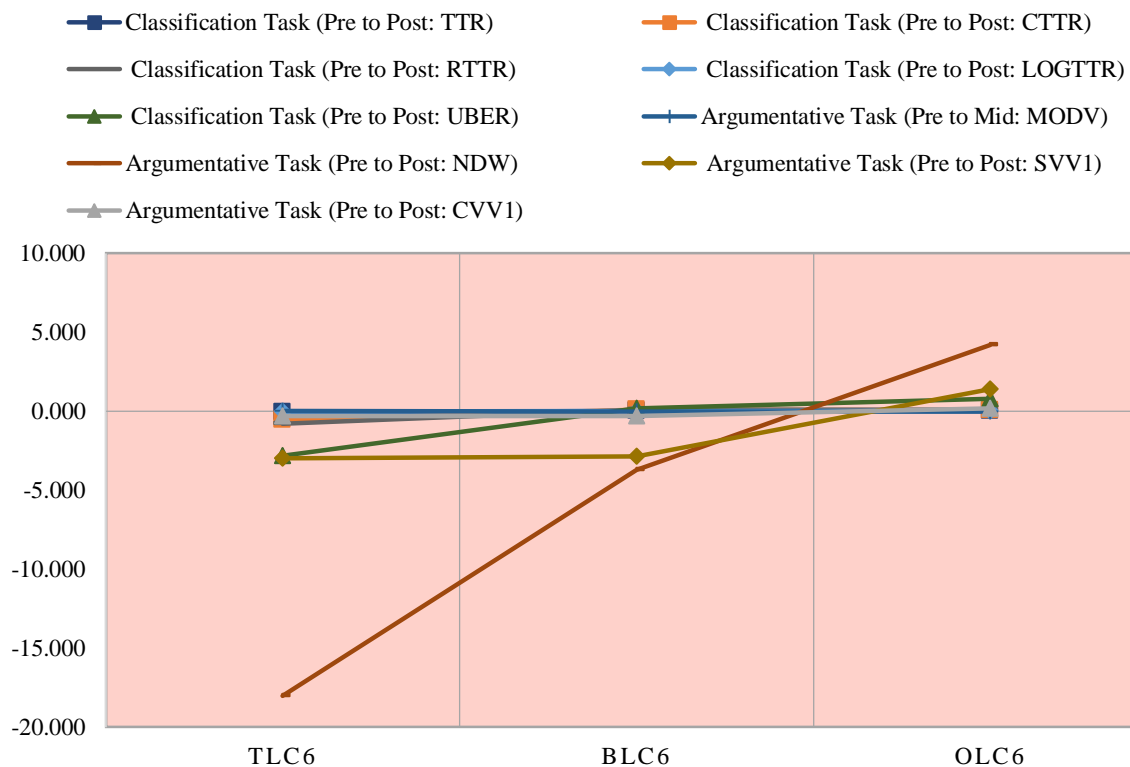
Briefly, in the classification, the OLC was the most effective in enhancing the lexical variation of the high-proficiency Saudi EFL undergraduates in the long term, whereas the TLC

was the least effective. The reason is that the OLC group improved in all five lexical variation measures (one of which the most), but the TLC group deteriorated in all of them. In the argumentative task, the TLC was the most beneficial in promoting its participants' lexical variation in the short term, while the BLC was the least beneficial. Conversely, in the long term, the OLC was the most successful in developing its learners' lexical variation in the long term, whereas the TLC was the least successful. The justification is that the OLC group increased in three measures of lexical variation, but those in TLC and BLC declined. The decreases in such measures, however, were the most for TLC.

4.1.3 Accuracy.

4.1.3.1 Comparative changes in accuracy within the low- and high-proficiency groups.

Table 20 presents the statistically significant t-test findings relating to accuracy for all six groups (TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6) in the two text types.

4.1.3.1.1 Low-proficiency groups.

At the low-proficiency level, TLC1's results revealed that there were no significant scores for the classification task in both the short term and long term. Nevertheless, in the argumentative task, three metrics (EFT/T, E/T, and E/W) out of four indicated significant scores in the short term, and one metric (E/W) showed a significant score in the long term. That is, in the short term, there was a significant development of EFT/T (M = 0.145, $p <$ .005), but there were significant positive deteriorations in both E/T and E/W (M = -0.693 and -3.740, respectively; $p <$ .002 and .000, respectively). [7] In the long term, E/W declined positively (M = -3.625, $p <$ .005). All of these findings mean that, the accuracy of TLC1 remained unaffected in the non-complex task, while it improved in the complex task.

As for BLC1, there were some significant scores in the classification task revealed by three measures (EFT, EFT/T, and E/T) in the short term. Both EFT (M =2.267, $p <$ .018) and EFT/T (M = 0.155, $p <$ .039) increased significantly, whereas E/T (M = -0.563, $p <$ .004) decreased positively. In the argumentative task, only one measure (E/W) indicated a

---

[7] In all of the measures of syntactic complexity, lexical density, lexical sophistication, lexical variation, and fluency, and in only two measures of accuracy (i.e., EFT and EFT/T), an increase indicates an increase in the constructs/sub-constructs and a decreased shows a decrease in the constructs/sub-constructs. However, in the accuracy other measures (i.e., E/W and E/T) there is a reversed result. Increasing in these measures indicates a decrease of the construct and decreasing in the measure shows an increase in the construct. Therefore, the term 'positive' was used with decrease and 'negative' with increase in order to clarify such reversed results.

significant score in deteriorating positively (M = -2.718, $p < .026$). Such results show that the BLC1's accuracy developed more in the simple task than in the difficult task. Yet, in the long term, all four measures (EFT, EFT/T, E/T, and E/W) revealed significant scores for both composition tasks. In the classification task, both EFT (M = 3.933, $p < .001$) and EFT/T (M = 0.213, $p < .006$) improved significantly, but E/T (M = -0.863, $p < .000$) and E/W (M = -4.519, $p < .001$) declined positively. In the argumentative task, EFT and EFT/T increased significantly (M = 4.333 and 0.247, respectively; $p < .000$), while E/T and E/W decreased positively (M = -0.637 and -5.383, respectively; $p < .000$). All of these findings imply that whether the task was non-complex or complex, the accuracy of BLC1 increased equally.

In the case of OLC1, there were some significant scores in the classification task as indicated by three metrics: EFT/T, E/T, and E/W, in the short term. EFT/T developed significantly (M = 0.139, $p < .034$), whereas both E/T (M = -0.609, $p < .046$) and E/W (M = -6.282, $p < .000$) deteriorated positively. In the argumentative task, only two metrics (EFT/T and E/W) showed significant scores in improving or declining positively (M = 0.139 and -4.859, respectively; $p < .013$ and .000, respectively). Likewise, in the long term, there were some significant scores in the classification task revealed by all metrics: EFT (M = 3.533, $p < .001$) and EFT/T (M = 0.181, $p < .000$) increased significantly, but E/T (M = -0.933, $p < .001$) and E/W (M = -5.511, $p < .000$) decreased positively. In the argumentative task, three metrics indicated significant scores of developments or a positive deterioration: EFT (M = 1.933, $p < 0.006$), EFT/T (M = 0.188, $p < .001$), and E/W (M = -4.713, $p < .004$). Such results mean that the OLC1's accuracy improved in the simple task more than in the difficult task.

In sum, the accuracy of the low-proficiency Saudi EFL participants in the three learning contexts was affected by the task complexity factor in one/two timescales. To demonstrate, in both the short term and long term, TLC1's accuracy remained unchanged in the non-complex task, while it increased in the complex task. The accuracy of OLC1 in the short term and long term, developed more in the simple task than in the difficult task. BLC1's accuracy improved in the non-complex task more than in the complex task in only the short term; nonetheless, no impact of level of task complexity was found on the accuracy of this group in the long term since it increased equally in both writing tasks.

4.1.3.1.2 High-proficiency groups.

At the high-proficiency level, TLC6's findings showed that there were no significant scores in both two text types in the short term and long term. This means that regardless of the task being non-complex or complex, the accuracy of TLC6 remained uninfluenced.

For BLC6, there were no significant scores in the classification task in both the short term and long term. However, in the argumentative task, two measures (E/T and E/W) declined positively (M = -0.453 and -3.048, respectively; p < .023 and .002, respectively) in the short term. In the long term, these measures furthermore decreased positively (M = -0.622 and -3.478, respectively; p < .006, and .003, respectively), whereas EFT (M = 2.400, p < .039) and EFT/T (M = 0.181, p < .022) developed significantly. All of these results imply that BLC6's accuracy remained unaffected in the simple task, but it improved in the difficult task.

With regard to OLC6, there were some significant scores for the classification task revealed by two metrics (EFT and EFT/T) in the short term. Both EFT and EFT/T increased significantly (M = 2.000 and 0.141, respectively; p < .030, and .005, respectively). Conversely, there were no significant scores for the argumentative task. Such findings indicate that the accuracy of OLC6 developed in the non-complex task, while it remained unchanged in the complex task. Nevertheless, in the long term, all four metrics showed significant scores in both classification and argumentative tasks. In the former task, EFT and EFT/T improved significantly (M = 3.300 and 0.264, respectively; p < .003 and .001, respectively), whereas E/T and E/W deteriorated positively (M = -0.956 and -5.774, respectively; p < .001 and .000, respectively). In the latter task, both EFT (M = 4.500, p < .012) and EFT/T (M = 0.261, p < .006) increased significantly, but both E/T (M = -0.881, p < .003) and E/W (M = -5.212, p < .002) declined positively. All of these results reveal that OLC6's accuracy developed equally in the simple and difficult tasks.

To conclude, both BLC6 and OLC6, in comparison to TLC6, were the only high-proficiency EFL learning environment groups whose accuracy was influenced by the task complexity factor in one/two timescales. Specifically, the accuracy of BLC6 remained uninfluenced in the non-complex task, while it improved in the complex task in both the short term and long term. OLC6's accuracy increased in the simple task, whereas it remained unaffected in the difficult task in only the short term. Yet, no effect of level of task complexity on accuracy was found for TLC6 in the short term and long term, and for OLC6 in only the long term as it remained unchanged for the former group and developed equally for the latter group in both composition tasks.

Table 20

TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6 Accuracy Metrics: Significant Results of the T-Test

Learning Contexts			Test Comparisons		Classification Task				Argumentative Task			
					(EFT)	(EFT/T)	(E/T)	(E/W)	(EFT)	(EFT/T)	(E/T)	(E/W)
Proficiency Levels	Low-Proficiency Groups	Traditional Learning Context (TLC1)	Pre-Test to Mid-Term Test	M						0.145	-0.693	-3.740
				SD						0.170	0.724	2.806
				P						.005	.002	.000
			Pre-Test to Post-Test	M								-3.625
				SD								4.192
				P								.005
		Blended Learning Context (BLC1)	Pre-Test to Mid-Term Test	M	2.267	0.155	-0.563					-2.718
				SD	3.283	0.263	0.642					4.232
				P	.018	.039	.004					.026
			Pre-Test to Post-Test	M	3.933	0.213	-0.863	-4.519	4.333	0.247	-0.637	-5.383
				SD	3.654	0.256	0.650	4.401	3.579	0.205	0.445	3.331
				P	.001	.006	.000	.001	.000	.000	.000	.000
		Online Learning Context (OLC1)	Pre-Test to Mid-Term Test	M		0.139	-0.609	-6.282		0.139		-4.859
				SD		0.229	1.079	4.786		0.189		4.085
				P		.034	.046	.000		.013		.000
			Pre-Test to Post-Test	M	3.533	0.181	-0.933	-5.511	1.933	0.188		-4.713
				SD	3.270	0.155	0.822	4.598	2.314	0.178		5.216
				P	.001	.000	.001	.000	.006	.001		.004

Learning Contexts		Test Comparisons		Classification Task				Argumentative Task			
				(EFT)	(EFT/T)	(E/T)	(E/W)	(EFT)	(EFT/T)	(E/T)	(E/W)
High-Proficiency Groups	Traditional Learning Context (TLC6)	Pre-Test to Mid-Term Test	M								
			SD								
			P								
		Pre-Test to Post-Test	M								
			SD								
			P								
	Blended Learning Context (BLC6)	Pre-Test to Mid-Term Test	M							-0.453	-3.048
			SD							0.525	2.200
			P							.023	.002
		Pre-Test to Post-Test	M					2.400	0.181	-0.622	-3.478
			SD					3.134	0.207	0.545	2.683
			P					.039	.022	.006	.003
	Online Learning Context (OLC6)	Pre-Test to Mid-Term Test	M	2.000	0.141						
			SD	2.449	0.123						
			p	.030	.005						
		Pre-Test to Post-Test	M	3.300	0.264	-0.956	-5.774	4.500	0.261	-0.881	-5.212
			SD	2.541	0.182	0.651	3.179	4.503	0.233	0.695	3.681
			P	.003	.001	.001	.000	.012	.006	.003	.002

Note. Only statistically significant metrics are included in the table.

4.1.3.2 Comparative changes in accuracy between the low- and high-proficiency groups.

4.1.3.2.1 Low-proficiency groups.

Table 21 indicates seven significant findings of the ANOVA test, showing which of the three low-proficiency groups' accuracy improved or decreased the most in the two writing tasks. As presented in the table, in the classification task, one measure (E/W) revealed significant scores in both the short term and long term, but three measures (E/T, EFT, and EFT/T) indicated such scores only in the long term. In the short term, E/W deteriorated (positively) the most for OLC1 (M = -6.282), followed by BLC1 (M = -2.775) and, lastly, TLC1 (M = -1.791). The p-value was < .031. In the long term, E/W continued to decline (positively) the most for OLC1 (M = -5.511), followed by BLC1 (M = -4.519), while TLC1 (M = 0.271) increased (negatively); the p-value was < .002. In addition, E/T decreased (positively) the most for OLC1 (M = -0.933), followed by BLC1 (M = -0.863) and, lastly, TLC1 (M = -0.107); the p-value was < .038. On the other hand, EFT developed the most for BLC1 (M = 3.933), followed by OLC1 (M = 3.533) and, finally, TLC1 (M = 0.267); the p-value was < .009. EFT/T, besides, improved significantly for BLC1 (M = 0.213), followed by OLC1 (M = 0.181), whereas it deteriorated for TLC1 (M = -0.014); the p-value was < .007. All of these results show that both the OLC (in the short term and long term) and the BLC (in only the long term) were the most effective in increasing accuracy in the classification task, but the TLC was the least effective.

In the argumentative type, two metrics (EFT and EFT/T) revealed significant scores in the long term: EFT developed the most for BLC1 (M = 4.333), followed by OLC1 (M = 1.933) and, lastly, TLC1 (M = 0.067) with a p-value of < .001. EFT/T also improved the most for BLC1 (M = 0.247), followed by OLC1 (M = 0.188) and, finally, TLC1 (M = 0.064); with a p-value of < .029. These findings indicate that in the long term, the BLC was the most beneficial, while the TLC was the least beneficial, in increasing accuracy in the argumentative task. All of these results are graphically presented in Figure 10.

Table 21

TLC1, BLC1, and OLCI Accuracy Metrics: Significant Results of the ANOVA Test

Writing Tasks	Classification Task							
Test Comparisons	Pre to Mid				Pre to Post			
Metrics	**(E/W)**				**(E/W)**			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	-1.791	-2.775	-6.282*	-3.616	0.271	-4.519	-5.511*	-3.253
SD	3.766	5.388	4.786	4.984	4.091	4.401	4.598	4.973
P-Value	0.031				0.002			
Test Comparisons	Pre to Post							
Metrics	**(E/T)**				**(EFT)**			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	-0.107	-0.863	-0.933*	-0.634	0.267	3.933*	3.533	2.578
SD	1.253	0.650	0.822	0.996	3.195	3.654	3.270	3.696
P-Value	0.038				0.009			
Writing Tasks	Classification Task				Argumentative Task			
Test Comparisons	Pre to Post				Pre to Post			
Metrics	**(EFT/T)**				**(EFT)**			
Groups	TLC1	BLC1	OLC1	Total	TLC1	BLC1	OLC1	Total
Mean	-0.014	0.213*	0.181	0.127	0.067	4.333*	1.933	2.111
SD	0.178	0.256	0.155	0.221	2.404	3.579	2.314	3.277
P-Value	0.007				0.001			
Writing Tasks	Argumentative Task							
Test Comparisons	Pre to Post							
Metrics	**(EFT/T)**							
Groups	TLC1	BLC1	OLC1	Total				
Mean	0.064	0.247*	0.188	0.166				
SD	0.167	0.205	0.178	0.196				
P-Value	0.029							
(*) indicates the most increase or decrease								

Figure 10. TLC1, BLC1, and OLC1 Accuracy Metrics: Significant Results of the ANOVA Test

Legend:
- Classification Task (Pre to Mid: E/W)
- Classification Task (Pre to Post: E/W)
- Classification Task (Pre to Post: E/T)
- Classification Task (Pre to Post: EFT)
- Classification Task (Pre to Post: EFT/T)
- Argumentative Task (Pre to Post: EFT)
- Argumentative Task (Pre to Post: EFT/T)
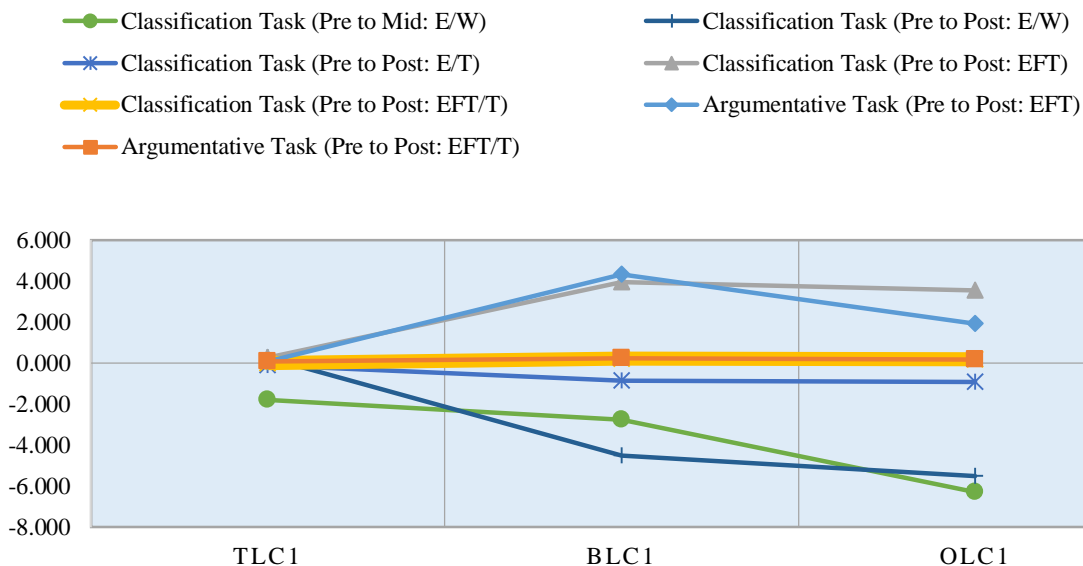
All things considered, in enhancing the accuracy of the low-proficiency Saudi EFL participants in the classification task, the OLC (in the short term and long term) and the BLC (in only the long term) were the most successful, whereas the TLC was the least successful. In

promoting the accuracy of such learners in the long term in the argumentative task, the BLC was found to be the most effective, but the TLC was the least effective.

4.1.3.2.2 High-proficiency groups.

Table 22 shows five statistically significant findings of the ANOVA test, revealing which of the three high-proficiency groups' accuracy developed or declined the most in the two text types. As indicated in the table, in the classification task, three measures (EFT, E/T, and E/W) showed significant scores in the long term: EFT improved the most for OLC6, then BLC6 and, lastly, TLC6 (M = 3.300, 1.900, and 0.041, respectively; $p < .034$). E/T decreased (positively) the most for OLC6 (M = -0.956), followed by BLC6 (M = -0.285), and then TLC6 (M = -0.119); the p-value was < .012. E/W deteriorated (positively) the most for OLC6 (M = -5.774), followed by BLC6 (M = -1.651), while it increased (negatively) for TLC6 (M = 0.600); the p-value was < .000. In the argumentative task, two measures (EFT and E/W) revealed significant scores in the long term: EFT developed the most for OLC6 (M = 4.500), then BLC6 (M = 2.400) and, finally, TLC6 (M = 0.019); the p-value was < .014. E/W declined (positively) the most for OLC6 (M = -5.212), followed by BLC6 (M = -3.478) and, lastly, TLC6 (M = -0.200); the p-value was < .005. These results mean that the OLC was the most beneficial, whereas the TLC was the least beneficial in improving accuracy in both the classification and argumentative tasks in the long term. The aforementioned findings are graphically shown in Figure 11.

Table 22

TLC6, BLC6, and OLC6 Accuracy Metrics: Significant Results of the ANOVA Test

Writing Tasks	Classification Task							
Test Comparisons	Pre to Post							
Metrics	(EFT)				(E/T)			
Groups	TLC6	BLC6	OLC6	Total	TLC6	BLC6	OLC6	Total
Mean	0.041	1.900	3.300*	1.747	-0.119	-0.285	-0.956*	-0.453
SD	0.230	3.784	2.541	2.882	0.705	0.467	0.651	0.699
P-Values	0.034				0.012			
Writing Tasks	Classification Task				Argumentative Task			
Test Comparisons	Pre to Post				Pre to Post			
Metrics	(E/W)				(EFT)			
Groups	TLC6	BLC6	OLC6	Total	TLC6	BLC6	OLC6	Total
Mean	0.600	-1.651	-5.774*	-2.275	0.019	2.400	4.500*	2.306
SD	3.34	2.781	3.179	4.025	0.222	3.134	4.503	3.581
P-Values	.000				0.014			
Writing Tasks	Argumentative Task							
Test Comparisons	Pre to Post							
Metrics	(E/W)							
Groups	TLC6	BLC6	OLC6	Total				
Mean	-0.200	-3.478	-5.212*	-2.963				
SD	2.936	2.683	3.681	3.685				
P-Values	0.005							
(*) indicates the most increase or decrease								

Figure 11. TLC6, BLC6, and OLC6 Accuracy Metrics: Significant Results of the ANOVA Test
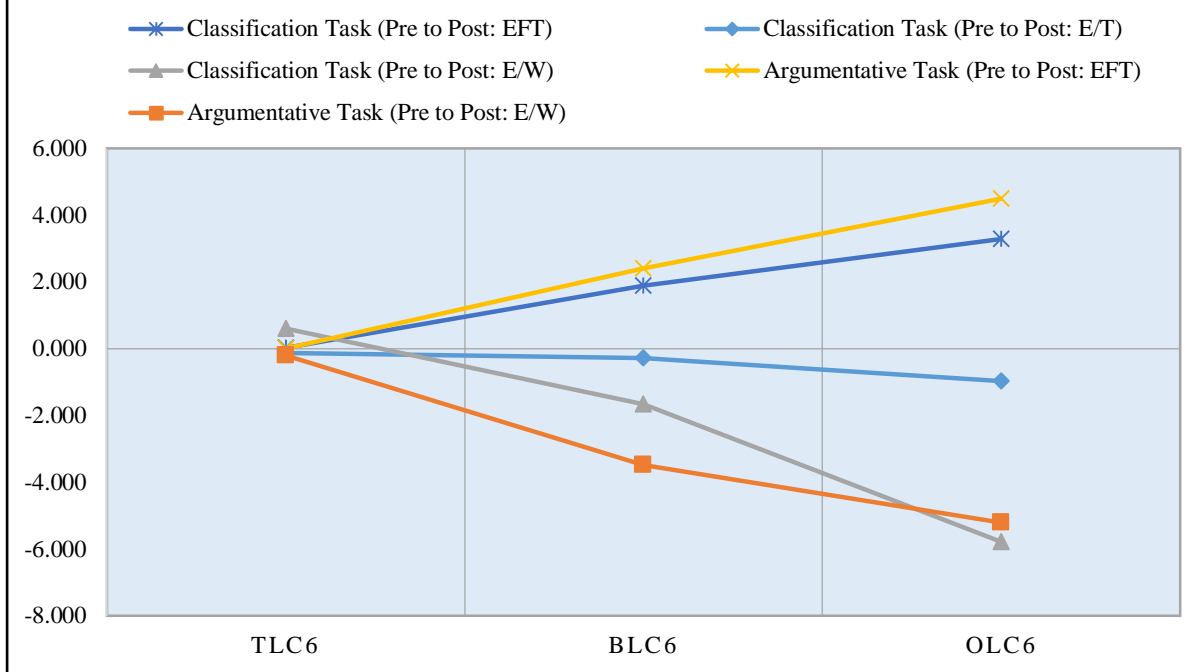
In brief, in enhancing the accuracy of the high-proficiency Saudi EFL undergraduates in the classification and argumentative task, in the long term, the OLC was the most successful, whereas the TLC was the least successful.

4.1.4 Fluency.

4.1.4.1 Comparative changes in fluency within the low- and high-proficiency groups.

Table 23 displays the statistically significant t-test results regarding the fluency for all six groups (TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6) in the two composition tasks.

4.1.4.1.1 Low-proficiency groups.

At the low-proficiency level, TLC1's findings indicated that out of the four fluency metrics (W/T, W/C, W/EFT, and W), only W/C in the classification task and W in the argumentative task revealed statistically significant scores of increasing or decreasing (M = 2.735 and -42.933, respectively; $p < .049$ and .014, respectively) in the short term. These results mean that the fluency of TLC1 developed in the non-complex task, but it deteriorated in the complex task. In contrast, in the long term, there were no statistically significant scores for both writing tasks, showing that whether the task was simple or difficult, TLC1's fluency remained unchanged.

For BLC1, there were no statistically significant scores for the classification task in the short term, while in the argumentative task only one significant score was indicated by one

measure: W/T (M = 1.097, $p < .006$). This implies that the fluency of BLC1 remained uninfluenced in the non-complex task, whereas it improved in the complex task. In the long term, only one measure (W/EFT: M = -4.271, $p < .015$) in the classification task, and one other measure (W/T: M = 1.642, $p < .005$) in the argumentative task revealed statistically significant scores of declining or increasing. Such findings show that BLC1's fluency decreased in the simple task, but it developed in the difficult task.

In the case of OLC1, there were some statistically significant scores in the classification task, as indicated by two metrics (W/T and W/C) in the short term, and by one metric (W/C) in the long term. All of these metrics improved significantly (M = 2.487, 4.263, and 2.483, respectively; $p < .003, .006,$ and $.033$, respectively). Nonetheless, there were no statistically significant scores for the argumentative task in both the short term and long term. All of the aforementioned results mean that the fluency of OLC1 increased in the non-complex task, while it remained unaffected in the complex task.

On the whole, the fluency of the low-proficiency Saudi EFL undergraduates in the three learning contexts was impacted differently by the task complexity factor in one/two timescales. To clarify, BLC6's fluency remained unchanged in the simple task, but it developed in the difficult task in the short term. In the long term, this construct deteriorated for BLC6 in the non-complex task, whereas it improved in the complex task. The fluency of OLC1 increased in the simple task, but it remained uninfluenced in the difficult task in both the short term and long term. Finally, TLC1's fluency developed in the non-complex task, while it declined in the complex task. However, no influence of level of task complexity on such construct was found for TLC1 in the long term because it remained unaffected in both text types.

4.1.4.1.2 High-proficiency groups.

At the high-proficiency level, TLC6's findings in the classification and argumentative tasks revealed no statistically significant scores in the short term, which shows that regardless of the task being simple or difficult, TLC6's fluency remained unchanged. Nevertheless, in the long term, although there were no statistically significant scores for the classification task, in the argumentative task, two measures either improved significantly, as in the case of W/T (M = 1.547, $p < .041$), or decreased significantly, as in the case of W (M = -38.700, $p < .004$). Such results imply that the fluency of TLC6 remained unchanged in the non-complex task, whereas it equally increased and deteriorated in the complex task.

As for BLC6, there were no statistically significant scores in the classification and argumentative tasks in the short term and long term. This means that BLC6's fluency remained uninfluenced in the simple and difficult tasks.

OLC6's findings for the classification task indicated that one metric (W/T) developed significantly (M = 2.020, p < .009) in the short term, but in the argumentative task, there were no statistically significant scores. This reveals that the fluency of OLC6 improved in the non-complex task, while it remained unaffected in the complex task. In contrast, in the long term, there were no statistically significant scores for the classification task, whereas for the argumentative task, one metric (W/C) increased significantly (M = 3.982, p < .036). Such results show that OLC6's fluency remained unchanged in the simple task, but it developed in the difficult task.

In short, both TLC6 and OLC6, in comparison to BLC6, were the only high-proficiency EFL learning environment groups in which fluency was affected by the task complexity factor in one/two timescales. To elucidate, the fluency of TLC6 in the long term remained uninfluenced in the non-complex task, while it equally improved and declined in the complex task. OLC6's fluency increased in simple task, whereas it remained unaffected in the difficult task in the short term. Again, in the long term, this construct remained unchanged for OLC6 in the non-complex task, but it developed in the complex task. Lastly, there was no impact of level of task complexity on the fluency of TLC6 in the short term, and BLC6 in the short term and long term since it remained uninfluenced in both composition tasks.

Table 23

TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6 Fluency Metrics: Significant Results of the T-Test

Learning Contexts			Test Comparisons		Classification Task				Argumentative Task			
					(W/T)	(W/C)	(W/EFT)	(W)	(W/T)	(W/C)	(W/EFT)	(W)
Proficiency Levels	Low-Proficiency Groups	Traditional Learning Context (TLC1)	Pre-Test to Mid-Term Test	M		2.735						-42.933
				SD		4.917						59.464
				P		.049						.014
			Pre-Test to Post-Test	M								
				SD								
				P								
		Blended Learning Context (BLC1)	Pre-Test to Mid-Term Test	M					1.097			
				SD					1.297			
				P					.006			
			Pre-Test to Post-Test	M			-4.271		1.642			
				SD			5.494		1.931			
				P			.015		.005			
		Online Learning Context (OLC1)	Pre-Test to Mid-Term Test	M	2.487	4.263						
				SD	2.744	5.074						
				P	.003	.006						
			Pre-Test to Post-Test	M		2.483						
				SD		4.070						
				P		.033						

	Learning Contexts	Test Comparisons		Classification Task				Argumentative Task			
				(W/T)	(W/C)	(W/EFT)	(W)	(W/T)	(W/C)	(W/EFT)	(W)
High-Proficiency Groups	Traditional Learning Context (TLC6)	Pre-Test to Mid-Term Test	M								
			SD								
			P								
		Pre-Test to Post-Test	M					1.547			-38.700
			SD					2.056			31.899
			P					.041			.004
	Blended Learning Context (BLC6)	Pre-Test to Mid-Term Test	M								
			SD								
			P								
		Pre-Test to Post-Test	M								
			SD								
			P								
	Online Learning Context (OLC6)	Pre-Test to Mid-Term Test	M	2.020							
			SD	1.933							
			P	.009							
		Pre-Test to Post-Test	M						3.982		
			SD						5.128		
			P						.036		

Note. Only statistically significant metrics are included in the table.

191

4.1.4.2 Comparative changes in fluency between the low- and high-proficiency groups.

4.1.4.2.1 Low-proficiency groups.

Table 24 presents the single statistically significant result of the ANOVA test for the argumentative task, indicating which of the three low-proficiency groups' fluency improved or decreased the most. As presented in the table, in the long term, the W measure revealed that BLC1 (M = 27.600) increased, while TLC1 (M = -38.067) deteriorated the most, followed by OLC1 (M = -6.467). The p-value was < .012. This means that the BLC was the most effective, whereas the TLC was the least effective, in developing fluency in the argumentative task in the long term. Conversely, there were no statistically significant differences between the three groups in fluency scores in the classification task in both timescales, showing that the three learning environments had the same level of effectiveness/ineffectiveness in improving such construct. Figure 12 graphically presents the above-mentioned finding.

Table 24

TLC1, BLC1, and OLC1 Fluency Metrics: Significant Results of the ANOVA Test

Writing Tasks	Argumentative Task			
Test Comparisons	Pre to Post			
Metrics	(W)			
Groups	TLC1	BLC1	OLC1	Total
Mean	-38.067*	27.600	-6.467	-5.644
SD	70.804	52.182	45.169	62.017
P-Values	.012			

(*) indicates the most increase or decrease

Note. There were no statistically significant differences between the three low-proficiency learning context groups in fluency in the classification task.

Figure 12. TLC1, BLC1, and OLC1 Fluency Metrics: Significant Results of the ANOVA Test
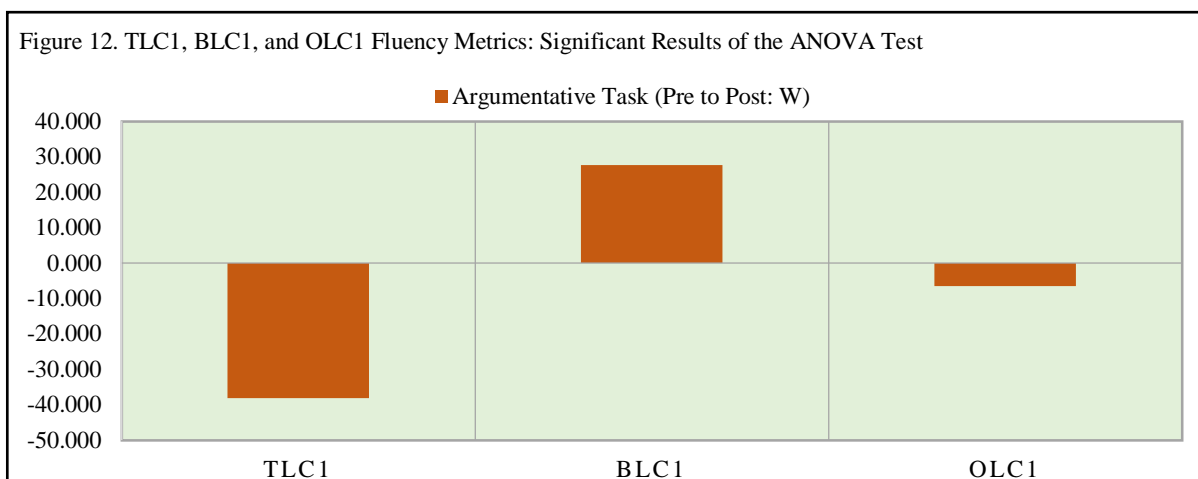
To sum up, the BLC appeared to be the most successful in promoting the fluency of the low-proficiency Saudi EFL participants in the argumentative task in the long term, but the TLC appeared to be the least successful. Yet, the TLC, BLC, and OLC had the same level of

success/unsuccess in enhancing the fluency of their learners in the classification task in the short term and long term.

4.1.4.2.2 High-proficiency groups.

The ANOVA test did not indicate any statistically significant differences between the three high-proficiency groups (TLC6, BLC6, and OLC6) across the two writing tasks, in the short term and long term, revealing that the level of effectiveness/ineffectiveness of the three learning contexts in increasing such construct was the same.

Chapter Five

Discussion

5.0 Introduction

This section presents a detailed discussion of the results in relation to the two research questions. It also compares them to other studies that addressed the task complexity factor and the impact of different learning contexts on the CAF writing of EFL learners.

5.1 Research Question One: 'How are the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates in the three learning environments: TLC, BLC, and OLC, affected longitudinally across two writing tasks (classification and argumentative) that differed in their level of complexity? And when?'

The first research question attempted to specify how the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates in the three learning environments (TLC, BLC, OLC) were influenced longitudinally across two text types (classification and argumentative) that differed in their level of complexity. According to Ellis (2003), "If task-based tests are to be used to infer the abilities of test-takers to predict performance and to generalize from context to context, it will be necessary to understand how the choice of task influences the way the testee performs" (p. 288). In order to achieve such a goal, two competing models that attempt to explain how L2 learners' attention is deployed during the performance of these tasks and how the manipulation of cognitive elements along the resource-directing dimension can affect L2 production were incorporated. The first model is the Limited Attentional Capacity Model (also known as the Trade-off Hypothesis) proposed by Skehan (1996a, 1996b, 1998, 2009), Skehan and Foster (1997, 1999; 2001, 2005, 2012), and Foster and Skehan (1996, 1999). In this model, the researchers predicted better performance in the simple task than in the more difficult task. In the latter task, participants will experience decreases in their syntactic complexity, lexical complexity, accuracy, and fluency. In contrast, the second model is the Multiple Attentional Resources Model (also known as the Cognition Hypothesis) proposed by Robinson (2001a, 2001b, 2003, 2005, 2007, 2011a, 2011b, 2015) and Robinson and Gilabert (2007). In their model, when increasing task complexity by resource-directing, learners will perform better than in the simple task in the sense that they would increase in accuracy and complexity simultaneously, while their fluency would decrease.

As important as broadening the scope of understanding when, how, and why the task complexity factor impacts the CAF writing constructs of EFL learners is in general, and Saudi learners in particular, the data analysis was undertaken in a way that is different from the other studies that investigated the topic. In these studies, the EFL participants' CAF productions were compared across the simple and difficult tasks, whereas the current study looks at the constructs for each composition task before comparing the two composition tasks. In other words, a statistical t-test was used to compare each of the TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6 groups' CAF scores in each writing task in the pre-test, first on the mid-term test (short term) and second, on the post-test (long term). The CAF results of these comparisons within each group were categorized with regard to their kinds of effect (i.e., unchanged, increasing, decreasing, or both) and the number of measures indicating such an effect, and without regard to the types of these measures. For example, when a similarity of influence on syntactic complexity was specified between the argumentative (complex) and the classification (non-complex) tasks for BLC1 in the long term, it was done on the basis that the construct decreased significantly across the same number, but not type, of measures in both texts. The construct's deterioration was across C/S, C/T, DC/T, and CT/T in the argumentative task and across MLS, VP/T, C/T, and CN/T in the classification task. The number of measures was emphasized as it plays a crucial role in interpreting the impact, that is, whether there was equal development and deterioration, more development or deterioration, more development than deterioration, or vice versa. To give one example, the lexical variation of TLC1 in the short term, increased across TTR, VV1, AdjV, and ModV, and decreased across NDW in the argumentative task, while it only declined across VV1 and AdjV in the classification task. This means that the participants' lexical variation increased more than it decreased in the former task, whereas it only declined in the latter task. Finally, the current research showed that the CAF constructs were affected similarly and differently in the two writing tasks, and therefore, were partially in line with Skehan and Foster's and Robinson's models.

5.1.1 Low-proficiency level participants: CAF effects in complex and non-complex tasks.

At the low-proficiency level, it has been seen that for TLC1, in terms of lexical density and lexical sophistication in the short term, no significant differences were found between the two text types, which neither supports the predictions of Skehan and Foster nor those of Robinson. However, an influence for task complexity on syntactic complexity, lexical

variation, accuracy, and fluency was found. Inconsistent with Skehan and Foster's and Robinson's models, syntactic complexity increased (CN/C) and decreased (T/S) equally in the simple task, but it remained uninfluenced in the difficult task. Moreover, lexical variation declined significantly (VV1 and AdjV) in the non-complex task, while it developed significantly (TTR, VV1, AdjV, and ModV) more than it deteriorated (NDW) in the complex task. In line with Robinson's model, accuracy remained unaffected in the simple task, whereas it improved significantly (EFT/T, E/T, and E/W) in the difficult task. Commensurate with this model as well as that of Skehan and Foster, fluency increased significantly (W/C) in the non-complex task, but it decreased significantly (W) in the complex task. All of these findings indicate that increasing task complexity in the short term leads the TLC low-proficiency Saudi EFL university students to pay no attention to content/meaning, less attention to one aspect of linguistic form, and more attention to two other aspects in that their written output becomes less fluent, remains unchanged in syntactic complexity, and becomes more accurate and (though there was a significant decrease in one measure) more lexically varied. It does not impact the lexical density and lexical sophistication of the output in a way that is different from that of the simple task. Similarly, there were no significant differences between the two composition tasks with regard to syntactic complexity, lexical density, and fluency in the long term. This does not support Skehan and Foster's and Robinson's predictions. Nevertheless, an effect for task complexity on lexical sophistication, lexical variation, and accuracy was found. Lexical sophistication declined less significantly in the non-complex task (LS2) than in the complex task (LS1 and LS2). This does not corroborate the Limited Attentional Capacity Model. Lexical variation deteriorated significantly (LV) in the simple task, while it developed significantly (VV1 and ModV) more than it decreased significantly (NDW) in the difficult task. As such, this influence was beyond Robinson's Cognition Hypothesis and Skehan and Foster's model. Consistent with Robinson's model, accuracy remained uninfluenced in the non-complex task, whereas it improved significantly (E/W) in the complex task. All of these results imply that elevating task complexity results in these TLC Saudi EFL undergraduates prioritizing two aspects of linguistic form and deprioritizing one other aspect in that their output becomes more accurate and (even though there was a significant decrease in one measure) more lexically varied, and less lexically sophisticated in the long term. Yet, it does not impact the syntactic complexity, lexical density, and fluency of their output in a way that varies from that of the simple task.

For BLC1, there were significant differences in the short term between the two writing tasks with respect to syntactic complexity, lexical density, lexical sophistication, lexical

variation, accuracy, and fluency; which are not in line with Skehan and Foster's and Robinson's predictions. In spite of the fact that syntactic complexity (C/T) declined significantly, and lexical density (LD) and lexical variation (NDWZ, NDWESZ, TTR, MSTTR, LogTTR, Uber, and VV1) increased significantly in the non-complex task, they all remained unaffected in the complex task. Lexical sophistication deteriorated significantly in the simple task (LS1, LS2, VS2, and CVS1), more so than in the difficult task (LS1 and LS2). Accuracy increased significantly in the non-complex task (EFT, EFT/T, and E/T), more so than in the complex task (E/W). Fluency remained unchanged in the simple task, but it developed significantly (W/T) in the difficult task. All these findings reveal that altering task complexity in the short term for the BLC low-proficiency Saudi EFL students contributes to their focusing little on three aspects of linguistic form and focusing more on content in that their written output remains uninfluenced in syntactic complexity, lexical density, and lexical variation, and becomes more fluent. It also contributes to their prioritization of one aspect of linguistic form and deprioritization of one other aspect in that their output becomes accurate and less lexically sophisticated (still not to the same level as in the non-complex task). Conversely, in the long term, there were no significant differences between the two text types in terms of syntactic complexity, lexical density, lexical sophistication, and accuracy, and despite identifying an effect for task complexity on lexical variation and fluency, only that on lexical variation that was commensurate with Robinson's Cognition Hypothesis. Lexical variation remained unaffected and fluency decreased significantly (W/EFT) in the simple task, while in the difficult task, both of which improved significantly (NDW and W/T, respectively). All of these results show that increasing task complexity in the long term for these BLC undergraduates causes them to pay more attention to one aspect of linguistic form and the content in that their output becomes more lexically varied and fluent; nonetheless, it does not influence their syntactic complexity, lexical density, lexical sophistication, and accuracy in a way that is distinctive from that in the non-complex task.

In the case of OLC1, there were no significant differences between the two composition tasks with regard to lexical density, which does not lend any support to Skehan and Foster's and Robinson's models. These models were also not supported in terms of the impact of task complexity that was found on syntactic complexity, lexical sophistication, lexical variation, accuracy, and fluency in the short term. In the simple task, syntactic complexity increased significantly (CP/C and CN/C) more than it declined significantly (DC/C); lexical variation developed significantly (NDWERZ, NDWESZ, MSTTR) more than it deteriorated significantly (AdjV and ModV); lexical sophistication decreased significantly (LS1 and LS2);

and fluency improved significantly (W/T and W/C), whereas in the difficult task, all of these remained unchanged. Accuracy significantly increased in the non-complex task (EFT/T, E/T, and E/W) and was lower in the complex task (EFT/T and E/W). All of these findings indicate that elevating task complexity for the OLC low-proficiency students leads them to less prioritize content and three aspects of linguistic form and prioritize one other aspect in that their written output remains uninfluenced in fluency, syntactic complexity, lexical sophistication, and lexical variation, and becomes accurate (but not to the same level as in the simple task) in the short term. However, this does not affect their output with regard to lexical density in a way that is different from that of the non-complex task. Similarly, in the long term, there were no significant differences in syntactic complexity and lexical density between the two writing tasks. This is not consistent with the two models discussed previously. Nevertheless, an influence for task complexity on lexical sophistication, lexical variation, accuracy, and fluency was found. In line with Skehan and Foster's predictions, lexical sophistication developed significantly (VS2 and CVS1) and so did lexical variation (NDWERZ and VV2) in the simple task, but the former sub-construct (LS1, VS1, VS2, and CVS1) and the latter sub-construct (SVV1 and CVV1) declined significantly in the difficult task. Accuracy significantly improved, which does not corroborate the two models, in the non-complex task (EFT, EFT/T, E/T, and E/W) and was more than that in the complex task (EFT, EFT/T, and E/W). Fluency increased significantly (W/C) in the simple task, while it remained unaffected in the difficult task. This is not commensurate with the two models. All of these results imply that altering task complexity for these OLC undergraduates results in their focusing on one aspect of linguistic form, not focusing on two other aspects, and focusing little on content in that their output becomes accurate (still not to the same level as in the non-complex task) and less lexically sophisticated and varied, and remains unchanged in fluency in the long term. Yet, it does not impact syntactic complexity and lexical density in a way that varies from that of the simple task.

5.1.2 High-proficiency level participants: CAF effects in complex and non-complex tasks.

At the high-proficiency level, for TLC6, there was no significant differences between the two text types in syntactic complexity, accuracy, and fluency in the short term, which does not support the predictions of Skehan and Foster or those of Robinson. Nonetheless, an effect for task complexity on lexical density, lexical sophistication, and lexical variation was found. Consistent with the predictions of the Limited Attentional Capacity Model, lexical density and

lexical sophistication remained uninfluenced in the non-complex task, whereas they deteriorated significantly (LD, LS1, and LS2, respectively) in the complex task. Lexical variation, which is not in line with either model, developed significantly (MSTTR and VV2) in the simple task, but it remained unaffected in the difficult task. All of these findings reveal that increasing task complexity in the short term for the TLC high-proficiency students contributes to their paying no attention to two aspects of linguistic form and less attention to one other aspect in that their written output becomes less lexically sophisticated and dense, and remained unchanged in lexical variation. However, this does not influence the syntactic complexity, accuracy, and fluency of the output in a way that is distinctive from that of the non-complex task. In contrast, in the long term, there were no significant differences in terms of lexical density and accuracy between the two composition tasks. This was not commensurate with either model. Nevertheless, an impact for task complexity on syntactic complexity, lexical sophistication, lexical variation, and fluency was found. In support of Robinson's Cognition Hypothesis, syntactic complexity remained uninfluenced in the simple task, while it improved significantly (C/S and C/T) in the difficult task. Consistent with the Limited Attentional Capacity Model, lexical sophistication remained unaffected and lexical variation decreased significantly (CTTR, RTTR, LogTTR, Uber, LV, and VV1) more than it increased (VV2) in the non-complex task, whereas both of which declined significantly (former sub-construct: LS1 and LS2, and latter sub-construct: NDW, CTTR, RTTR, SVV1, and CVV1) in the complex task. Fluency, which was not commensurate with either model, remained unchanged in the simple task, but it developed (W/T) and deteriorated (W) equally in the difficult task. All of these results show that elevating task complexity in the long term causes these TLC undergraduates to prioritize more, one aspect of linguistic form, deprioritize two others, and both prioritize and deprioritize content in that their output becomes more syntactically complex, less lexically sophisticated and varied, and equally more and less fluent. Yet, it does not affect the lexical density and accuracy of their output in a way that is different from that of the non-complex task.

With regard to BLC6, there were no significant differences in lexical density and fluency between the two writing tasks in the short term, which does not lend any support to Skehan and Foster's and Robinson's models. Nonetheless, an influence for task complexity on syntactic complexity, lexical sophistication, lexical variation, and accuracy was found. Inconsistent with the Cognition Hypothesis and the Limited Attentional Capacity Model, syntactic complexity decreased significantly (CP/T and CP/C) in the simple task, while it remained uninfluenced in the difficult task. Furthermore, lexical sophistication declined

significantly less in the non-complex task (LS1) than in the complex task (LS1 and LS2). In line with Skehan and Foster's model, lexical variation remained unaffected in the simple task, whereas it deteriorated significantly (NDWERZ, CTTR, RTTR, Uber, NV, AdjV, and ModV) in the difficult task. Commensurate with Robinson's Cognition Hypothesis, accuracy remained unchanged in the non-complex task, but it improved significantly (E/T and E/W) in the complex task. All of these findings indicate that altering task complexity for the BLC high-proficiency students leads to their having less, more, or no focus on four aspects of linguistic form in that their written output remains uninfluenced in syntactic complexity and becomes less lexically sophisticated and varied and more accurate in the short term. However, it does not impact the lexical density and fluency of the output in a way that varies from that of the simple task. Likewise, in the long term, there were no significant differences between the two text types in terms of syntactic complexity, lexical density, or fluency. This did not support the two models. Nevertheless, an effect for task complexity on lexical sophistication, lexical variation, and accuracy was found. Lexical sophistication, which does not corroborate Skehan and Foster's model, decreased significantly less in the non-complex task (LS2) than in the complex task (LS1 and LS2). Consistent with Skehan and Foster's predictions, lexical variation increased significantly (NDWERZ and VV2) in the simple task, while it declined significantly (VV1) in the difficult task. In line with Robinson's predictions, accuracy remained unaffected in the non-complex task, whereas it developed significantly (EFT, EFT/T, E/T, and E/W) in the complex task. All of these results imply that increasing task complexity for these BLC undergraduates results in their paying more or no attention to three aspects of linguistic form in that their output becomes more accurate and less lexically sophisticated and varied in the long term. It does not influence the syntactic complexity, lexical density, or fluency of the output in a way that is distinctive from that in the simple task.

In OLC6, there were no significant differences in lexical density between the two composition tasks in the short term, which is not commensurate with the models of Robinson and Skehan and Foster. Yet, an impact for task complexity on syntactic complexity, lexical sophistication, lexical variation, accuracy, and fluency was found. Not in support of the two models, syntactic complexity deteriorated significantly more in the non-complex task (CP/C and CN/C) than in the complex task (CN/C). Besides, lexical sophistication decreased significantly (LS1) and both accuracy (EFT and EFT/T) and fluency (W/T) improved significantly in the simple task, but in the difficult task, all of these remained unchanged. In line with the predictions of Skehan and Foster, lexical variation increased significantly

(NDWZ, TTR, CTTR, RTTR, LogTTR, and LV) in the non-complex task, while it declined significantly (NV) in the complex task. All of these findings reveal that elevating task complexity in the short term for the OLC high-proficiency students contributes to lessening their prioritization of content and two aspects of linguistic form, and no prioritization of two other aspects in that their written output remains uninfluenced in fluency, lexical sophistication, and accuracy and becomes less syntactically complex (but not to the same level as in the simple task) and less lexically varied. This does not affect the lexical density of their output in a way that is different from that of the non-complex task. Conversely, in the long term, there were no significant differences between the two writing tasks with regard to lexical density, lexical sophistication, and accuracy. This is not commensurate with the two models. Nonetheless, an influence for task complexity was found on syntactic complexity, lexical variation, and fluency. In support of the Limited Attentional Capacity Model, syntactic complexity remained unaffected and lexical variation developed significantly (NDWZ and VV2) in the simple task, whereas both of these deteriorated significantly (CN/T and NV, respectively) in the difficult task. Inconsistent with the two models, fluency remained unchanged in the non-complex task, but it improved significantly (W/C) in the complex task. All of these results show that altering task complexity in the long term for these OLC undergraduates causes them not to focus on two aspects of linguistic form and to focus more on content in that their output becomes less syntactically complex and lexically varied and more fluent. It does not impact the lexical density, lexical sophistication, and accuracy of the output in a way that varies from that of the simple task.

5.1.3 Summary of the findings and comparison to other task complexity studies.

To sum up, all of the previously mentioned short- and long-term findings show that the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL university students were affected to varying degrees in the classification and argumentative tasks. With regard to the similarities, which were inconsistent with the predictions of the Cognition Hypothesis and the Limited Attentional Capacity Model, there were no short-term significant differences between the complex and non-complex tasks in syntactic complexity (e.g., TLC6), lexical density (e.g., TLC1, OLC1, BLC6, and OLC6), lexical sophistication (e.g., TLC1), accuracy (e.g., TLC6), and fluency (e.g., TLC6 and BLC6). Again, there were no long-term significant differences between the two tasks in syntactic complexity (e.g., TLC1, BLC1, OLC1, and BLC6), lexical density (e.g., TLC1, BLC1, OLC1, TLC6, BLC6, and OLC6), lexical sophistication (e.g., BLC1 and OLC6), accuracy (e.g., BLC1, TLC6, and OLC6), and

fluency (e.g., TLC1 and BLC6). All of such were partially in line with the findings of some previous studies (e.g., Abdali & Fatemipour, 2014; Abdollahzadeh & Kashani, 2011; Jong, 2009; Kuiken & Vedder, 2007a, 2008, 2011; Kuiken et al., 2005; Li, 2000; Mohammadzadeh Mohammadabadi et al., 2013; Nikou & Eskandarsefat, 2012; Pourdana & Behbahani, 2011; Pourdana, Behbahani, & Safdari, 2011; Rahimpour & Hosseini, 2010; Salimi & Dadashpour, 2012; Sasayama, 2011). For instance, Kuiken and Vedder (2007a, 2008, 2011) and Kuiken et al. (2005) found no significant differences between the here-and-now task and the there-and-then task in syntactic complexity and lexical variation. Mohammadzadeh Mohammadabadi et al. (2013) observed a similarity in fluency and syntactic complexity between +/- planning here-and-now and +/- planning there-and-then tasks. Nikou and Eskandarsefat (2012) maintained that all simple and difficult decision-making and information-gap tasks had the same level of accuracy and syntactic complexity. Jong (2009) contended that productions of picture-description and story-creation tasks were similar in lexical density and accuracy, and that the texts of story-creation and opinion-expression were similar in fluency.

Kuiken et al. (2005) pointed out that, sometimes, CAF constructs might not indicate any significant differences between text types because their levels of complexity were not that different from one another. Such a claim cannot be made since the aforementioned research, which either incorporated various tasks; however, did not elaborate on their levels of complexity, or manipulated the complexity of the same task, obtained different CAF results. There might be other reasons why there were no differences between the composition tasks in some CAF constructs. For example, in the case of the current research, the similarity of the influence on CAF constructs in the classification and argumentative tasks could be a consequence of how the writing tests (pre-test, mid-term test, and post-test) were carried out.

In each writing test, the participants of all groups were required to write the two text types in one session. In many studies (e.g., Abdali & Fatemipour, 2014; Frear & Bitchener, 2015; Jiaxin, 2015; Kuiken & Vedder, 2007a, 2007b, 2008, 2011, Kuiken et al., 2005), researchers assigned a time interval between performing the different tasks (e.g., Salimi & Dadashpour, 2012); required half of the participants to do the non-complex task and the other half to do the complex task (e.g., Kuiken et al., 2005); or assigned each composition task to one group (e.g., Farahani & Meraji, 2011) in order to eliminate any factor that would jeopardize the findings, that is, leading participants to obtain similar CAF findings (cf. Nikou & Eskandarsefat, 2012). The application of such time intervals in the present study was not feasible because all of the recruited traditional, blended, and online learning environment

participants were full-time students who had to fulfill the various demands of the writing course as well as the other mandatory courses.

In addition, whether the similarity of impact on each CAF construct was determined based on its increase, decrease, or remained uninfluenced in both writing tasks, it did not appear in a systematic way through the same types and number of metrics in both the short and long terms for all the groups within or across proficiency levels. In only the long term, for instance, did lexical sophistication decline significantly through the same number and types of metrics for OLC6 in the argumentative and classification tasks (LS1 and LS2), and through the same number; nevertheless, different types of metrics, for BLC1 in the argumentative (LS1) and classification (LS2) tasks. Finally, what substantiated that the level of complexity differed between the text types used is the fact that each group had some CAF constructs that were affected differently. This was partially commensurate with the argument made by Yule (1997), Ellis (2008), and Rahimpour (2007) that discourse demands of composition tasks such as narrative, opinion, instructive, and descriptive force distinct and developmental related linguistic demands on L2 students (Yule, 1997).

Before summarizing the influence on CAF constructs between the writing tasks, it is worthwhile to mention that they were mostly group-specific. It is true that the various CAF impact could be similar among learning context groups in the complex task, but they would be different among all (or at least two) of these groups in the non-complex task. To clarify, the three low-proficiency groups' syntactic complexity remained unchanged in the argumentative task in the short term. In contrast, in the classification task, TLC1 equally increased and decreased, BLC1 deteriorated, and OLC1 developed more than they declined. Moreover, the three high-proficiency groups' lexical variation decreased in the long term, whereas it deteriorated more than it improved for TLC6, and increased for both BLC6 and OLC6 in the classification task. Furthermore, the manipulation of the task complexity in the short term led some constructs to remain unchanged as in the case of syntactic complexity (e.g., TLC1, BLC1, OLC1, and BLC6), lexical density (e.g., BLC1), lexical sophistication (e.g., OLC1 and OLC6), lexical variation (e.g., BLC1, OLC1, and TLC6), accuracy (e.g., OLC6), and fluency (OLC1 and OLC6). It also resulted in more increases than decreases in lexical variation (e.g., TLC1), improving more or less in accuracy (e.g., TLC1, BLC6, BLC1, and OLC1), increasing or decreasing more in fluency (e.g., BLC1 and TLC1), declining less in syntactic complexity (e.g., OLC6) and more in lexical density (e.g., TLC6) and lexical variation (e.g., BLC6 and OLC6), and more or less in lexical sophistication (e.g., TLC6, BLC6, and BLC1). Besides, increasing task complexity in the long term caused some

constructs to increase more than decrease (e.g., lexical variation: TLC1), equally develop and deteriorate (e.g., fluency: TLC6), improve more or less (e.g., accuracy: TLC1, BLC6, and OLC1), increase more (e.g., lexical variation: BLC1; syntactic complexity: TLC6; fluency: BLC1 and OLC6). It further contributed to making some constructs decline more (e.g., lexical sophistication: TLC1, OLC1, TLC6, and BLC6; lexical variation: OLC1, TLC6, BLC6, and OLC6; syntactic complexity: OLC6) and remain uninfluenced (e.g., fluency: OLC1).

All of the above-mentioned results only partially corroborate those of other studies that addressed task complexity and the predictions of the Cognition Hypothesis and the Limited Attentional Capacity Model. In other words, the findings were inconsistent with the results of other research (e.g., Abdali & Fatemipour, 2014; Abdollahzadeh & Kashani, 2011; Jong, 2009; Kuiken & Vedder, 2007a, 2008; 2011; Kuiken et al., 2005; and others), and with Robinson's and Skehan and Foster's models, but some CAF constructs and sub-constructs remained unaffected (e.g., syntactic complexity, lexical density, lexical sophistication, lexical variation, accuracy, and fluency), increased more than decreased (e.g., lexical variation), equally developed and deteriorated or only developed (e.g., fluency), decreased or increased more or less (e.g., syntactic complexity, lexical sophistication, and accuracy)[8] in the difficult task. Skehan and Foster (2001) maintained that since EFL learners have limited attentional capacity, they would prioritize one/two constructs at the expense of other constructs. The findings here prove that the attentional capacity of the Saudi EFL undergraduates is limited but from a different angle in that it could occur within the same construct leading these learners to both increase and decrease or could lead them to focus or not focus on a construct; yet, not to the level of improving or declining in it significantly or more than in the simple task. Such results would not have been found if the effect on CAF constructs were not specified first within each task before comparing them across tasks, let alone the fact of increasing the number of metrics incorporated for each construct, which was not found in any of the cited studies.

Nonetheless, in line with the Cognition Hypothesis, some CAF constructs developed in the complex task as in the case of accuracy (e.g., Abdali & Fatemipour, 2014; Abdollahzadeh & Kashani, 2011; Ishikawa, 2006, 2007; Jong, 2009; Kuiken & Vedder, 2007a, 2008; 2011; Kuiken et al., 2005; Mohammadzadeh Mohammadabadi et al., 2013; Nikou & Eskandarsefat,

---

[8] We mean by this the few cases in which a construct either increases or decreases on both the complex and non-complex tasks but to varying degrees. The essence of Skehan and Foster's predictions and those of Robinson is that when a construct increases in one task, it would decreases in the other task.

2012; Pourdana & Behbahani, 2011; Pourdana et al., 2011), syntactic complexity (e.g., Abdali & Fatemipour, 2014; Abdollahzadeh & Kashani, 2011; Ishikawa, 2006, 2007; Ismail et al., 2012; Li, 2000; Masrom et al., 2015; Nikou & Eskandarsefat, 2012; Pourdana & Behbahani, 2011; Pourdana et al., 2011; Rezazadeh et al., 2011; Salimi & Dadshpour, 2012; Sasayama, 2011), and lexical variation (e.g., Frear & Bitchener, 2015; Ishikawa, 2006, 2007; Jiaxin, 2015; Jong, 2009; Kuiken & Vedder, 2007b; Li, 2000). Ishikawa (2007) found that the EFL learners in the there-and-then task outperformed their counterparts in the here-and-now task in accuracy, syntactic complexity, and lexical variation. Again, what was not commensurate with the prediction of Robinson and those of Skehan and Foster, is the increase of fluency (e.g., Ishikawa, 2006, 2007; Rahimpour & Hosseini, 2010; Salimi & Dadshpour, 2012; Nikou & Eskandarsefat, 2012) that was found in the argumentative task. Rahimpour and Hosseini (2010) concluded that the EFL students were significantly more fluent when they performed a difficult narrative task than when they wrote a simple narrative task.

In support of the Limited Attentional Capacity, some CAF constructs deteriorated in the difficult task, such as lexical sophistication (e.g., Masrom et al., 2015), lexical variation (e.g., Ishikawa, 2006; Kuiken & Vedder, 2007b; Masrom et al., 2015), lexical density (e.g., Ishikawa, 2007), syntactic complexity (e.g., Frear & Bitchener, 2015; Jiaxin, 2015; Nikou & Eskandarsefat, 2012), and fluency (e.g., Bonzo, 2008; Dickinson, 2014; Ferreira, 2013; Jiaxin, 2015; Rezazadeh et al., 2011; Sponseller & Wilkins, 2015; Pourdana & Behbahani, 2011; Pourdana et al., 2011). The decrease of fluency was predicted by Robinson, as well. Masrom et al. (2015) stressed that in spite of EFL students in the most difficult condition (i.e., + causal reasoning demand − task instruction) outscoring the other four simpler conditions (i.e., + causal reasoning demand + task instruction, and − causal reasoning demand +/- task instruction) in syntactic complexity, they obtained lower scores in lexical sophistication than (+ causal reasoning demand + task instruction) condition and in lexical variation than (- causal reasoning demand − task instruction) condition. In Ishikawa's (2006) study, the high-proficiency participants decreased their lexical variation in the there-and-then task (TTR: 47.15 / 8.2), and elsewhere (2007) the there-and-then subjects scored less, close to reach significance, in lexical density in two measures (lexical to function word [L/F]: 82.05 / 10.56, and Lexical words to words[L/W]: 44.90 / 3.09) than the here-and-now subjects (L/F: 85.76 / 12.35 , and L/W: 45.95 / 3.51). Jiaxin (2015) contended that increasing the difficulty of the task led to a significant deterioration of syntactic complexity, fluency, and accuracy. The deterioration of accuracy was; however, not found in current study.

5.2 Research Question Two: 'Which of the three learning environments: TLC, BLC, and OLC, lead to the most/least increase or decrease in the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates across two writing tasks: classification and argumentative? And when?'

The second research question of this study aimed to determine when and which of the TLC, BLC, and OLC led to the most/least increase or decrease in the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates across two text types: classification and argumentative. Thus, two comparisons were carried out using the ANOVA statistical tool: first between the short-term results of each proficiency level, and then between their long-term results. In any composition task and for any CAF construct, the learning context whose group increased the most, increased in more measures than those of the other learning context groups, or decreased positively (i.e., in accuracy E/W and E/T measures) the most, would be regarded as the most effective. The same procedure had been applied toward designating the learning context that could be the least effective, that is, when its group increased the least or negatively (i.e., increased in accuracy E/W and E/T measures), or decreased the most or in more measures than those of the other groups. Nevertheless, the short-term and long-term findings here indicate that, regardless of how proficient the Saudi EFL university students were, none of the three learning environments led its participants to obtain the most/least development/deterioration in all CAF constructs in either task, and that, in each writing task, the level of development and deterioration in some CAF constructs was the same between groups across all measures, while in some others it was different in only one or a few measures.

For example, at the low-proficiency level, the results show that there were no significant differences between the scores of the three groups' (TLC1, BLC1, and OLC1) lexical density and fluency in the classification task and their lexical density and lexical sophistication in the argumentative task in the short term and long term. In the classification task, the findings revealed that in the short term, TLC1 outperformed BLC1 and OLC1 in syntactic complexity; BLC1 outscored TLC1 and OLC1 in lexical variation, and OLC1 performed better than TLC1 and BLC1 in accuracy. In the long term, both BLC1 and OLC1 obtained the highest scores in accuracy (across various measures) and in lexical sophistication for only the latter group. In the short term, to elucidate, TLC1 was the only group to increase in one measure of syntactic complexity (DC/C = 0.096), whereas its counterparts in both BLC1 and OLC1 declined. The deterioration of BLC1; yet, was the greatest (DC/C = -0.114*). Conversely, in the short term, only BLC1 developed in six measures of lexical variation, one of which was the most

(NDWERZ = 1.213; NDWESZ = 2.420; MSTTR = 0.045; Uber = 2.235*; VV1 = 0.133, and ModV = 0.013); TLC1 was the only group to decline in five of these measures, one of which the greatest (NDWESZ = -0.220; MSTTR = -0.003; Uber = -1.136; VV1 = -0.091*; and ModV = -0.033). In addition, all of the participants in TLC1, BLC1, and OLC1 decreased positively in one measure (E/W), accuracy, with OLC1 deteriorating the most (-6.282*) and TLC1 declining the least (-1.791) in the short term. In the long term, OLC1 continued to decline positively the most in the previous measure (-5.511) and in E/T (-0.933), but TLC1 increased negatively in E/W (0.271) and deteriorated positively the least in E/T (-0.107). BLC1, on the other hand, improved the most in two other measures (EFT and EFT/T) of accuracy (3.933*, and 0.213*, respectively), while TLC1 increased the least in EFT (0.267) and decreased in EFT/T (-0.014). In the long term, OLC1 developed more than TLC1 in two measures (VS2 and CVS1), lexical sophistication (0.0630* and 0.282*, respectively), whereas BLC1 deteriorated in these measures (-0.447 and -0.151, respectively).

In the argumentative task, the results indicate that in the long term, BLC1 stood alone in outperforming TLC1 and OLC1 in syntactic complexity, fluency, accuracy, and lexical variation. Specifically, BLC1 was the only group to improve in one metric of syntactic complexity (CP/C = 0.103) and fluency (W = 27.600), but both TLC1 and OLC1 declined. The deteriorations of TLC1 in these constructs were; nonetheless, the greatest (CP/C= -0.054* and W = -38.067*). Moreover, only BLC1 increased in three metrics of lexical variation (NDW = 15.133; SVV1 = 2.338; and CVV1 = 0.272), and even its decrease in VV1 (-0.006) was lower than that of OLC1 (-0.065*). The latter group also deteriorated in the previous metrics, two of which the most (NDW = -4.400; SVV1 = -2.065*; and CVV1 = -0.226*). Furthermore, although the three groups developed in two metrics of accuracy, the developments of BLC1 were the greatest (EFT = 4.333*, and EFT/T = 0.247*), while they were the least for TLC1 (EFT = 0.067 and EFT/T = 0.064).

On the other hand, at the high-proficiency level, there were no significant differences between the scores of the three groups' (TLC6, BLC6, and OLC6) lexical density and fluency in the classification task, and their syntactic complexity, lexical density, lexical sophistication, and fluency in the argumentative task in the short term and long term. However, in the classification task, BLC6 performed better than TLC6 and OLC6 in syntactic complexity in the short term, whereas in the long term, OLC6 performed the best compared to TLC6 and BLC6 in lexical variation and accuracy and the worst in lexical sophistication. Only BLC6 improved in two measures (DC/C and CN/C) of syntactic complexity (0.058 and 0.132, respectively), and TLC6 and OLC6 increased the most in one measure and decreased in the

other in the short term. The development was in DC/C (0.065*) for TLC6 and in CN/C (0.286*) for OLC6, and the deterioration was in CN/C (-0.181) for the former group and in DC/C (-0.097) for the latter group. In the long term, all of three groups declined in one measure (LS2), lexical sophistication, with OLC6 decreasing the most (-0.151*) and TLC6 deteriorating the least (-0.045). Besides, in the long term, OLC6 was the only group to increase in five measures (TTR, CTTR, RTTR, LogTTR, and Uber), one of which the greatest, of lexical variation (0.034; 0.065; 0.093; 0.010; and 0.792*), but TLC6 deceased in all of them (-0.059*; -0.549; -0.778; -0.022; and -2.812). Both OLC6 and BLC6 increased and decreased positively in three metrics of accuracy, while TLC6 did the same in only two measures and increased negatively in one measure. In other words, OLC6's improvement and positive deteriorations were the greatest in EFT (3.300*), E/T (-0.956*), and E/W (-5.774*), as TLC6's increase and positive decrease were the least in EFT (0.041), and E/T (-0.119) and its negative development was in E/W (0.600).

In the argumentative task, TLC6 outscored BLC6 and OLC6 in lexical variation only in the short term, and OLC6 outperformed TLC6 and BLC6 in lexical variation as well as accuracy in the long term. TLC6 was the only group to improve in one metric (ModV), lexical variation (0.012) in the short term, whereas both BLC6 and OLC6 decreased. The deterioration of BLC1; nevertheless, was the greatest (-0.038*). In contrast, in the long term, only OLC6 increased in three other metrics (NDW, SVV1, and CVV1) of lexical variation (4.200, 1.406, and 0.167, respectively), but both TLC6 and BLC6 declined. The decreases of TLC6 in such metrics were the greatest (-18.000*, -2.982*, and -0.305*, respectively). Again, all of these groups increased and positively decreased in two metrics (EFT and E/W) of accuracy in the long term. Yet, OLC6's development and positive deterioration were the greatest (4.500* and -5.212*, respectively), while they were the least for TLC6 (0.019 and -0.200, respectively).

From the above findings, it can be inferred that none of the learning environments was found to be the most or the least effective in enhancing CAF constructs and sub-constructs of the participants in either task. As a matter of fact, in each proficiency level, there were both insignificant and significant differences in CAF scores between the six groups in both text types, indicating that the level of effectiveness and ineffectiveness of the three learning contexts varied. That is, the level of success/unsuccess was similar in the short term and long term between the three learning contexts in terms of enhancing the low-proficiency students' lexical density and fluency in the classification task and their lexical density and lexical sophistication in the argumentative task. It was also of the same level of

effectiveness/ineffectiveness with regard to promoting the high-proficiency undergraduates' lexical density and fluency in the classification task and their syntactic complexity, lexical density, lexical sophistication, and fluency in the argumentative task. The reason for these results could be attributed to the writing instructors, the students themselves, and/or their fellow students, as we will discuss in the following.

To clarify, in this research, various types of feedback methods (i.e., teacher-correction, peer-correction, and self-correction) were incorporated as a part of the process genre approach applied on all of the three learning environment participants. Nonetheless, these editing types differed substantially from one another (Ferris, 2007) with respect to their influence on writing skills (Lundstrom & Baker, 2009; Peterson, 2010; Rana & Perveen, 2013), were appreciated by and deemed beneficial for EFL learners in general (e.g., Abdollahifam, 2014; Chiramanee & Kulprasit, 2014; Ferris, Liu, Sinha, & Senna, 2013; Ibarrola, 2013; Jamalinesari, Rahimi, Gowhary, & Azizifar, 2015; Miao, Badger, & Zhen, 2006; Morgan, Fuisting, & White, 2014), and for Saudi EFL students in particular (e.g., Al-Hazmi & Scholfield, 2007; Alshahrani & Storch, 2014; Alqurashi, 2015; Faqeih, 2015; Grami, 2005; Halimi, 2008; Hamouda, 2011; Jahin, 2012; Mahfoodh & Pandian, 2011; Mahmoud & Oraby, 2015). In the bulk of the latter studies, researchers found that not only did Saudi EFL learners view these methods as the perfect means to improve their writing skills (cf. Mustafa, 2012; M. Saba, 2013), but they also took the comments and suggestions made by the teachers and fellow-students on board. However, the peer-correction and self-correction are based on the teacher-correction. That is, Alghizzi (2011, 2012) asserted that the types of composition aspects/features that Saudi EFL students in the traditional classrooms would most likely try to perfect while learning writing skills would be done by those who had been focused on by their instructors (see Al-Hazmi & Scholfield, 2007). The issue here is that, during the experiment, despite enlarging the learning contexts to include blended and online and all teachers being forced to view themselves as writing skill instructors (cf. Zamel, 1985; Gascoigne, 2004) whose main concern was guiding students through how to perfect their language (i.e., content, clarity, capitalization, spelling, grammar, organization, style, and punctuation)—as they are the criteria adopted for correcting their writing course final exams—the instruction and discussion of how these students could develop their CAF constructs were neglected or disregarded. With the lack of teacher support and guidance for such matters, these undergraduates were left to exploit and employ their learning environments' strengths independently to enhance their writing skills. This might have led them to continually and unconsciously prioritize (Knoch, 2008; Skehan, 1996, 1998, 2009; Skehan & Foster, 1996,

1997, 2001) the same constructs and sub-constructs—even though they were in different learning contexts—and therefore, performed similarly in both the short term and long term. Still, the other CAF construct scores showed that the level of success and unsuccess of the three learning environments was different and dependent on the timescales (i.e., short term and/or long term), CAF constructs, proficiency levels, and writing tasks.

In other words, the TLC was the most beneficial in developing the syntactic complexity of the low-proficiency learners in the classification task, and the lexical variation of the high-proficiency students in the argumentative task, but only in the short term. The justification could be that these participants are accustomed to the traditional learning environment atmosphere (Deveci, 2015), which led them to outperform the learners in the other learning environments (BLC and OLC). The success of the TLC further implies that it could be equally beneficial for both low- and high-proficiency participants, while its benefit would be temporary and limited to one construct (syntactic complexity or lexical variation), in one writing task (classification or argumentative) for each proficiency level group (low or high). Such temporariness could mean that when adopting a novel approach to teaching writing (i.e., process genre) to Saudi EFL undergraduates who are accustomed to a product approach and whose pre-university instructors taught them the skills by means of learning grammar, vocabulary (Alharbi, 2015; Alrasheed, 2008; Al-Seghayer, 2014a, 2015), and linguistic features at the sentence level (Al-Hazmi, 2006), and memorizing ad hoc samples (Alghizzi, 2011, 2012; Elyas & Picard, 2010; Shukri, 2014), the TLC participants would need less time to adjust themselves to the new method—and perform better, at least in these constructs—than their counterparts in the other learning contexts (BLC and OLC).

In addition, in constructs other than the BLC and OLC, the TLC was the least effective in improving the low-proficiency students' lexical variation in the short term and their accuracy in both the short term and long term in the classification task, and their syntactic complexity, fluency, and accuracy in the long term in the argumentative task. It was also unsuccessful in increasing the lexical sophistication of the high-proficiency learners in the long term in the classification task, though not to the level of OLC and BLC. [9] Moreover, it was the least successful, along with OLC, in enhancing the high-proficiency students' syntactic complexity in the short term in the classification task, and was solely the least successful in promoting their lexical variation and accuracy in the long term in both tasks.

---

[9] All of the three learning environment groups decreased significantly in lexical sophistication, but the deterioration of OLC6 was the most, followed by BLC6, and lastly TLC6.

The reason could be that the participants lost their motivation and autonomy and were restricted by the limited resources the TLC provided them. In other words, both autonomy and motivation are essential for better and long-lasting acquisition and development of the writing skills (Honsa, 2013), and although Saudi EFL learners are motivated to learn English (Al-Jarf, 2008; Faruk, 2013) and can be taught to be autonomous toward learning writing skills (Al-Hazmi & Scholfield, 2007), it appeared that the TLC instructor's effort to increase his students' autonomy and motivation (Alshehri, 2012; Deneme, 2011; Jahin & Idrees, 2012) by adopting the process genre approach was not sufficient to keep students motivated. One common characteristic of Saudi EFL learners, besides considering the writing skills to be hardest, is that they can easily lose their learning motivation for a variety of reasons (Liton, 2012; Alshehri, 2012), one of which is when they are being confined to two sources of information (i.e., teacher and fellow-classmates) with no exposure to technological tools (Internet) by which to access information provided by other writing experts.

On the other hand, the BLC was the most beneficial in promoting the low-proficiency learners' lexical variation in the short term and their accuracy in the long term in the classification task, and their syntactic complexity, fluency, accuracy, and lexical variation in the long term in the argumentative task. Furthermore, it was only the most effective in developing the syntactic complexity of the high-proficiency students in the short term in the classification task. The significant performance of these participants could be attributed to the possibility that they made full use of the strengths of their learning environment. The blended learning context is regarded as convenient, effective, and efficient in improving meaningful learning experiences (Garrison & Kanuka, 2004; Vaughan & Garrison, 2006). What might be the case here is that this learning context actually accommodated its Saudi EFL participants' various learning styles; increased their learning engagement, self-confidence, and the time allotted to studying; provided them with a less stressful environment to practice writing inside and outside the classroom with more personalized and individualized learning support and experiences; and helped them to develop necessary and valuable updated learning skills (Klímová, 2008; Marsh, 2012; Pardo-Gonzalez, 2013; Peachey, 2013). The context—with its usage of different platforms (i.e., blogs and text chat)—might not have only aided these students in establishing a sense of collaborative community to share their learning experiences, address writing issues, and access extra learning materials (Aborisade, 2013), it might also have increased their motivation, self-esteem, enjoyment, focus, autonomy, and acceptance of writing (Alanazi, 2013; Aljumah, 2012; Eydelman, 2013; Larsen, 2012).

All of these findings are generally commensurate with the body of research that has emphasized the positive impact of the blended learning environments on the writing skills (e.g., Abdelrahman, 2013; Al Fadda et al., 2011; Ahmed, 2015; Al-Hammadi, 2010; Al-Jarf, 2002, 2004a, 2004b, 2006, 2007a, 2009a; Al-Menei, 2008; Alsaleem 2014; Alshumaimeri, 2011; AbuSeileek; 2006; Fageeh, 2011; Ghahari & Ameri-Golestan, 2014; Janfaza et al., 2014; Liu, 2013; Mahmoud, 2014; Miyazoe & Anderson, 2010; Niazi & Pourgharib, 2013; Roy, 2012; Shih, 2011; Yoon & Lee, 2010). For example, Ghahari and Ameri-Golestan (2014) found that the Iranian blended learning group significantly outperformed the traditional group in their writing performances. Miyazoe and Anderson (2010) concluded that the integration of wiki, blogs, and forums were effective in increasing Japanese EFL learners' ability in distinguishing between English writing styles. Liu (2013) maintained that employing blended learning in academic English writing courses helped develop EFL Chinese academic writing ability. Ahmed (2015) reported that the writing of Saudi EFL female undergraduates who were exposed to blended learning (via Twitter) improved more significantly in ideas and content, organization, style, and voice than their counterparts taught only in traditional classrooms. Alsaleem (2014) found that the use of the WhatsApp tool by Saudi EFL female university students resulted in significant development in word choice and voice. A study by Al Fadda et al. (2011) showed that fifth grade Saudi female EFL students who were exposed to the WebQuests tool significantly outperformed their fellow students who were taught only in a traditional classroom in terms of writing organization, content, length, vocabulary, grammar, and total writing overall score. Mahmoud (2014) asserted that the incorporation of social networks (i.e., email and Facebook) on Saudi EFL male university students led to a positive attitude toward such technologies and improved their writing achievements. Fageeh (2011) concluded that the use of weblogs led to better writing production by the Saudi EFL male undergraduates in comparison to their counterparts in the traditional classroom.

The results are also partially in line with some studies in which some CAF constructs were part of the data analyses. Fellner and Apple (2006), for instance, found that using weblogs with traditional classroom resulted in a significant increase in the text length and lexical complexity of Japanese EFL learners. Shang (2007) reported that incorporating email exchanges between Taiwanese EFL students led to significant improvement in the written syntactic complexity and grammatical accuracy, whereas lexical density remained unaffected. Fitze (2006) concluded that there were no differences between the EFL participants' production in class conferences and in the electronic conferences in terms of word count, but

the lexical range increased significantly in the electronic-conference writing. Finally, although Zaid (2011) cited no differences among the three groups of Saudi EFL undergraduates he studied, the participants in the blended learning contexts (i.e., online concept-mapping and online reading before writing) produced larger and richer drafts than the participants in the traditional learning classrooms.

Nevertheless, the best performance of the two BLC groups in the current study appeared in many constructs (short term and/or long term) in the two text types for the low-proficiency participants (BLC1), while it was in only one construct (short term) in one text type for high-proficiency learners (BLC6). This implies that the lower the writing competency of the participants was, the higher the effect of the blended learning environment on their written production. This could be attributed to the possibility that the low-proficiency learners—as compared to the high-proficiency students—were not skeptical about the benefits of BLC on their writing ability and had already acquired the elements that are important for the successful usage of that learning context, such as computer and Internet competency, learning skills, owning a computer, and overall positive attitudes (Al Zumor et al., 2013; Al-Harbi, 2011). Again, the best short-term production of lexical variation by BLC1 and that of syntactic complexity by BLC6 in the classification task reveal that the Saudi EFL undergraduates can quickly familiarize themselves with the new blended learning environment and make the best of it. Alpala and Flórez (2011) have maintained that the implementation of the blended course should take place gradually as EFL learners need time to acquaint themselves with its technologies, and this could be true in the case of the CAF constructs where the low-proficiency participants surpassed their rivals in the other learning contexts (TLC and OLC) only in the long term and those constructs where such participants and those of a high-proficiency level underperformed in the short term and/or the long term. The BLC was the least successful in improving the low-proficiency learners' syntactic complexity in the short term, and their lexical sophistication in the long term for the classification task, whereas it was the least effective in doing so for the high-proficiency students' lexical variation in the short term for the argumentative task. The reason for this, which could also be used to justify the aforementioned, could be the incapability of the low- and high-proficiency undergraduates to keep up with the demands of such a learning environment, writing skills, and those of the other courses.

The OLC was the most beneficial in increasing the low-proficiency learners' accuracy in both the short and long terms, as well as their lexical sophistication in the long term, but only for the classification task. It was also the most effective in enhancing the lexical

variation and accuracy of the high-proficiency students in the long term for both tasks. All of this could be linked to what many scholars have called the perspective advantages of teaching using technologies. That is, the online learning context with its technological tools has actually helped the students devote more time to focusing on grammar and linguistic accuracy, lessen their learning anxiety, and increase their motivation, responsibility (Hurd, 2005), activeness, confidence, enjoyment, and involvement in their learning (Ilter, 2009; Yang & Chen, 2007; Wu et al., 2011). It could also have provided them with not only access to expert knowledge and input (Roux et al., 2014), but knowledge that is genuinely contextualized and that is not restricted by the borders and boundaries of the traditional classrooms, which eventually develops their pragmatic, cultural, and linguistic competence (Arnold & Paulus, 2012).

The aforementioned could be said to be in line with the studies that investigated the influence of teaching writing remotely (via different tools) on EFL student productions (e.g., Ahmed, 2013; Al-Abed Al-Haq & Al-Sobh, 2010; Alhassan & Gashan, 2013; Alzu'bi & Sabha, 2013; Abdul Fattah, 2015; Hayati & Gooran, 2014; Sayadi & Khoshsima, 2016; Xiao's; 2008). Ahmed (2013) found that teaching composition through email to Saudi EFL undergraduates contributed significantly to better sentence structure, text organization, paragraphing, and style/register than those taught in the traditional classroom. Al-Abed Al-Haq and Al-Sobh's (2010) study revealed that Jordanian pre-university EFL students instructed in a web-based writing program outperformed their counterparts who were taught in regular classrooms. Alzu'bi and Sabha (2013) showed that mobile-based email teaching resulted in significant development of the Jordanian EFL students' vocabulary acquisition and writing ability. Abdul Fattah's (2015) research indicated that the writing of the experimental Saudi EFL undergraduate group instructed via the WhatsApp program improved significantly in terms of writing punctuation and vocabulary and outscored the control group who were taught in a traditional classroom in punctuation and sentence structure. Hayati and Gooran (2014) investigated the impact of teaching writing short passages in classrooms and through email and maintained that there was a positive relationship between using email and Persian EFL learners' writing production. Sayadi and Khoshsima (2016) found that the experimental Persian EFL students who were taught the skills mainly by receiving instructions via PowerPoint (virtual learning) and the control group who were instructed in traditional classroom both developed in writing; yet, the improvement of the former group was greater than the latter group. Xiao (2008) examined the effect of peer review implemented in distance-taught EFL writing and concluded that it led to positive influence in terms of

evolving Chinese EFL's writings and developing their transferrable skills like self-efficacy, collaboration, problem-solving, and self-management skills. The results of the OLC groups were partially commensurate with those of Alhassan and Gashan (2013), who concluded that the Saudi EFL learners who were exposed only to WebQuest writing instruction improved significantly in text length, vocabulary, and grammar and outscored those who received only traditional writing instruction.

The significant performance of the two OLC groups in the current study was observed to be in two constructs (short term and/or long term) in one writing task for low-proficiency participants (OLC1) and in both writing tasks for the high-proficiency learners (OLC6). This implies that regardless of how competent the Saudi EFL undergraduates were in writing, the online learning environment would be beneficial for developing their writing ability. This would not have happened if the low- and high-proficiency EFL learners had not had the indispensable factors necessary for the ultimate benefit of the OLC like a positive attitude toward it, enjoyment, motivation, confidence, intention to use, and computer and Internet proficiency (Bendania, 2011). In addition, OLC1's production of accuracy in the short term in the classification task indicates that the low-proficiency Saudi EFL university student could adjust to the novel online learning context and benefit from it quickly. This only partially supports Wang, Calandra, and Yi's (2010) finding that even if the EFL learners had a high computer and Internet knowledge and had finished a general computer competency course, they would need more time to familiarize themselves with the technology adopted in an online course. The evidence of such a claim is found in the other CAF constructs in which the OLC groups outscored their counterparts in the other learning environments (TLC and BLC) in the long term and those in which they either performed the worst or lower than one of the latter learning environment groups. The OLC was found to be the least successful merely in promoting the low-proficiency Saudi EFL learners' lexical variation in the long term in the argumentative task. It was also the most unsuccessful in developing the lexical sophistication in the long term for the high-proficiency students, and was the least beneficial, along with TLC, in improving their syntactic complexity in the short term in the classification task.

Chapter Six
Conclusion

6.0 Introduction

This chapter recaps and summarizes the results of this study. It then points out some pedagogical implications, recommendations, limitations, and suggestions for future research.

6.1 Summary of the Results

The purpose of the current research was to build on the literature of second language writing skills by investigating two areas that have either never been investigated or have rarely been investigated in a Saudi context. For this investigation, two research questions were generated. The first research question attempted to determine when and how the CAF constructs and sub-constructs of the low- and high-proficiency Saudi EFL undergraduates in the three learning environments (TLC, BLC, and OLC) were affected longitudinally across two writing tasks (classification and argumentative) that differed in their level of complexity. The second research question was designed to specify when and which of these learning environments led to the highest and lowest increase and/or decrease of their low- and high-proficiency Saudi EFL participants' CAF in the two writing tasks. A review of the literature posited some pieces of evidence for two competing models (the Limited Attentional Capacity and the Multiple Attentional Resources), which were proposed to explain how the EFL learners' attention is deployed during the performance of these tasks, and how these tasks' level of complexity would lead to a particular influence on CAF constructs. The review also highlighted the effectiveness and ineffectiveness of each learning context—sometimes, in comparison to one other learning context—on the EFL learners' written productions in general, and on their CAF constructs in particular. To answer the above-mentioned questions, 75 Saudi EFL university students were recruited from a pool of two proficiency levels (low and high) and were divided into six groups, each of which was exposed to one learning context that was either traditional, blended, or online. These students' written productions in the pre-test, mid-term test, and post-test were collected and analyzed using two statistical tests: the t-test and the ANOVA test.

The findings of the t-test for the first research question showed that:

1) In each group, there were both similar and different impacts on CAF constructs between the classification and argumentative tasks in the short term and long

term. The only exception was the BLC1 group, in which there were only different effects on all CAF constructs between the two writing tasks in the short term.

2) However, the similarities and differences of influence on CAF constructs between these composition tasks appeared to be group-specific, in that they were dependent on the learning context, proficiency level, and timescales (i.e., short term and long term). In other words, regarding whether a construct in the two writing tasks was impacted similarly or differently, such impact did not generally occur in a systematic way and across the same number and types of measures for the same group—except for TLC1, OLC1, BLC6, and OLC6's lexical density, BLC6's fluency, and TLC6's accuracy—in both the short term and long term, or even across the groups with the same or different proficiency levels.

3) The predictions of Skehan and Foster, and those of Robinson, were not fully supported not just because of the similar effect on CAF constructs, but also because of some of the different effects identified for each group. This means that none of the six groups' total results was consistent with the two-model hypotheses. For instance:

   a) In line with Robinson's model, increasing task complexity led to increases in accuracy (e.g., TLC1 and BLC6 = short term and long term), lexical variation (e.g., BLC1 = long term), and syntactic complexity (e.g., TLC6 = long term).

   b) Commensurate with Skehan and Foster's model, elevating task complexity resulted in decreases in lexical density (e.g., TLC6 = short term), lexical sophistication (e.g., TLC6 and OLC1 = short term and/or long term), lexical variation (e.g., BLC6, OLC6, TLC6, and OLC1= short term and/or long term), and syntactic complexity (e.g., OLC6 = long term).

   c) In support of the two models was the finding that altering task complexity contributed to deteriorations in fluency (e.g., TLC1 = short term).

   d) Nevertheless, it did not corroborate the predictions of Skehan and Foster, and those of Robinson, when increasing task complexity caused some constructs to remain unchanged (e.g., syntactic complexity: TLC1, BLC1, OLC1, and BLC6 = short term; Lexical density: BLC1 = short term; lexical sophistication: OLC1 and OLC6 = short term; lexical variation: BLC1, OLC1, and TLC6 = short term; accuracy: OLC6 = short term; and fluency: OLC1 and OLC6 = long term and/or short term). It was also inconsistent with the two models when elevating task complexity led to increases more than

217

decreases in lexical variation (e.g., TLC1 = short term and long term) and equal development and deterioration or only development in fluency (e.g., TLC6, BLC1, and OLC6 = short term and/or long term), less increases in accuracy (e.g., OLC1 and BLC1 = long term and/or short term), less or more decreases in lexical sophistication (e.g., BLC1, TLC1, and BLC6 = short term and/or long term), and less decreases in syntactic complexity (e.g., OLC6 = short term).

On the other hand, the findings of the ANOVA test indicated mixed results, since each of the three learning environments led to benefits in some ways. In other words, in the two proficiency levels, the TLC, BLC, and OLC had the same level of success/unsuccess in developing all the measures of some CAF constructs in both writing tasks in the short term and long term. Yet, in the other CAF constructs, there was no uniform linear development or deterioration of all measures across the six groups. In each of these constructs, the differences between these groups emerged from one or more measures, but not from all measures. Each of these learning contexts stood alone in being the most or least successful in increasing some constructs. Nonetheless, this was dependent on participants' proficiency levels, text types, and timescales. Some conclusions are drawn below:

1) The level of effectiveness/ineffectiveness of TLC, BLC, and OLC, in both the short term and long term, was the same in terms of enhancing the low-proficiency participants' lexical density and fluency in the classification task, and their lexical density and lexical sophistication in the argumentative task. It was also the same, in both the short and long term, in promoting the high-proficiency students' lexical density and fluency in the classification task, and their syntactic complexity, lexical density, lexical sophistication, and fluency in the argumentative task.

2) The TLC was found to be effective for both the low- and high-proficiency Saudi undergraduates, and the effectiveness was temporary and limited to one construct in one writing task in both the low- and high-proficiency participants:

   a) The TLC was the most beneficial in the short term in developing the syntactic complexity of the low-proficiency learners in the classification task, and the lexical variation of the high-proficiency EFL students in the argumentative task.

   b) However, the TLC was the least effective in improving the low-proficiency students' lexical variation in the short term and their accuracy in both the

short term and long term in the classification task, and their syntactic complexity, fluency, and accuracy in the long term in the argumentative task. It was also unsuccessful in increasing the lexical sophistication of the high-proficiency learners in the long term in the classification task, though not to the level of OLC and BLC. Moreover, it was the least successful, along with OLC, in enhancing the high-proficiency students' syntactic complexity in the short term in the classification task, and was solely the least successful in promoting their lexical variation and accuracy in the long term in both tasks.

3) The BLC appeared to be more effective for the low-proficiency EFL participants than the high-proficiency learners. Such effectiveness was evident in many constructs (short term and/or long term) in both text types for the former group, while it was in only one construct (short term) in one text type for the latter group:

   a) The BLC was the most beneficial in developing the low-proficiency learners' lexical variation in the short term and their accuracy in the long term in the classification task, and their syntactic complexity, fluency, accuracy, and lexical variation in the long term in the argumentative task. Furthermore, it was only the most effective in improving the syntactic complexity of the high-proficiency students in the short term in the classification task.

   b) Yet, the BLC was the least successful in increasing the low-proficiency learners' syntactic complexity in the short term, and their lexical sophistication in the long term for the classification task, whereas it was the least effective in doing so for the high-proficiency students' lexical variation in the short term for the argumentative task.

4) The OLC was observed to be effective for both proficiency levels, in the sense that the effectiveness was in two constructs (short term and/or long term) in one writing task for the low-proficiency participants, and in both writing tasks for the high-proficiency learners:

   a) The OLC was the most beneficial in enhancing the low-proficiency learners' accuracy in both the short and long terms, as well as their lexical sophistication in the long term, but only for the classification task. It was also the most effective in promoting the lexical variation and accuracy of the high-proficiency students in the long term for both tasks.

   b) Nonetheless, the OLC was found to be the least successful merely in developing the low-proficiency Saudi EFL learners' lexical variation in the

long term in the argumentative task. It was also the most unsuccessful in improving the lexical sophistication in the long term for the high-proficiency students, and was the least beneficial, along with TLC, in increasing their syntactic complexity in the short term in the classification task.

6.2 Implications and Recommendations

This research provides several pedagogical implications for academic research, EFL writing instructors (both native and Arabs) at pre-university and university levels, and task-based investigators. First, the significance of the academic research is that this study expands the scope of the writing skill field literature by exploring the effect of four factors—learning environments, proficiency levels, timescales, and text types—on the CAF constructs of the Saudi EFL undergraduates. Unfortunately, there has been no previous research that addressed the aforementioned factors altogether. The topic is important not only because EFL learners in general, and Saudi EFL students in particular, regard writing as the most difficult skill and would most likely have negative attitudes towards it, but also because they are still motivated to learn it as it has been deemed essential for their success in their current studies and future careers. The Saudi EFL learners—whether they are high school graduates, undergraduates, or postgraduates—have the opportunity to be sponsored by the MoE to pursue their education abroad, mostly in English-speaking countries. The success of such a journey would not be accomplished with a poor writing competency. To be able to determine the impact of the learning contexts, proficiency levels, and text types would be of great usefulness in understanding how these variables would contribute to the development of the EFL learners' CAF constructs as aspects that are believed to be of significant importance for academic writing.

Second, this research has ramifications for EFL composition teachers. It is especially vital for teachers who have no academic and professional training in teaching writing skills, and whose specialties and expertise are in teaching other language skills (i.e., reading, speaking, and listening), or fields (i.e., literature, translation, and theoretical or applied linguistics). The study shows that although the low- and high-proficiency students in all three learning environments developed and/or deteriorated in CAF constructs and sub-constructs in the classification and argumentative tasks, when comparing their scores across each other, the level of effectiveness of the TLC, BLC, and OLC were either the same or different in the short term and the long term. This means that inasmuch as all of these learning contexts could be equally beneficial, each context could also be the most productive and counter-productive

to improving the students' writing ability. However, all of this is dependent on the CAF constructs, proficiency levels, writing tasks, and timescales. Larsen-Freeman (2009) emphasized that the CAF constructs interact with one another, their interaction changes with time, and that the constructs should be examined—and possibly taught—as a whole and not separately. Thus, it is proposed that EFL writing instructors should educate themselves on how to teach CAF constructs and work on the factors that may affect the development of students. These teachers should know when to apply a learning context for a particular proficiency level, CAF construct, and text type, and when the prospective positive findings will occur. By doing so, teachers would be required to constantly change their students' learning environment within one semester, but it is hoped that the teaching and learning method would be enjoyable, motivating, time-, effort-, and cost-effective, and, most importantly, systematic. One possible hurdle that may stand against the above-mentioned perspective is when teachers rely on the pre-service and in-service training courses that are not governed by standardized regulations and practices. As explained in the literature, these courses are not effective and lack the basic elements to produce writing teachers who are equipped to overcome unforeseen teaching issues and who must be able to continually update their teaching approaches and contexts to meet their students' learning-style preferences and needs.

Finally, the study reveals some other successful techniques for the task-based researchers on how to investigate the impact of task complexity factor. In all of the cited studies that examined the CAF written productions of EFL learners, there are some issues: incorporating limited numbers of measures, overlooking the result of one measure if it is not supported by the finding of its correct-form measure (e.g., lexical variation measures: TTR and CTTR, in Kuiken et al., 2005; Kuiken & Vedder, 2007a, 2008, 2011), not articulating the complexity level of each of the writing tasks used, and manipulating the complexity level of the same task. The other related problems are that researchers compared students' CAF scores in one learning context (traditional) across tasks in one or two particular points of time, and they took the suggestions of Skehan and Foster and those of Robinson as orthodox guidelines that cannot be alternated or amended. It is true that a few researchers did make some amendments towards these recommendations, such as manipulating the writing task complexity along both the resource-directing and resource-dispersing dimensions, but this is still not enough to address the task complexity factor. That is, by extending the learning contexts to include BLC and OLC, increasing the number of CAF measures, specifying the level of difficulty of each task with concrete reasons, and analyzing the CAF performances

first within each writing task in the short term and long term and then comparing the yielded results across the two writing tasks, the current research is able to maintain other types of influence of task complexity on CAF constructs that were beyond the predictions of Skehan and Foster and those of Robinson. All of these authors' models were partially supported. Nevertheless, the impact of task complexity did not prove that the students—regardless of their learning contexts or proficiency levels—have multiple attentional resources, but rather a limited attentional capacity. The latter was even different from what Skehan and Foster anticipated. In other words, in each learning context, when students performed the complex task, there could be various ways that CAF constructs were impacted, as well as more than one scenario on how students' attention was deployed. They could pay more, less, or no attention to different constructs, leading them to increase, remain unchanged, or decrease on these constructs. They could also prioritize some aspects of one construct while deprioritizing its other aspects, which would result in either equal development and deterioration, or more development than deterioration. Furthermore, in both the difficult and simple tasks, learners could focus on one construct and/or neglect one other, but to varying degrees. This would cause them to increase and/or decrease in the two tasks, yet the increases and decreases would be higher or lower in one task than the other. Most importantly, there are some pieces of evidence showing that even in the non-complex task, the limited attentional capacity exists, and it is therefore inadequate to describe the students' CAF performances as the previous researchers did—using basic terminologies of increasing or decreasing. Based on the learning context and proficiency level, some CAF constructs of participants may increase, decrease, remain uninfluenced, equally develop and deteriorate, or develop more or less than deteriorate. Thus, it is recommended to reconceptualize the understanding of how the attention of EFL learners is deployed during their execution of simple and difficult tasks that differ along the resource-directing dimension, and possibly the resource-dispersing dimension, as well. It is by considering other types of learning environments, texts, proficiency levels, CAF measures, and analyses, that task-based scholars would most likely be capable of providing writing skill textbook designers, researchers, and teachers with more solid models and hypotheses that can be relied on even if more factors than those already mentioned come to surface.

6.3 Limitations and Suggestions for Future Research

Despite the current goal of the study to provide outcomes that are relevant and important for L2 writing skill researchers, writing skill instructors (both native and Arabs),

and task-based investigators, it is acknowledged that there are some limitations from the research design that was adopted and the participants that were recruited, the teaching method, writing tasks, and learning contexts applied, and the CAF measures used. The first limitation emerged from the research design and selection of participants concerning conducting quantitative research without including novice, intermediate, and very advanced participants or examining the current participants' perceptions and attitudes towards the writing skill learning environments (PTWSCC), their multicultural personality (MP [Van der Zee & Van Oudenhoven, 2000; 2001; Van der Zee, Van Oudenhoven, Ponterotto, & Fietzer, 2013), and their perceptual learning-styles preference (PLSP [Reid, 1987]). Some of these different variables might have influenced the current research results in some ways. Therefore, in order to enrich the reliability and validity of the current findings, future studies are recommended to take place in more Saudi English colleges and departments, and in English-speaking-country universities where Saudi learners attend. The study would have an emphasis on adopting a triangulation method (i.e., qualitative and quantitative) by increasing the number of subjects and proficiency levels and carrying out the above mentioned types of questionnaires. By doing so, not only will the results be representative of Saudi EFL university students, but they will also draw a picture of how such participants develop in English CAF writing, and how the factors of PTWSCC, MP, and PLSP correlate with and affect the CAF development of the Saudi EFL undergraduates.

The second limitation arose from the teaching approach, text types, and learning environments concerning the fact that they were of a limited number and specific versions. In other words, in this research, there was one teaching method (process genre) and two writing tasks (classification and argumentative) used, as well as some specific types of technologies (blogs, text chat, forums, and virtual classes) by which the TLC was transformed to BLC and OLC. There is also the fact that the comparisons were only within each learning context group's performance and across the productions of the three learning context groups with the same and different proficiency levels (low and high). In the field, there is still a dire need to expand the investigations by adopting other teaching approaches (e.g., product, process, and genre), writing tasks (e.g., expository, narrative, descriptive, information-gap, decision-making, text-reconstruction, and instructional), and other technologies (e.g., Second Life, Twitter, and Facebook) for the BLC and OLC, and drawing on other types of comparisons, such as that between the novice, intermediate, and very advanced participants exposed to the same learning context (TLC, BLC, and OLC). Therefore, it is suggested that researchers should take all of these into consideration.

Finally, the third limitation that emerged from the CAF metrics is in regards to their findings and numbers. In other words, participants' writing development in terms of CAF constructs was traced longitudinally for approximately four months. There is a possibility that all of the metrics that did not yield any significant results were due to the fact that they need more time than was designated to be captured (Norris & Ortega, 2009; Pallotti, 2009). Also, the number of metrics used in this study is somehow justifiable, as they are considered to be the best predictors of CAF constructs, but what about the other metrics cited by other researchers, such as Wolfe-Quintero et al. (1998)? Although such researchers maintained the ineffectiveness of such metrics, their conclusions were still based on the findings of L2 European learners rather than Arabic speakers. As a result, future studies in the Arab world in general, and in Saudi Arabia in particular, should consider incorporating these measures to verify their validity in capturing Arab and Saudi EFL learners' writing development, and to determine if the factors of proficiency levels, learning contexts, and texts types would or would not contribute to the diversity of the findings yielded.

References

Abdali, A., & Fatemipour, H. (2014). The impact of different writing tasks on intermediate EFL learners' writing performance. *Theory and Practice in Language Studies*, *4*(4), 730–735. doi:10.4304/tpls.4.4.730-735

Abdel Latif, M. M (2009). What do we mean by writing fluency? Proposing a new measure for assessing fluent written language production. In M. Edwardes (Ed.), *Proceedings of the BAAL Annual Conference 2008* (pp. 3–4). London, UK: The British Association for Applied Linguistics.

Abdel Latif, M. M. (2011). What do we know and what do we need to know about Arab gulf EFL/ESL students' writing. *TESOL Arabia* Perspectives, *18*(2), 6–14.

Abdel Latif, M. M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, *34*(1), 99–105. doi:10.1093/applin/ams073

Abdelrahman, O. N. B. (2013). Use of dialogue through e-mail technology in developing writing skills to English as foreign language learners (EFL) at Al-Imam Mohammad Ben Saud Islamic University. *Journal of Educational & Psychological Sciences*, *14*(4), 665–688.

Abdollahifam, S. (2014). Investigating the effects of interactional feedback on EFL students' writings. *Procedia-Social and Behavioral Sciences*, *98*, 16–21.

Abdollahzadeh, S., & Kashani, A. F. (2011). The effect of task complexity on EFL learners' narrative writing task performance. *Journal of English Language Teaching and Learning*, *3*(8), 1–28.

Abdul Fattah, S. F. E. S. (2015). The effectiveness of using WhatsApp messenger as one of mobile learning techniques to develop students' writing skills. *Journal of Education and Practice*, *6*(32), 115–127.

Abedi, R., Latifi, M., & Moinzaden, A. (2010). The effect of error correction vs. error detection on Iranian pre-intermediate EFL learners' writing achievement. *English Language Teaching*, *3*(4), 168–174.

Aborisade, P. A. (2013). Blended learning in English for academic purposes courses: A Nigerian case study. In B. Tomlinson & and C. Whittaker (Eds.), *Blended learning in English language teaching: Course design and implementation* (pp. 35–42). London, UK: British Council 2013 Brand and Design.

AbuSeileek, A. F. (2006). The use of word processor for teaching writing to EFL learners in King Saud University. *Journal of King Saud University (Education, Science, and Islamic Studies)*, *19*(2), 1–15.

AbuSeileek, A. F. (2012). The effect of computer-assisted cooperative learning methods and group size on the EFL learners' achievement in communication skills. *Computer & Education*, *58*, 231–239. doi:10.1016/j.compedu.2011.07.011

Adas, D., & Abu Shmais, W. (2011). Students' perceptions towards blended learning environment using the OCC. *An - Najah Univ. J. Res. (Humanities)*, *25*(6), 1681– 1710.

Adas, D., & Bakir, A. (2013). Writing difficulties and new solutions: Blended learning as an approach to improve writing abilities. *International Journal of Humanities and Social Science*, *3*(9), 254–266.

Adobe Systems Incorporated. (2016). Adobe connect learning. Retrieved from http://www.adobe.com/products/adobeconnect/learning.html

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425–438.

Ågren, M., Granfeldt, J., & Schlyter, S. (2012). The growth of complexity and accuracy in L2 French: past observations and recent applications of developmental stages. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 95–120). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Ahmed, A. (2010). Contextual challenges to Egyptian students' writing development. *International Journal of Arts and Sciences*, *3*(14), 503–522.

Ahmed, M. A. E. A. S. (2013). The effect of using e-mail on developing EFL writing skills for secondary stage students at the experimental schools. *Arab World English Journal (AWEJ)*, *4*(3), 135–150.

Ahmed, M. A. E. A. S. (2015). The effect of Twitter on developing writing skill in English as a foreign language. *Arab World English Journal (AWEJ)*, *2*, 134-149.

Ahmed, O. N. (2012). The effect of different learning styles on developing writing skills of EFL Saudi learners. *British Journal of Arts and Social Sciences*, *5*(2), 220–233.

Ai, H. (2016a). Web-based L2 syntactical complexity analyzer. Retrieved from http://aihaiyang.com/software/l2sca/

Ai, H. (2016b). Web-based lexical complexity analyzer. Retrieved from http://aihaiyang.com/software/lca/

Akkoyunlu, B., & Soylu, M. Y. (2008). A study of student's perceptions in a blended learning environment based on different learning styles. *Educational Technology & Society*, *11*(1), 183–193.

Al-Abed Al-Haq, F., & Al-Sobh, M. A. (2010). The effect of a web-based writing instructional EFL program on enhancing the performance of Jordanian secondary students. *The JALT CALL Journal*, *6*(3), 189–218.

Al-Abed Al-Haq, F., & Smadi, O. (1996). Spread of English and westernization in Saudi Arabia. *World Englishes*, *15*(3), 307–317.

Al-Abed-AI-Haq, F., & Ahmed, A. S. E. A. (1994). Discourse problems in argumentative writing. *World Englishes*, *13*, 307–323.

Al-Ahaydib, M. E. A. (1986). *Teaching English as a foreign language in the intermediate and secondary schools of Saudi Arabia: diagnostic study* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global. (Oder No. 8619876)

Al-Ahdal, A. A. M. H., Alfallaj, F. S., Al-Awaied, S. A., & Al-Hattami, A. A. (2014). A comparative study of proficiency in speaking and writing among EFL learners in Saudi Arabia. *American International Journal of Contemporary Research*, *4*(2), 141–149.

Alamin, A., & Ahmed, S. (2012). Syntactical and punctuation errors: An analysis of technical writing of university students Science College, Taif University, KSA. *English Language Teaching*, *5*(5), 2–8. doi:10.5539/elt.v5n5p2

Alamri, A. A. (2008). *An evaluation of the sixth grade English language textbook for Saudi boys' schools* (Master's thesis). Retrieved from http://faculty.ksu.edu.sa/amri/Documents/MA%20thesis.pdf

Alanazi, M. A. (2013). *Using electronic writing to promote students' writing* (Master's thesis). Retrieved from http://dc.ewu.edu/cgi/viewcontent.cgi?article=1115&context=theses

Al Asmari, A. (2013). Saudi university undergraduates' language learning attitudes: a preparatory year perspective. *International Journal of Asian Social Science*, *13*(11), 2288–2306.

Al-Asmari, A. M., & Rabb Khan, M. S. (2014). E-learning in Saudi Arabia: Past, present and future. *Near and Middle Eastern Journal of Research in Education*, *1*(2), 1–11. doi:10.5339/nmejre.2014.2

Alaugab, A. M. (2007). *Benefits, barriers, and attitudes of Saudi female faculty and students toward online learning in higher education* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global. (Oder No. 3258686)

Al-Dosari, H. (2011). Faculty members and students perceptions of E-learning in the English department: A project evaluation. *Journal of Social Sciences*, *7*(3), 391–407.

Alebaikan, R. A. (2012). The future of blended learning. *World Academy of Science, Engineering and Technology*, *63*, 484–488.

Alebaikan, R., & Salah Troudi, S. (2010a). Blended learning in Saudi universities: Challenges and perspectives. *ALT-J: Research in Learning Technology*, *18*(1), 49–59.

Alebaikan, R., & Salah Troudi, S. (2010b). Online discussion in blended courses at Saudi Universities. *Procedia Social and Behavioral Sciences*, *2*, 507–514.

Alenezi, A. R., Abdul Karim, M. A., & Veloo, A. (2010). An empirical investigation into the role of enjoyment, computer anxiety, computer self-efficacy and internet experience in influencing the students' intention to use e-learning: A case study from Saudi Arabian governmental universities. *TOJET: The Turkish Online Journal of Educational Technology*, *9*(4). 22–34.

Al Fadda, H. (2012). Difficulties in academic writing: From the perspective of King Saud University postgraduate students. *English Language Teaching*, *5*(3), 123–130.

Al Fadda, H., Almasri, M., & ALShumaimeri, Y. (2011). A preliminary study of the effect Of WebQuests on the writing performance of Saudi female EFL elementary school students. *The JALT CALL journal*, *7*(3), 373–390.

Al-Fahad, F. N. (2009). Students' attitudes and perceptions towards the effectiveness of mobile learning in King Saud University, Saudi Arabia. *The Turkish Online Journal of Educational Technology - TOJET*, *8*(2). Retrieved from http://files.eric.ed.gov/fulltext/ED505940.pdf

Alghamidi, A. (2015, January 7). تدريب 8700 معلم لغة إنجليزية [Training 8700 EFL teachers]. Okadh Newspaper. Retrieved from http://www.okaz.com.sa/new/mobile/20150107/Con20150107745435.htm

Alghizzi, T. M. (2011). *The role of English writing instruction methodologies on the types of written mistakes/errors EFL graduate diploma students can identify in their writings* (Unpublished graduate diploma thesis). Dublin International Foundation College, Dublin, Ireland.

Alghizzi, T. M. (2012). *The role of English writing instruction methodologies on the types of written mistakes/errors Saudi EFL pre-university students can identify in their writings* (Unpublished Master's thesis). University College Cork, Cork, Ireland.

Alhajailan, T. (2015). The history of teaching English in the Kingdom of Saudi Arabia. Retrieved from http://hajailan.blog.com/2009/01/15/%D8%AA%D8%A7%D8%B1%D9%8A%D8%AE%D8D9%85%D9%85%D9%84%D9%83/

Al-Hammadi, F. (2010). The impact of multimedia on critical thinking and writing of Saudi secondary school students. *Information Technology Journal*, *9*(1), 11–19.

Al-Hammadi, F. S. (2011). Investigating the interaction between cognitive complexity and amount of L2 exposure in Saudi college students in the English department at King Faisal University. *Journal of Applied Sciences*, 1–6. doi:10.3923/jas.2011. Retrieved from http://docsdrive.com/pdfs/ansinet/jas/0000/32603-32603.pdf

Alharbi, H. A. (2015). Improving students' English speaking proficiency in Saudi public schools. *International Journal of Instruction*, *8*(1), 105–117.

Al-Harbi, K. A. (2011). E-Learning in the Saudi tertiary education: Potential and challenges. *Applied Computing and Informatics*, *9*(1), 31–46. doi:10.1016/j.aci.2010.03.002

Al-Harthi, A. S. (2005). Distance higher education experiences of Arab Gulf students in the United States: A cultural perspective. *The International Review of Research in Open and Distributed Learning*, *6*(3). Retrieved from http://www.irrodl.org/index.php/irrodl/article/viewArticle/263/406

Alhassan, R. A., & Gashan, A. K. A. (2013). Exploring WebQuests in EFL Writing Classrooms. 1–20. Retrieved from http://s3.amazonaws.com/academia.edu.documents/38920954/The_Effects_of_Using_WebQuests_on_the_Writing_Performance_of_Saudi_Male_English_as_Foreign_Language_Learners.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1466027711&Signature=Fvm3jscsKmguRuja8cOSJzJx5wo%3D&response-content-disposition=inline%3B%20filename%3DExploring_WebQuests_in_EFL_Writing_Clas s.pdf

Al-Hazmi, S. (2003). EFL teacher preparation programs in Saudi Arabia: Trends and challenges. *TESOL in the Gulf*, *37*(2), 341–344.

Al-Hazmi, S. (2006). Writing and reflection: Perceptions of Arab EFL learners. *South Asian Language Review*, *XVI*(2), 36–52.

Al-Hazmi, S., & P. Scholfield (2007). Enforced revision with checklist and peer feedback in EFL writing: The example of Saudi university students. *Scientific Journal of King Faisal University (Humanities and Management Sciences)*, *8*(2), 237–267.

Alhumaidi, M. (2013, November, 8). تطبيقات إلى الورق حيز من...السعودية في الإنجليزية اللغة تدريس [Teaching المقبل العام طالب ملايين أربعة يستهدف إلكتروني مشروع لتطبيق جارية تحضيرات :التقنية English language in Saudi Arabia…from the paper-based phase to electronic-based phase: In preparation for the application of the electronic project targeting four million

students next year]. Asharq Al-Awsat Newspaper. Retrieved from
http://aawsat.com/home/article/8730

Al-Humaidi, M. M. (2008). *Planning in the writings of Saudi EFL learners at the College of Languages and Translation at KSU* (Master's thesis). Retrieved from
http://repository.ksu.edu.sa/jspui/bitstream/123456789/19289/1/%d8%a7%d9%84%d8%b1%d8%b3%d8%a7%d9%84%d8%a9.pdf

Ali, S. H., Sait, S. M, & Al-Tawil, K. M. (2003). Perceptions about e-learning in Saudi Arabia. Retrieved from
https://www.researchgate.net/profile/Sadiq_Sait/publication/258119425_Perceptions_about_eLearning_in_Saudi_Arabia/links/00b4952868caf1a323000000.pdf

Aljabre, A. (2012). An exploration of distance learning in Saudi Arabian universities: Current practices and future possibilities. *International Journal of Business, Humanities and Technology*, *2*(2), 132–137.

Al-Jamhoor, A. (2001). A cross-cultural analysis of written discourse of Arabic-speaking learners of English. *Journal of King Saud University (Language and Translation)*, *13*, 25–44.

Al-Jarf, R. S. (2002). Effect of online learning on struggling ESL college writers. Retrieved from http://files.eric.ed.gov/fulltext/ED475920.pdf

Al-Jarf, R. S. (2004a). The effect of web-based learning on struggling EFL college writers. *Foreign Language Annals*, *37*(1), 49–56.

Al-Jarf, R. S. (2004b, November). *Differential effects of online instruction on a variety of EFL courses*. Paper presented at the 3rd annual meeting of the Asian Association of Computer Assisted Language Learning (AsiaCALL) at Universiti Sains Malaysia, Penang, Malaysia.

Al-Jarf, R. S. (2006). Cross-cultural communities: Saudi, Ukrainian, and Russian students online. *The Asian EFL Journal Quarterly*, *8*(2), 7–32.

Al-Jarf, R. S. (2007a). Online instruction and creative writing by Saudi EFL freshman students. *The Asian EFL Journal Quarterly*, *Professional Teaching Articles*, *22*. Retrieved from http://asian-efl-journal.com/1838/teaching-articles/2007/08/online-instruction-and-creative-writing-by-saudi-efl-freshman-students/

Al-Jarf, R. S. (2007b). Impact of blended learning on EFL college readers. Retrieved from http://repository.ksu.edu.sa/jspui/bitstream/123456789/5861/1/Impact%20of%20Blended%20Learning%20on%20EFL%20College%20Readers.pdf

Al-Jarf, R. (2008). The impact of English as an international language (EIL) upon Arabic in Saudi Arabia. *Asian EFL Journal*, *10*(4), 193–210.

Al-Jarf, R. (2009a, April). *Enhancing freshman students' writing skills with a mind mapping software*. Paper presented at the 5th International Scientific Conference, eLearning and Software for Education, Bucharest, Romania.

Al-Jarf, R. (2009b). *Phonological and orthographic problems in EFL college spellers*. Paper presented at TELLIS Conference, Azad Islamic University-Roudehen, Iran.

Al-Jarf, R. (2013). Enhancing freshman students' performance with online reading and writing activities. In I. Roceanu (Ed.), *Quality and Efficiency in eLearning: Proceedings of the 9th International Scientific Conference "eLearning and Software for Education"* (Vol. 2, pp. 524–530). doi:10.12753/2066-026X-13-193

Aljumah, F. H. (2012). Saudi learner perceptions and attitudes towards the use of blogs in teaching English writing course for EFL majors at Qassim University. *English Language Teaching*, *5*(1), 100–116. doi:10.5539/elt.v5n1p100

Al-Khairy, M. A. (2013). Saudi English-major undergraduates' academic writing problems: A Taif University perspective. *English Language Teaching*, *6*(6), 1–12.

Al-Khalifa, H. S. (2009, October). The state of distance education in Saudi Arabia. *The eLearn Magazine: Where Thought and Practice Meet*. Retrieved from http://elearnmag.acm.org/archive.cfm?aid=1642193

Al-Khasawneh, F. M. S. (2010). Writing for academic purposes: Problems faced by Arab postgraduate students of the college of business, UUM. *ESP World*, *9*(2), 1–23.

Alkhatnai, M. (2011). *Learning styles of EFL Saudi college-level students in on-line and traditional educational environments* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global. (Oder No. 3453565)

Alkhawaldeh, A. (2011). The contribution of reading comprehension to writing skill development in English among university students in Jordan. *European Journal of Social Science*, *25*(2), 260–274.

Alkubaidi, M. A. (2014). The relationship between Saudi English major university students' writing performance and their learning style and strategy use. *English Language Teaching*, *7*(4), 83–95. doi:10.5539/elt.v7n4p83

Allen, M., Omori, K., Burrell, N., Mabry, E., & Timmerman, E. (2013). Satisfaction with distance education. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 143–154) New York, USA: Routledge Taylor & Francis Group.

Al-Mahrooqi, R., & Troudi, S. (Eds.). (2014). *Using technology in foreign language teaching*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Al-Mansour, N. S., & Al-Shorman, R. A. (2014). The effect of an extensive reading program on the writing performance of Saudi EFL university students. *International Journal of Linguistics*, *6*(2), 258–275. doi:10.5296/ijl.v6i2.5551

Almelhi, A. M. (2014). Effects of teaching argumentative reading and writing by integration in an e-learning environment on literacy development in EFL college students. *International Journal of Humanities and Social Science*, *4*(5), 85–102.

Al-Menei, A. M. O. (2008). *An investigation of the effect of computer-assisted writing instruction on EFL Saudi learners' ability* (Master's thesis). Retrieved from http://repository.ksu.edu.sa/jspui/bitstream/123456789/19275/1/Thesis%20by%20Ahmed%20M.%20O.%20Al-Menei.pdf

Al-Nofaie, H. (2010). The attitudes of teachers and students towards using Arabic in EFL classrooms in Saudi public schools—a case study. *Novitas-ROYAL (Research on Youth and Language)*, *4*(1), 64–95.

Al Noursi, O. (2013). Attitude towards learning English: The case of the UAE technological high school. *Educational Research*, *4*(1), 21–30.

Alnufaie, M., & Grenfell, M. (2012). EFL students' writing strategies in Saudi Arabian ESP writing classes: Perspectives on learning strategies in self-access language learning. *Studies in Self-Access Learning Journal*, *3*(4), 407–422.

Alodwan, T. A. A., & Ibnian, S. S. K. (2014). The effect of using the process approach to writing on developing university students' essay writing skills in EFL. *Review of Arts and Humanities*, *3*(2), 139–155.

Alp, P., Kerge, K., & Pajupuu, H. (2013). Measuring lexical proficiency in L2 creative writing. In J. Colpaert, M. Simons, A. Aerts, & M. Oberhofer (Eds.), *Language Testing in Europe: Time for a New Framework* (pp. 274–286).

Alpala, C. A. O., & Flórez, E. E. R. (2011). Blended Learning in the teaching of English as a foreign language: An educational challenge. *HOW*, *18*(1), 154–168.

Al-Quran, M. (2010). Concept-based grammatical errors of Arab EFL learners. *Journal of Instructional Psychology*, *37*(1), 3–12.

Alqurashi, F. (2015). Perspectives of Saudi EFL learners towards teacher response in writing courses. *International Journal of English Linguistics*, *5*(5), 37–46. doi:10.5539/ijel.v5n5p37

Alrashidi, O., & Phan, H. (2015). Education context and English teaching and learning in the Kingdom of Saudi Arabia: An overview. *English Language Teaching*, *8*(5), 33–44. doi:10.5539/elt.v8n5p33

Alresheed, S. (2008). EFL program of the secondary schools in Qassim region (Saudi Arabia): Problems, causes and solutions. *Newcastle University Journal*. Retrieved from http://faculty.mu.edu.sa/public/uploads/1357382794.1003EFL%20P%20R%20Q%20%D8%A8%D8%AD%D8%AB%20%D9%83%D8%A7%D9%85%D9%84%20PdF.pdf

Alsaleem, B. I. A. (2014). The effect of "WhatsApp" electronic dialogue journaling on improving writing vocabulary word choice and voice of EFL undergraduate Saudi students. *Proceedings of the 21st Century Academic Forum Conference at Harvard University*, *USA*, *1*(1), 32–47.

Alsamadani, H. A. (2010). The relationship between Saudi EFL students' writing competence, L1 writing proficiency, and self-regulation. *European Journal of Social Science*, *16*(1), 53–63.

Al Samadani, H. A., & Ibnian, S. S. (2015). The relationship between Saudi EFL students' attitudes towards learning English and their academic achievement. *International Journal of Education and Social Science*, *2*(1), 92–102.

Al-Seghayer, K. (2011). *English teaching in Saudi Arabia: Status, issues, and challenges*. Riyadh, Saudi Arabia: Hala Printed Co.

Al-Seghayer, K. (2014a). The four most common constraints affecting English teaching in Saudi Arabia. *International Journal of English Linguistics*, *4*(5), 17–26. doi:10.5539/ijel.v4n5p17

Al-Seghayer, K. (2014b). The actuality, inefficiency, and needs of EFL teacher preparation programs in Saudi Arabia. *International Journal of Applied Linguistics and English Literature*, *3*(1), 143–151. doi:10.7575/aiac.ijalel.v.3n.1p.143

Al-Seghayer, K. (2015). Salient key features of actual English instructional practices in Saudi Arabia. *English Language Teaching*, *8*(6), 89–99. doi:10.5539/elt.v8n6p89

Al-Shaer, I. M. R. (2013). Effects of a blended learning module on EFL students' attitudes in an introductory reading course in Al-quds Open University setting. *International Journal of Language Learning and Applied Linguistics World (IJLLALW)*, *3*(4), 224–242.

Alshahrani, A., & Storch, N. (2014). Investigating teachers' written corrective feedback practices in a Saudi EFL context: How do they align with their beliefs, institutional

guidelines, and students' preferences? *Australian Review of Applied Linguistics*, *37*(2), 101–122.

Al-Shehri, A. M. (2010). E-learning in Saudi Arabia: 'To E or not to E, that is the question.' *Journal of Family and Community Medicine*, *17*(3), 147–150.

Alshehri, E. (2012). Saudi EFL teachers' and students' perceptions of motivational strategies. In K. Donnelly & F. Formato (Eds.), *Papers from the 7th Lancaster University Postgraduate Conference in Linguistics & Language Teaching 2012: Vol. 7.* (pp. 26–54).

Alshumaimeri, Y. (2011). The effects of wikis on foreign language students writing performance. *Procedia-Social and Behavioral Sciences*, *28*, 755–763. doi:10.1016/j.sbspro.2011.11.139

Alsulayyi, M. N. (2015).The use of grammatical collocations by advanced Saudi EFL learners in the UK and KSA. *International Journal of English Linguistics*, *5*(1), 32–43. doi:10.5539/ijel.v5n1p32

Alsunbul, A. (2002). Issues relating to distance education in the Arab world. *Convergence*, *XXXV*(1), 59–80.

Alswaier, A., Youssef, A., & Emam, A. (2012). A new trend for e-learning in KSA using educational clouds. *Advanced Computing: An International Journal (ACIJ)*, *3*(1), 81–97. doi:10.5121/acij.2012.3107

Altameem, A. (2013). What drives successful e-learning? An empirical investigation of the key technical issues in Saudi Arabian universities. *Journal of Theoretical and Applied Information Technology*, *53*(1), 63–70.

Al-Tamimi, A., & Shuib, M. (2009). Motivation and attitudes towards learning English: A study of petroleum engineering undergraduates at Hadhramout University of Sciences and Technology. *GEMA Online Journal of Language Studies*, *9*(2), 29–55.

Altowjry, A. (2005). *Reforming higher education in Saudi Arabia: The use of telecommunications technology* (Master's thesis). Retrieved from http://scholarworks.rit.edu/cgi/viewcontent.cgi?article=1459&context=theses

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

Alzu'bi, M. A. M., & Sabha, M. R. N. (2013). Using mobile-based email for English foreign language learners. *TOJET: The Turkish Online Journal of Educational Technology*, *12*(1), 178–186.

Al Zumor, A. W. Q., Al Refaai, I. K., Bader Eddin, E. A., & Aziz Al-Rahman, F. H. (2013). EFL students' perceptions of a blended learning environment: Advantages, limitations and suggestions for improvement. *English Language Teaching*, *6*(10), 95–110, doi:10.5539/elt.v6n10p95

Anker, S. (2010). *Real writing with readings: Paragraphs and essays for college, work, and everyday life* (5th ed.). Boston: Bedford/St. Martin's.

Anson, C. M. (1999). Distant voices: Teaching and writing in a culture of technology. *College English*, *61*(3), 261–280.

Arafeh, S. (2004). *The implications of information and communications technologies for distance education: Looking toward the future* (Final Report No. P11913). Arlington, VA: SRI International.

Armstrong, K. M. (2010). Fluency, accuracy, and complexity in graded and ungraded writing. *Foreign Language Annals*, *43*(4), 690–702.

Arnaud, P. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate component tests. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 133–145). London: Macmillan Academic and Professional LTD.

Arnold, N., & Paulus, T. (2010). Using a social networking site for experiential learning: Appropriating, lurking, modeling and community building. *Internet and Higher Education*, *13*, 188–196. doi:10.1016/j.iheduc.2010.04.002

Badawi, M. F. (2009, March). *Using blended learning for enhancing EFL prospective teachers' pedagogical knowledge and performance*. Paper presented at the Learning & Language - The Spirit of the Age, Ain Shams University Guest House, Cairo, Egypt.

Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, *54*(2), 153–160.

Bailey, N., Madden, C., & Krashen, S. D. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning*, *24*(2), 235–243.

Bañados, E. (2006). A blended-learning pedagogical model for teaching and learning EFL successfully through an online interactive multimedia environment. *CALICO Journal*, *23*(3), 533–550.

Bani-Khaled, T. A. A. (2014).The role of English as perceived by student of Applied English at the University of Jordan. *European Scientific Journal*, *10*(5), 400–420.

Baralt, M., Gilabert, R., & Robinson, P. (2014). An introduction to theory and research in task sequencing and instructed second language learning. In M. Baralt, R. Gilabert, & P.

Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 1–34). London: Bloomsbury Academic.

Bardovi-Harlig, K. & Bofman, T. (1988, March). *A second look at T-unit analysis*. Paper presented at the annual meeting of the Teachers of English to Speakers of Other Languages, Chicago, IL.

Barkaoui, K. (2007). Teaching writing to second language learners: Insights from theory and research. *TESL Reporter*, *40*(1), 35–48.

Barnett, M. A. (1989). Writing as a process. *The French Review*, *63*(1), 31–44.

Baró, À. L., & Serrano, R. S. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Española de Lingüística Aplicada*, (24), 95–110.

Bataineh, R. F. (2005). Jordanian undergraduate EFL students' errors in the use of the indefinite article. *Asian EFL Journal*, *7*(1), 56–76.

Bayat, N. (2014). The effect of the process writing approach on writing success and Anxiety. *Educational Sciences: Theory & Practice*,*14*(3), 1133–1141. doi:10.12738/estp.2014.3.17

Becker, A. (2011). Examining rubrics used to measure writing performance in U.S. intensive English programs. *The CATESOL Journal*, *22*(1) 113–130.

Bendania, A. (2011). Instructors' and learners' attitudes toward teaching and learning online: King Fahd University of Petroleum and Minerals (KFUPM) (Saudi Arabia) case study. *International Journal of Arts & Sciences*, *4*(8), 223–241.

Bijeikienė, V., Rašinskienė, S., & Zutkienė, L. (2011). Teachers' attitudes towards the use of blended learning in general English classroom. *Studies About Languages*, (18), 122–127.

Black, L. M. (2013). A history of scholarship. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 3–20) New York, USA: Routledge Taylor & Francis Group.

Blackmore-Squires, S. (2010). *An investigation into the use of a blog as a tool to improve writing in the second language classroom* (Master's thesis). Available from The Asian EFL Journal.

Blakelock, J., & Smith, T. E. (2006). Distance learning: From multiple snapshots, a composite portrait. *Computers and Composition*, *23*(1), 139–161. doi:10.1016/j.compcom.2005.12.008

Bonzo, J. D. (2008). To assign a topic or not: Observing fluency and complexity in intermediate foreign language writing. *Foreign Language Annals*, *41*(4), 722.

Breland, H. M. (1983). *The direct assessment of writing skill: A measurement review* (College Board Report No. 83–6). New York: College Entrance Examination Board.

Breland, H. M., & Jones, R. J. (1982). *Perceptions of writing skills* (Report No. 82–4). New York: College Entrance Examination Board. Retrieved from http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-1982-4-perceptions-writing-skill.pdf

Breuch, L. -A. M. K. (2011). Post-process "pedagogy:" A philosophical exercise. In V. Villanueva & K. L. Arola (Eds.), *Cross-talk in comp theory: A reader* (3rd ed., pp. 97–125). Urbana, Illinois: National Council of Teachers of English.

Brumfit, C. J. (1979). Communicative language teaching: An educational perspective. In C. J. Brumfit, & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 183–191). Oxford: Oxford University Press.

Bruton, D. L., & Kirby, D. R. (1987). Research in the classroom: Written fluency: Didn't we do that last year? *The English Journal*, *76*(7), 89–92.

Btoosh, M. A., & Taweel, A. Q. (2011). Contrastive rhetoric: Inflation, verbal voices and polyphonic visibility in learners and native speakers' academic writing. *Asian EFL Journal*, *13*(3), 205– 228.

Buckley, J. (1997). The invisible audience and the disembodied voice: Online teaching and the loss of body image. *Computers and Composition*, *14*(2), 179–187.

Buckley, M. H. (1986). When teachers decide to integrate the language arts. *Language Arts*, *63*(4), 369–377.

Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–40). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Bulté, B., & Housen, A. (2015). Evaluating short-term changes in L2 complexity development. *Círculo de Lingüística Aplicada a la Comunicación*, *63*, 42–76.

Bulté, B., Housen, A., Pierrard, M., & Van Daele, S. (2008). Investigating lexical proficiency development over time – the case of Dutch-speaking learners of French in Brussels. *Journal of French Language Studies*, *18*, 277–298. doi:10.1017/S0959269508003451

Byrne, D. (1993). *Teaching writing skills: Longman handbooks for language teachers* (new edition). Essex, England: Longman Group UK Limited.

Celaya, M. L., Torras, M. R., & Pérez-Vidal, C. (2001). Short and mid-term effects of an earlier start: An analysis of EFL written production. In S. H. Foster-Cohen & A.

Nizegorodcew (Eds.), *EUROSLA yearbook* (Vol. 1, pp. 195–209). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Chambers, F. (1997). What do we mean by oral fluency? *System*, *25*, 535–44.

Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge, UK: Cambridge University Press.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, *18*(1), 65–81.

Chen, L. -L. (2006). The effect of the use of L1 in a multimedia tutorial on grammar learning: An error analysis of Taiwanese beginning EFL learners' English essays. *The Asian EFL Journal Quarterly*, *8*(2), 76–110.

Chen, M., & Zechner, K. (2011). Computing and evaluating syntactic complexity features for automated scoring of spontaneous non-native speech. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, *Portland*, *Oregon*, 722–731.

Chiang, C. S., & Dunkel, P. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly*, *26*(2), 345–374.

Chiramanee, T., & Kulprasit, W. (2014). Journal writing with peer feedback: A friend or a foe for EFL learners. *International Journal of English Language Education*, *2*(2), 142–153. doi:10.5296/ijele.v2i2.6038

Cho, H. (2015). Effects of task complexity on English argumentative writing. *English Teaching*, *70*(2), 107–131. doi:10.15858/engtea.70.2.201506.107

Chrichton, S., & Kinash, S. (2013). Enabling learning for disabled students. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 216–230) New York, USA: Routledge Taylor & Francis Group.

Chu, R. (2011). Effects of teacher's corrective feedback on accuracy in the oral English of English-majors college students. *Theory and Practice in Language Studies*, *1*(5), 454–459.

Clause. (2015). In *Merriam-Webster.com*. Retrieved from http://www.merriam-webster.com/dictionary/clause

Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (Vol, 10, pp. 15–38). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Conti, G. (2015, August 10). Crucial issues in the assessment of speaking and writing (Part 1) [Web long post]. Retrieved from https://gianfrancoconti.wordpress.com/2015/08/10/ten-crucial-issues-in-the-assessment-of-speaking-and-writing-part-1/

Cope, B., & Kalantzis, M. (2012). Introduction: How a genre approach to literacy can transform the way writing is taught. In B. Cope & M. Kalantiz (Eds.), *The powers of literacy: A genre approach to teaching writing* (pp. 1–21). New York, USA: Routledge Library Editions: Education.

Coulmas, F. (2003). *Writing systems: An introduction to their linguistic analysis*. Cambridge, UK: Cambridge University Press.

Covington, M. A., & McFall, J. D. (2010) Cutting the Gordian knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, *17*(2), 94–100. doi:10.1080/09296171003643098

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, California: SAGE Publications, Inc.

Cronogue, C. E. A. (1985). *Oral and written language of low level ESL students* (Master's thesis) Available from Retrospective Theses and Dissertation. (Paper No. 151)

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics*, *11*(2), 183–199.

Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, *18*(2), 119–135. doi:10.1016/j.jslw.2009.02.002

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, *26*, 66–79.

Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, *59*(2), 307–334.

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, *45*(1), 182–193. doi: 10.5054/tq.2010.244019

Crossley, S., Salsbury, T., Titak, A., & McNamara, D. (2014). Frequency effects and second language lexical acquisition: Word types, word tokens, and word production. *International Journal of Corpus Linguistics*, *19*(3), 301–332.

Crowhurst, M. (1990). Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education*, *15*(4), 348–359.

Crystal, D. (1987). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.

Crystal, D. (2005). *The stories of English*. London: Penguin Books.

Cumming, A. (2002). Assessing L2 writing: Alternative constructs and ethical dilemmas. *Assessing Writing*, *8*, 73–83.

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology*, *100*(4), 907–919.

Dahmardeh, M., & Shahmirzadi, N. (2016). Measuring the written performance quality in terms of (CAF) complexity, accuracy and fluency constructs. *Social Sciences and Humanities*, *24*(2), 639–654.

Dalir, M., Jafarigohar, M., & Soleimani, H. (2013). The effect of using weblogs on Iranian EFL learners writing skills. *International Journal of Language Learning and Applied Linguistics World (IJLLALW)*, *3*(4), 213–223.

Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, *24*, 197–222.

Dashtestani, R. (2014). English as a foreign language—teachers' perspectives on implementing online instruction in the Iranian EFL context. *Research in Learning Technology*, 22. Retrieved from http://www.researchinlearningtechnology.net/index.php/rlt/article/view/20142

Davies, A., Hamp-Lyons, L., & Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes*, *22*(4), 571–584.

Deane, P. (2014). *Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks* (ETS Research Report No. RR 14–03). Princeton, NJ: ETS.

Delialioglu, O., & Yildirim, Z. (2007). Students' perceptions on effective dimensions of interactive learning in a blended learning environment. *Educational Technology & Society*, *10*(2), 133–146.

Deneme, S. (2011). A lesson model for enhancing motivation in EFL writing classes. *Theory and Practice in Language Studies*, *1*(7), 785–788.

Dennen, V. P. (2013). Activity design and instruction in online learning. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 282–298) New York, USA: Routledge Taylor & Francis Group.

Deveci, C. (2015). Teachers' and students' awareness level of their roles in distance education. *International Journal on New Trends in Education and their Implications (IJONTE)*, *6*(3), 10–17.

Dewaele, J. -M. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics 19*(4), 471–490.

Dickinson, P. (2014). The effect of topic-selection control on EFL writing fluency. 新潟国際情報大学情報文化学部紀要, *17*, 15–25.

Diehl, W. C. (2013). University correspondence study: A revised historiographic perspective. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 21–37) New York, USA: Routledge Taylor & Francis Group.

Dobao, A. F. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, *21*(1), 40–58.

Dokhykh, R. A. (2014). Students' perceptions of benefits and challenges in EFL distance learning in Saudi Arabia: A case study [term paper]. Retrieved from https://beautiful2as0a1rose4.files.wordpress.com/2014/12/students-perceptions-of-benefits-and-challenges-in-efl-distance-learning-in-saudi-arabia-a-case-study-term-paper-raihanah-dokhykh.pdf

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford, UK: Oxford University Press.

Doró, K. (2014). The lexical richness of undergraduate student essays: The influence of writing experience and essay writing strategies. *AFinLA-e Soveltavan Kielitieteen tutkimuksia*, (6), 7–19. Retrieved from http://ojs.tsv.fi/index.php/afinla/article/download/46277/12360

Driscoll, M. (2002). Blended learning: Let's get beyond the hype. *IBM Global Services*, 1–3. Retrieved from https://www-07.ibm.com/services/pdf/blended_learning.pdf

Dubiner, D., Freed, B., & Segalowitz, N. (2006). Native speakers' perceptions of fluency acquired by study abroad students and their implications for the classroom at home. In S. Wilkinson (Ed.), *Insights from study abroad for language programs* (pp. 2–21). Boston: Thompson Heinle.

Dudley-Evans, T. (1997). Genre models for the teaching of academic writing to second language speakers: Advantages and disadvantages. In T. Miller (Ed.), *Functional approaches to written text: Classroom applications* (pp. 150–159). Washington, DC: United States Information Agency.

Dudley-Evans, T. (2002). The teaching of the academic essay: Is a genre approach possible? In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 225–236). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*(2), 245–258.

Dulay, H. C., & Burt, M. K. (1974a). Natural sequences in child second language acquisition. *Language Learning 24*(1), 37–53.

Dulay, H. C., & Burt, M. K. (1974b). Errors and strategies in child second language acquisition. *TESOL Quarterly 8*(2), 129–136.

Dulay, H. C., & Burt, M. K. (1975). A new approach to discovering universal strategies of child second language acquisition. In D. P. Dato (Ed.), *Developmental psycholinguistics: Theory and application* (*Georgetown University Round Table on Languages and Linguistics)* (pp. 209–233). Washington, D. C.: Georgetown University Press.

Durán, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, *25*(2), 220–242.

Dušková, L. (1969). On sources of errors in foreign language learning. *IRAL-International Review of Applied Linguistics in Language Teaching*, *7*(1), 11–36.

Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer?, *Language Testing*, *19*(4), 347-368.

El Ebaikan, R., & Troudi, S. (2014). Blended learning in Saudi Arabia: Potential for its use in EFL at the tertiary level. In R. Al-Mahrooqi & S. Troudi (Ed.), *Using technology in foreign language teaching* (pp. 30–44). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Ellis, R. (1994). *The study of second language acquisition*. Oxford, UK: Oxford University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, *30*(4), 474–509. doi:10.1093/applin/amp042

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

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition (SSLA)*, *26*(01), 59–84.

Elola, I., & Oskoz, A. (2010). Collaborative writing: Fostering foreign language and writing conventions development. *Language Learning & Technology*, *14*(3), 51–71.

Elyas, T., & Picard, M. (2010). Saudi Arabian educational history: Impacts on English language teaching. *Education, Business and Society: Contemporary Middle Eastern Issues*, *3*(2), 136–145.

Emig, J. (2011). Writing as a mode of learning. In V. Villanueva & K. L. Arola (Eds.), *Cross-talk in comp theory: A reader* (3rd ed., pp. 7–15). Urbana, Illinois: National Council of Teachers of English.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, *4*(2), 139–155.

Erkan, D. Y. (2013). Effect of cross-cultural e-mail exchange on self-efficacy in EFL writing. *Ç.Ü. Sosyal Bilimler Enstitüsü* Dergisi, *22*(1), 25–42.

Error. (2010). In *Oxford Advanced Learner's Dictionary* (8th ed., International Student's Edition, p. 1345). Oxford: Oxford University Press.

Eydelman, N. (2013). A blended English as a foreign language academic writing course. In B. Tomlinson & C. Whittaker (Eds.), *Blended learning in English language teaching: Course design and implementation* (pp. 43–50). London, UK: British Council 2013 Brand and Design.

Ezza, E. (2010). Arab EFL learners' writing dilemma at tertiary level. *Journal of English Language Teaching*, *3*(4), 33–39.

Fageeh, A. I. (2011). EFL learners' use of blogging for developing writing skills and enhancing attitudes towards English learning: An exploratory study. *Journal of Language and Literature*, *2*(1), 31–48.

Faqeih, H. I. (2015). Learners' Attitudes towards corrective feedback. *Procedia-Social and Behavioral Sciences*, *192*, 664–671. doi: 10.1016/j.sbspro.2015.06.101

Farahani, A. A. K., & Meraji, S. R. (2011). Cognitive task complexity and L2 narrative writing performance. *Journal of Language Teaching and Research*, *2*(2), 445–456.

Farooq, M. S., Uzair-UI-Hassan, M., & Wahid, S. (2012).Opinion of second language learners about writing difficulties in English language. *South Asian Studies*, *27*(1),183–194.

Faruk, S. (2013). English language teaching in Saudi Arabia: A world system perspective. *Scientific Bulletin of the Politehnica University of Timişoara Transactions on Modern Languages*, *12*(2-2), 73–80

Faruk, S. M. G. (2014a). English textbooks and the objectives of ELT in Saudi Arabia: Gaps and rationale. *Scientific Bulletin of the Politehnica University of Timişoara Transactions on Modern Languages*, *13*(1), 47–56.

Faruk, S. M. G. (2014b). Saudis' attitudes towards English: Trends and rationale. *Professional Communication and Translation Studies*, *7*(1-2), 173–180.

Fazilatfar, A. M., Fallah, N., Hamavandi, M., & Rostamian, M. (2014). The effect of unfocused written corrective feedback on syntactic and lexical complexity of L2 writing. *Procedia-Social and Behavioral Sciences*, *98*, 482–488.

Fellner, T., & Apple, M. (2006). Developing writing fluency and lexical complexity with blogs. *The JALT CALL Journal*, *2*(1), 15–26.

Ferrari, S. (2012). A longitudinal study of complexity, accuracy and fluency variation in second language development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 277–297). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Ferreira, D. (2013). Researching the effect of students' self-selected topics on writing fluency. フェリス女学院大学文学部紀要, *48*, 297–306.

Ferris, D. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing*, *16*(3), 165–193. doi:10.1016/j.jslw.2007.07.003

Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, *22*(3), 307–329.

Finn, A. & Bucceri, M. (2006). *A case study approach to blended learning*. Los Angeles: Centra Software.

Fischer, R. A. (1984). Testing written communicative competence in French. *The Modern Language Journal*, *68*(1), 13–20.

Fitze, M. (2006). Discourse and participation in ESL face-to-face and written electronic conferences. *Language Learning & Technology*, *10*(1), 67–86.

Flower, L., & Hayes, J. R. (2011). A cognitive process theory of writing. In V. Villanueva & K. L. Arola (Eds.), *Cross-talk in comp theory: A reader* (3rd ed., pp. 253–277). Urbana, Illinois: National Council of Teachers of English.

Folse, K. S., Muchmore-Vokoun, A., & Solomon, E. V. (2010). *Great writing 2: Great paragraphs* (International student 3rd ed.). Boston, USA: Heinle Cengage Learning.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*, 299–323.

Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, *3*, 215–247. doi:10.1177/136216889900300303

Franken, M., & S. Haslett, 2002. When and why talking can make writing harder. In S. Ransdell & M. -L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 208–229). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing*, *30*, 45–57.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(02), 275–301.

Freed, B., So, S., & Lazar, N. A. (2003). Language learning abroad: How do gains in written fluency compare with gains in oral fluency in French as a second language? *ADFL Bulletin*, *34*(3), 34–40.

Gaies, S. J. (1980). T-Unit analysis in second language research: Applications, problems and limitations. *TESOL Quarterly*, *14*(1), 53–60.

Gallavan, N. P., Bowles, F. A., & Young, C. T. (2007). Learning to write and writing to learn: Insights from teacher candidates. *Action in Teacher Education*, *29*(2), 61–70.

Gallick-Jackson, S. A. (1997). *Improving narrative writing skills, composition skills, and related attitudes among second grade students by integrating word processing, graphic organizers, and art into a process approach to writing*. (Master's thesis). Available from ERIC database. (ERIC No. ED420064).

Garrison, D. R., & Akyol, Z. (2013). The community of inquiry theoretical framework. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 104–119) New York, USA: Routledge Taylor & Francis Group.

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *Internet and Higher Education*, *7*, 95–105. doi:10.1016/j.iheduc.2004.02.001

Gascoigne, C. (2004). Examining the effect of feedback in beginning L2 composition. *Foreign Language Annals*, *37*(1), 71–76.

Gaudelli, W. (2006). Convergence of technology and diversity: Experiences of two beginning teachers in web-based distance learning for global/multicultural education. *Teacher Education Quarterly*, *33*(1), 97–116.

Ghahari, S., & Ameri-Golestan, A. (2014). The effect of blended learning vs. classroom learning techniques on Iranian EFL learners' writing. *International Journal of Foreign Language Teaching & Research*, *1*(3). Retrieved from http://jfl.iaun.ac.ir/pdf_5591_174e9bc312bf98f092f1ddbd340ccf9d.html

Ghavamnia, M., Tavakoli, M., & Esteki, M. (2013). The effect of pre-task and online planning conditions on complexity, accuracy, and fluency on EFL learners written production. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (20), 31–43.

Ghorbani, M. R., Gangeraj, A. A., & Alavi, S. Z. (2013).Reciprocal teaching of comprehension strategies improves EFL learners' writing ability. *Current Issues in Education (CIE)*, *16*(1), 1–13.

Gillam, K., & Wooden, S. R. (2013). Re-embodying online composition: Ecologies of writing in unreal time and space. *Computers and Composition*, *30*(1), 24–36.

Glasgow, R., & Fitze, M. (2008). Input enhancement and tense formation in Arab EFL writing. In C. Coombe, A. Jendli, & P. Davidson (Eds.), *Teaching writing skills in English: Theory, research and pedagogy* (pp. 193–205). Dubai, United Arab Emirates: TESOL Arabia Publications.

Gleason, J. (2013). Dilemmas of blended language learning: Learner and teacher experiences. *CALICO Journal*, *30*(3), 323– 341. doi: 10.11139/cj.30.3.323-341

Grabe, W. (2002). Narrative and expository macro-genres. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 249–267). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Graham, C. R. (2013). Emerging practice and research in blended learning. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 333–350) New York, USA: Routledge Taylor & Francis Group.

Grami, M. (2005). The effect of teachers' written feedback on ESL students' perception: a study in a Saudi ESL university-level context. *Annual Review of Education, Communication and Language Sciences*, *2*. Retrieved from https://scholar.google.com/scholar?q=Grami+2005+feedback&btnG=&hl=en&as_sdt=0%2C5

Greener, S. L. (2008). Self-aware and self-directed: Student conceptions of blended learning. *MERLOT Journal of Online Learning and Teaching*, *4*(2), 243–253.

Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analyzing lexical density and lexical diversity in university students' written discourse. *Procedia-Social and Behavioral Sciences*, *198*, 546–556.

Grgurović, M. (2011). Blended learning in an ESL class: A case study. *CALICO Journal*, *29*(1), 100–117.

Guleff, V. (2002). Approaching genre: Prewriting as apprenticeship to communities of practice. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 211–224). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Gunnarsson, C. (2012). The development of complexity, accuracy and fluency in the written production of L2 French. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 247–276). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Ha, J. H. (2001). English-speaking students' written language development in Korean in a Korean/English two-way immersion program. In J. J. Ree (Ed.), *The Korean language in America* (Vol, 6, pp. 299–307). USA: the American Association of Teachers of Korean.

Hakuta, K. (1974). A preliminary report on the development of grammatical morphemes in a Japanese girl learning English as a second language. *Working Papers on Bilingualism*, *3*, 18–38.

Hakuta, K. (1975a). *Becoming bilingual at age five: the story of Uguisu*. (Unpublished doctoral dissertation). Harvard University, USA.

Hakuta, K. (1975b). Learning to speak a second language: What exactly does the child learn? In D. P. Dato (Ed.), *Developmental psycholinguistics: Theory and application* (*Georgetown University Round Table on Languages and Linguistics)* (pp. 193–207). Washington, D. C.: Georgetown University Press.

Hakuta, K. (1976). A case study of a Japanese child learning English. *Language Learning*, *26*, 321–351.

Hakuta, K. (1983). English language acquisition by speakers of Asian languages. In M. C. Chang (Ed.), *Comparative research in bilingual education: Asian Pacific American perspectives* (pp. 31–55). New York: Teachers College Press.

Halimi, S. S. (2008). Indonesian teachers' and students' preferences for error correction. *WACANA*, *10*(1), 50–71.

Halliday, M. A. K. (1994). Spoken and written modes of meaning. In D. Graddol & O. Boyd-Barret (Eds.), *Media texts: Authors and readers* (pp. 51–73). Clevedon, UK: Multilingual Matters LTD.

Halliday, M. A. K. (2009). Methods – techniques – problems. In M. A. K. Halliday & J. J. Webster (Eds.), *Continuum companion to systemic functional linguistics* (pp. 59–86). London: Continuum International Publishing Group.

Hamdan, A. K. (2014). The reciprocal and correlative relationship between learning culture and online education: A case from Saudi Arabia. *The International Review of Research in Open and Distributed Learning*, *15*(1). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1408/2773

Hammad, E. A. (2014). Palestinian university students' problems with EFL essay writing in an instructional setting. *Journal of Second and Multiple Language Acquisition (JSMULA)*, *2*(1), 1–21.

Hammill, M. J. (2014). *Second language writing in intensive English programs and first year composition* (Doctoral dissertation, Arizona State University). Retrieved from https://repository.asu.edu/attachments/137333/content/hammill_asu_0010e_14240.pdf

Hamouda, A. (2011). A study of students and teachers' preferences and attitudes towards correction of classroom written errors in Saudi EFL context. *English Language Teaching*, *4*(3), 128–141. doi:10.5539/elt.v4n3p128

Hamp-Lyons, L. (1995). Rating nonnative writing: The trouble with holistic scoring. *Teachers of English to Speakers of Other Languages (TESOL)*, *29*(4),759–762.

Hamp-Lyons, L., & Heasley, B. (2006). *Study writing: A course in written English for academic purposes* (2nd ed.). Cambridge: Cambridge University Press.

Harrington, M. (1986). The t-unit as a measure of JSL oral proficiency. *Descriptive and Applied Linguistics 19*, 49–56.

Hassan, M. K., & Akhand, M. M. (2010). Approaches to writing in EFL/ESL context: Balancing product and process in writing class at tertiary level. *Journal of NELTA*, *15*(1-2), 77–88.

Hastie, M., Hung, I. -C., Chen, N. -S., & Kinshuk (2010). A blended synchronous learning model for educational international collaboration. *Innovations in Education and Teaching International*, *47*(1), 9–24. doi: 10.1080/14703290903525812

Haswell, R. H. (2011). Researching teacher evaluation of second language writing via prototype theory. In P. K. Matsuda & T. Silva (Eds.), Second language writing research: Perspectives on the process of knowledge construction (pp. 105–120). New York: Routledge Taylor & Francis Group.

Hayati, A. M., & Gooran, M. (2014). Email and its effects on Iranian EFL learners' writing ability. *MEXTESOL Journal*, *38*(1), 1–14.

Hirano, K. (1989). Research in T-unit measures in ESL. *Bull. Jeotsu. Univ. Educ.*, *8*(2). 67–77. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&ved=0ahUKEwiJj-r1gZLPAhUJL8AKHV_lCn4QFghAMAU&url=https%3A%2F%2Fjuen.repo.nii.ac.jp%2Findex.php%3Faction%3Dpages_view_main%26active_action%3Drepository_action_common_download%26item_id%3D5433%26item_no%3D1%26attribute_id%3D22%26file_no%3D1%26page_id%3D13%26block_id%3D30&usg=AFQjCNEwJiu-Ul790oYM03O1RlQJlaofFA&cad=rja

Hong, K. H., & Samimy, K. K. (2010). The influence of L2 teachers' use of CALL modes on language learners' reactions to blended learning. *CALICO Journal*, *27*(2), 328–348.

Honsa J. S. (2013). Self-assessment in EFL writing: A study of intermediate EFL students at a Thai university. *Voices in Asia Journal*, *1*(1), 34–57.

House, J. (2003). English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics*, *7*(4), 556–578.

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

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

Housen, A., Schoonjans, E., Janssens, S., Welcomme, A., Schoonheere, E., & Pierrard, M. (2011). Conceptualizing and measuring the impact of contextual factors in instructed SLA–the role of language prominence. *IRAL-International Review of Applied Linguistics in Language Teaching*, *49*(2), 83–112.

Howell, S. L., Saba, F., Lindsay, N. K., & Williams, P. B. (2004). Seven strategies for enabling faculty success in distance education. *The Internet and Higher Education*, *7*(1), 33–49. doi:10.1016/j.iheduc.2003.11.005

Hrastinski, S. (2008). Asynchronous and synchronous e-learning . *EDUCAUSE Quarterly*, 31(4), 51–55.

Huiting, Z. (2008). Assessing L2 learners' lexical richness and syntactic complexity: A longitudinal study [Word document]. Retrieved from http://english.njtech.edu.cn/myEditor/uploadfile/20140117123554657.doc

Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Roport No. 3). Champaign, IL: National Council of Teachers of English.

Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English*, *43*(7), 732–739.

Hunt, K. W. (1970a). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development (Serial No. 143)*, *35*(1), iii-iv+1–67.

Hunt, K. W. (1970b). Do sentences in the second language grow like those in the first?. *TESOL Quarterly*, *4*(3), 195–202.

Hurd, S. (2005). Autonomy and the distance language learner. In B. Holmberg, M. Shelley, & C. J. White (Eds.), *Distance education and languages: Evolution and change (new perspectives on language and education)* (pp. 1–19). Clevedon, UK: Multilingual Matters.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Language Learning*, *27*(2), 383–410.

Hyon, S. (2002). Genre and ESL reading: A classroom study. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 121–144). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Ibarrola, A. L. (2013). Reformulation and self-correction: Insights into correction strategies for EFL writing in a school context. *Vigo. International Journal of Applied Linguistics*, *29*, 29–49.

Idol, L., & Croll, V. J. (1987). Story-mapping training as a means of improving reading comprehension. *Learning Disability Quarterly*, *10*(3), 214–229. doi: 10.2307/1510494

Ilter, B. G. (2009). Effect of technology on motivation in EFL classrooms. *Turkish Online Journal of Distance Education-TOJDE*, *10*(4), 136–158.

Irele, M. E. (2013). Evaluating distance education in an era of internationalization. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 493–506) New York, USA: Routledge Taylor & Francis Group.

Irimiea, S. (2012). Assessment through forms of academic writing of business experts and translators. In A. Pelea (Ed.), *International review of studies in applied modern languages* (No, 5, pp. 305–317). Retrieved from http://193.231.20.24/rielma/RIELMA_no5_2012.pdf#page=306

Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *The Journal of Asia TEFL*, *3*(4), 193–225.

Ishikawa, T. (2007). The effect of manipulating task complexity along the [+/-Here-and-Now] dimension on L2 written narrative discourse. In M. D. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136–156). Clevedon, UK: Multilingual Matters LTD.

Ismail, L., Abd. Samad, A. A., Eng, W. B., & Noordin, N. (2012). Effects of task reasoning demand and task condition on learner written output in ESL classrooms. *International Journal of Education*, *4*(4), 157–175. doi:10.5296/ije.v4i4.2249

Iwashita, N. (2010). Features of oral proficiency in task performance by EFL and JFL learners. In M. T. Prior, Y. Watanabe, & S. -K. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices* (pp. 32–47). Somerville, MA: Cascadilla Proceedings Project.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*(1), 24–49.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition; A practical approach*. Rowley, MA: Newbury House.

Jafari, N., & Ansari, D. N. (2012). The effect of collaboration on Iranian EFL learners' writing accuracy. *International Education Studies*, *5*(2), 125–131. doi:10.5539/ies.v5n2p125

Jahin, J. H. (2012). The effect of peer reviewing on writing apprehension and essay writing ability of prospective EFL teachers. *Australian Journal of Teacher Education*, *37*(11), 60–84.

Jahin, J. H., & Idrees, M. W. (2012). EFL major student teachers' writing proficiency and attitudes towards learning English. *Umm Al-Qura University Journal of Educational & Psychologic Sciences*, *4*(1), 10–72.

Jamalinesari, A., Rahimi, F., Gowhary, H., & Azizifar, A. (2015). The effects of teacher-written direct vs. indirect feedback on students' writing. *Procedia-Social and Behavioral Sciences*, *192*, 116–123. doi: 10.1016/j.sbspro.2015.06.018

Janfaza, A., Shahsavari, K., & Soori, A. (2014). The impact of using email on improving the writing skills among Iranian students. *Advances in Language and Literary Studies*, *5*(5), 75–79. doi:10.7575/aiac.alls.v.5n.5p.75

Javid, C. Z., & Umer, M. (2014, March). *Saudi EFL learners' writing problems: A move towards solution*. Paper presented at the Global Summit on Education GSE, Kuala Lumpur, Malaysia.

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *Teachers of English to Speakers of Other Languages (TESOL Quarterly)*, *40*(1), 157–181.

Jiang, W. (2013). Measurements of development in L2 written production: The case of L2 Chinese. *Applied Linguistics*, *34*(1), 1–24. doi:10.1093/applin/ams019

Jiaxin, X. I. N. G. (2015). The effects of increasing task complexity on EFL learners' writing performance. *Studies in Literature and Language*, *11*(4), 34–39. doi:10.3968/7711

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers*, *53*, 61–79.

Johns, A. M. (1986). Coherence and academic writing: Some definitions and suggestions for teaching. *TESOL Quarterly*, *20*(2), 247–265.

Johns, A. M. (1990). L1 composition theories: Implications for developing theories of L2 composition. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 24–36). Cambridge, UK: Cambridge University Press.

Johns, A. M. (2002). Introduction: Genre in the classroom. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 3–13). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Johnson, G. M. (2006). Synchronous and asynchronous text-based CMC in educational contexts: A review of recent research. *TechTrends*, *50*(4), 46–53.

Johnson, M. C., & Graham, C. R. (2015). Current status and future directions of blended learning models. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (pp. 2470–2480). Hershey, PA: Idea Group Inc.

Jones, R. L., & Spolsky, B. (Eds.). (1975). *Testing language proficiency.* Virginia, USA: Center for Applied Linguistics.

Jong, Y. O. (2009). *An investigation into the benefits of collaborative writing for the development of EFL children's communication skills: A reflective report of a teacher researcher* (Doctoral dissertation). Retrieved from http://wrap.warwick.ac.uk/3743/1/WRAP_THESIS_Jong_2009.pdf

Kanuka, H., & Conrad, D. (2003). The name of the game: Why "distance education" says it all. *Quarterly Review of Distance Education*, *4*(4), 385–393.

Kaya, M. (2012). Distance education systems used in universities of Turkey and Northern Cyprus. *Procedia Social and Behavioral Sciences*, *31*, 676–680. doi:10.1016/j.sbspro.2011.12.123

Kellogg, R. T. (2006). Professional writing expertise. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 389–402). New York: Cambridge University Press.

Khan, A. I., Noor-ul-Quyyum, Shaik, M. S., Ali, A. M., & Bebi, C. V. (2012). Study of blended learning process in education context. *I.J.Modern Education and Computer Science*, *9*, 23–29. doi:10.5815/ijmecs.2012.09.03

Khan, I. A. (2011a). Learning difficulties in English: Diagnosis and pedagogy in Saudi Arabia. *International Research Journals*, *2*(7), 1248–1257.

Khan, I. A. (2011b). Challenges of teaching/learning English and management. Global *Journal of Human Social Science*, *11*(8), 68–80.

Khongput, S. (2010, August). *EFL writing assessment practices: Teachers' perspectives*. Paper presented at the 36th International Association for Educational Assessment (IAEA) Annual Conference, Bangkok, Thailand.

Khorasani, R., Pandian, A., & Ismail, S. A. M. M. (2012). The effect of strategic planning on accuracy, fluency, and complexity of written narrative task production. *Language in India*, *12*(7), 422–446.

Kim, J. -Y. (2014). Predicting L2 writing proficiency using linguistic complexity measures: A corpus-based study. *English Teaching*, *69*(4), 27–51.

Kim, M. C., Kozan, K., Kim, W., & Koehler, A. A. (2013). Technology integration: From implementation to dynamic scaffolding. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 299–315) New York, USA: Routledge Taylor & Francis Group.

Kitchakarn, O. (2014). Developing writing abilities of EFL students through blogging. *Turkish Online Journal of Distance Education*, *15*(4), 34–47.

Klein, H. J., Noe, R. A., & Wang, C. (2006). Motivation to learn and course outcomes: The impact of delivery mode, learning goal orientation, and perceived barriers and enablers. *Personnel Psychology*, *59*, 665–702.

Klímová, B. F. (2008). Blended learning and teaching foreign languages. *Problems of Education in the 21st Century*, *5*, 69–74.

Knoch, U. (2008). Diagnostic writing ability: A rating scale for accuracy, fluency and complexity. *New Zealand Studies in Applied Linguistics*, *14*(2), 1–24.

Kobayashi, K., & Little, A. (2011). Learner perceptions on the usefulness of a blended learning EFL program. *The JALT CALL Journal*, *7*(1), 103–117.

Koizumi, R. (2012). Relationships between text length and lexical diversity measures: can we use short texts of less than 100 tokens. *Vocabulary Learning and Instruction*, *1*(1), 60–69. doi:10.7820/vli.v01.1.koizumi

Koizumi, R., & In'nami, Y. (2014). Modeling complexity, accuracy, and fluency of Japanese learners of English: A structural equation modeling approach. *JALT Journal: The Asian Association for Language Teaching*, *36*(1), 25–46.

Kong, J., Powers, S., Starr, L., & Williams, N. (2012). Connecting English language learning and academic performance: A prediction study. *American Educational Research Association*, 1–22. Retrieved from http://images.pearsonassessments.com/images/tmrs/EnglishLanguageLearningandAcademicPerformanc_AERA.pdf

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). Ann Arbor, Michigan: University of Michigan Press.

Kowal, I. (2014). Fluency in second language writing: A developmental perspective. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, *131*, 229–246. doi:10.4467/20834624SL.14.013.2321

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

Kuiken, F., & Vedder, I. (2007a). Cognitive task complexity and linguistic performance in French L2 writing. In M. D. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 117–135). Clevedon, UK: Multilingual Matters LTD.

Kuiken, F., & Vedder, I. (2007b). Task complexity and measures of linguistic performance in L2 writing. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(3), 261–284.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, *17*(1), 48-60. doi:10.1016/j.jslw.2007.08.003

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

Kuiken, F., Mos, M. & Vedder, I. (2005). Cognitive task complexity and second language writing performance. In S. H. Foster-Cohen, M. D. P. G. Mayo, & J. Cenoz (Eds.), *EUROSLA yearbook* (Vol. 5, pp. 195–222). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research* (EUROSLA Monographs series, 1, pp. 81–99). European Second Language Association.

Larsen, L. J. E. (2012). *Teacher and student perspectives on a blended learning intensive English program writing course* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global. (Oder No. 3511431)

Larsen-Freeman, D. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly*, *9*, 409–419.

Larsen-Freeman, D. (1978a). Evidence of the need for a second language acquisition index of development. In W. C. Ritchie (Ed.), *Second language acquisition research: Issues and implications* (pp. 127–136). New York: Academic Press.

Larsen-Freeman, D. (1978b). An ESL index of development. *TESOL Quarterly*, *12*(4), 439–448.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Seliger & M. H. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 287–306). Rowley, Massachusetts: Newbury House Publishers, Inc.

Larsen-Freeman, D. (1990). *Language teaching methods: Teacher's handbook for the video series*. Retrieved from https://americanenglish.state.gov/files/ae/resource_files/language_teaching_methods_teachers_handbook.pdf

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

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 579–589. doi:10.1093/applin/amp043

Larsen-Freeman, D., & Long, M. H. (1997). *An introduction to second language acquisition research*. New York, USA: Addison Wesley Longman Limited.

Larsen-Freeman, D., & Storm, V. (1977). The construction of a second language acquisition index of development. *Language Learning*, *27*(1), 123–134.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–22.

Leech, G., Rayson, P., & Wilson, A. (2014). *Word frequencies in written and spoken English: Based on the British national corpus*. New York, USA: Routledge.

Lemmouh, Z. (2008). The relationship between grades and the lexical richness of student essays. *Nordic Journal of English Studies*, *7*(3), 163–180.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Michigan, USA: The University of Michigan Press.

Levy, M. C., & Ransdell, S. (1995). Is writing as difficult as it seems? *Memory & Cognition*, *23*(6), 767–779.

Li, W. (2013). A study of assessment mechanism and training approach of college English writing through sharing & improving online learning community. *Research Inventy: International Journal Of Engineering And Science*, *3*(12), 1–5.

Li, Y. (2000). Linguistic characteristics of ESL writing in task-based e-mail activities. *System*, *28*(2), 229–245.

Liang, M. Y., & Bonk, C. J. (2009). Interaction in blended EFL learning: Principles and practice. *International Journal of Instructional Technology and Distance Learning*, *6*(1), 3–16.

Liaw, M. L., & Johnson, R. J. (2001). E-mail writing as a cross-cultural learning experience. *System*, *29*(2), 235–251.

Lin, H. (2002). The effect of e-mail project on enhancing EFL learners' academic writing skills. In M. Driscoll & T. Reeves (Eds.), *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2002* (pp. 1810–1813). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Lindqvist, C., Gudmundson, A., & Bardel, C. (2013). A new approach to measuring lexical sophistication in L2 oral production. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (EUROSLA Monographs series, 2, 109–126). European Second Language Association.

Lintunen, P., & Mäkilä, M. (2014). Measuring syntactic complexity in spoken and written learner language: Comparing the incomparable? *Research in Language*, *12*(4), 377–399. doi:10.1515/rela-2015-0005

Lipinski, T. A. (2013). Legal and recent copyright issues. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 533–549) New York, USA: Routledge Taylor & Francis Group.

Liss, R., & Davis, J. (2012). *The researched essay: Effective academic writing* (2nd ed.). New York, USA: Oxford University Press.

Liton, H. A. (2012). Developing EFL teaching and learning practices in Saudi colleges: A review. *International Journal of Instruction*, *5*(2), 129–152.

Liton, H. A. (2013). EFL teachers' perceptions, evaluations and expectations about English language courses as EFL in Saudi universities. *International Journal of Instruction*, *6*(2), 19–34.

Liu, J., & Sadler, R. W. (2003). The effect and affect of peer review in electronic versus traditional modes on L2 writing. *Journal of English for Academic Purposes*, *2*(3), 193–227. doi:10.1016/S1475-1585(03)00025-0

Liu, M. (2013). Blended learning in a university EFL writing course: Description and evaluation. *Journal of Language Teaching and Research*, *4*(2), 301–309. doi:10.4304/jltr.4.2.301-309

LoCoco, V. (1975). An analysis of Spanish and German learner's errors. *Working Papers on Bilingualism*, *7*, 96–124.

Long, R., & Tabuki, M. (2014). Comparing EFL learners and native speaker fluency. In H. Y. -H. Lee (Ed.), *Proceedings of the International Conference on Language and Communication: Innovative Inquiries and Emerging Paradigms in Language, Media*

and Communication (pp. 14–35). Bangkok, Thailand: The Graduate School of Language and Communication and The National Institute of Development Administration.

Lord, G. (2009). Second-language awareness and development during study abroad: A case study. *Hispania*, *92*(1), 127–141.

Lorenzo, F., & Rodríguez, L. (2014). Onset and expansion of L2 cognitive academic language proficiency in bilingual settings: CALP in CLIL. *System*, *47*, 64–72.

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *SSLA*, *16*, 303–323.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. 1–28. Retrieved from http://personal.psu.edu/xxl13/papers/Lu_inpress_ijcl.pdf

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. 1–26. Retrieved from http://www.personal.psu.edu/xxl13/papers/Lu_inpress_mlj.pdf

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43. doi:10.1016/j.jslw.2008.06.002

Macken-Horarik, M. (2002). "Something to shoot for": A systemic functional approach to teaching genre in secondary school science. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 17–42). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Mahboob, A., & Elyas, T. (2014). English in the Kingdom of Saudi Arabia. *World Englishes*, *33*(1), 128–142.

Mahfoodh, O. H. A., & Pandian, A. (2011). A qualitative case study of EFL students' affective reactions to and perceptions of their teachers' written feedback. *English Language Teaching*, *4*(3), 14–25.

Mahmoud, A. (2011). The role of interlingual and intralingual transfer in learner-centered EFL vocabulary instruction. *Arab World English Journal*, *2*(3), 28–49.

Mahmoud, S. A. (2014). Email and Facebook to promote foundation year students' EFL writing at King Abdul-Aziz University. *International Review of Social Sciences and Humanities*, *6*(2), 157–172.

Mahmoud, S. S., & Oraby, K. K. (2015). Let Them Toil to Learn: Implicit feedback, self-correction and performance in EFL writing. *Theory and Practice in Language Studies*, *5*(8), 1672–1681. doi: http://dx.doi.org/10.17507/tpls.0508.18

Maloney, K. B., & Hopkins, B. L. (1973). The modification of sentence structure and its relationship to subjective judgements of creativity in writing. *Journal of Applied Behavior* Analysis, *6*, 425–433.

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

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills, England: Palgrave MacMillan.

Marsh, D. (2012). *Blended learning creating learning opportunities for language learners*. Cambridge: Cambridge University Press.

Mart, C. T. (2013) The grammar-translation method and the use of translation to facilitate learning in ESL classes. *Journal of Advances in English Language Teaching*, *1*(4), 103–105.

Masrom, U. K., Alwi, N. A. N. M., & Daud, N. S. M. (2015). The effects of task complexity on the complexity of the second language written production. *Journal of Second Language Teaching & Research*, *4*(1), 38–66.

Mata, H. M. (2003). *The clauses in a pear tree: Syntactic complexity in the narratives of Mexican women of different educational levels* (Doctoral dissertation). Retrieved from http://athenaeum.libs.uga.edu/xmlui/handle/10724/21110

Matsuda, P. K. (2003). Basic writing and second language writers: Toward an inclusive definition. *Journal of Basic Writing*, *22*(2), 67–89.

Matsuda, P. K. (2011). Composing studies and ESL writing: A disciplinary division of labor. In V. Villanueva & K. L. Arola (Eds.), *Cross-talk in comp theory: A reader* (3rd ed., pp. 673–696). Urbana, Illinois: National Council of Teachers of English.

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for academic purposes programme. *Journal of Second Language Writing*, *29*, 3–15.

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Global. (Oder No. 3199485)

McCarthy, P., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, *24*, 459–488.

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. doi:10.3758/BRM.42.2.381

McCarthy, S. J., Guo, Y. -H., & Cummins, S. (2005). Understanding changes in elementary Mandarin students' L1 and L2 writing. *Journal of Second Language Writing*, *14*, 71–104. doi:10.1016/j.jslw.2005.05.003

McCarthy, T. (1998). *Descriptive writing*. New York, US: Scholastic Professional books.

McGivern, Y. (2009). *The practice of market research* (3ʳᵈ ed.). Gosport: Ashford Colour Press LTD.

McLaughlin, B. (1987). *Theories of second-language learning*. London, Great Britain: Edward Arnold.

McMullen, M. G. (2009). Using language learning strategies to improve the writing skills of Saudi EFL students: Will it really work? *ScienceDirect*, *37*, 418–433.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1) 57–86.

Mehrotra, C. M., Hollister, C. D., & McGahey, L. (2001). *Distance learning: Principles for effective design, delivery, and evaluation*. Thousand oaks, CA: Sage Publications.

Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisitions*, *3*(2), 109–135.

Mellor, A. (2006). Using graphs to estimate lexical diversity. *四日市大学総合政策学部論集*, *5*(1), 185–195. Retrieved from http://ci.nii.ac.jp/els/110006161209.pdf?id=ART0008128874&type=pdf&lang=en&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1475231659&cp=

Miao, Y., Badger, R., & Zhen, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing*, *15*(3), 179–200. doi:10.1016/j.jslw.2006.09.004

Mihaljević Djigunović, J., & Letica Krevelj, S. (2011). Lexical diversity in early learning of English as L2. *SRAZ LVI*, 253–270. Retrieved from http://hrcak.srce.hr/file/150133

Miller, S. K. (2001). A review of research on distance education in computers and composition. *Computers and Composition*, *18*(4), 423–430.

Ministry eyes ways to utilize newly hired English teachers. (2002, August). Arab News. Retrieved from http://www.arabnews.com/node/223593

Miniwatts Marketing Group (2016, September 12). Internet world stats usage and population statistics: Middle east. Retrieved from http://www.internetworldstats.com/

Miralpeix, I. (2006). Age and vocabulary acquisition in English as a foreign language (EFL). In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 89–106). Clevedon, UK: Multilingual Matters LTD.

Mirici, I. H. (2003). The factors affecting the success in English proficiency exams and possible contributions of the internet. *Turkish Online Journal of Distance Education - TOJDE*, *4*(1), 1–8.

Mirza, A. A. (2007). Is e-learning finally gaining legitimacy in Saudi Arabia? *Saudi Computer Journal*, *6*(2), 1–14.

Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2nd ed.). London: Hodder Arnold.

Miyazoe, T., & Anderson, T. (2010). Learning outcomes and students' perceptions of online writing: Simultaneous implementation of a forum, blog, and wiki in an EFL blended learning setting. *System*, *38*(2), 185–199. doi:10.1016/j.system.2010.03.006

Mohammadzadeh Mohammadabadi, A., Dabaghi, A., & Tavakoli, M. (2013). The effects of simultaneous use of pre-planning along+/-Here-and-Now dimension on fluency, complexity, and accuracy of Iranian EFL learners' written performance. *International Journal of Research Studies in Language Learning*, *2*(3), 49–65. doi:10.5861/ijrsll.2012.168

Mohar, T., Sraka, D., & Kaučič, B. (2012). Analyzing blended based learning systems. *Central European Conference on Information and Intelligent System*, 107–113. Retrieved from http://www.ceciis.foi.hr/app/public/conferences/1/papers2012/eis4.pdf

Mojavezi, A. (2014). The relationship between task repetition and language proficiency. *Applied Research on English Language*, *3*(1), 29-40.

Mongillo, G., & Wilder, H. (2012). An examination of at-risk college freshmen's expository literacy skills using interactive online writing activities. *Journal of College Reading and Learning*, *42*(2), 27–50.

Montague, N. (1995). The process oriented approach to teaching writing to second language learners. *New York State Association for Bilingual Education Journal*, *10*, 13–24.

Moore, M. G. (2013). Design and teaching: An overview. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 247–250) New York, USA: Routledge Taylor & Francis Group.

Moore, M., & Kearsley, G. (2012). *Distance education: A system view of online learning* (3rd ed.). Belmont, USA: Wadsworth Cengage Learning.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, *46*(4), 610–641.

Moran, C. (2003). Computers and Composition 1983–2002: What we have hoped for. *Computers and Composition*, *20*, 343–358. doi:10.1016/j.compcom.2003.08.011

Morgan, B., Fuisting, B., & White, J. (2014). University student attitudes towards peer review in EFL writing: A quantitative study. *Language Education in Asia*, *5*(1), 93–116.

Mota, M. B. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Fragmentos: Revista de Línguae Literatura Estrangeiras*, *24*, 69–104.

Mourtaga, K. R. (2010). Poor writing in English: A case of the Palestinian EFL learners in Gaza strip. Retrieved from http://www.qou.edu/english/conferences/firstNationalConference/pdfFiles/kamalMourtraga.pdf

Murray, D. M. (2011). Teach writing as a process not product. In V. Villanueva & K. L. Arola (Eds.), *Cross-talk in comp theory: A reader* (3rd ed., pp. 3–6). Urbana, Illinois: National Council of Teachers of English.

Mustafa, R. F. (2012). Feedback on the feedback: Sociocultural interpretation of Saudi ESL learners' opinions about writing feedback. *English Language Teaching*, *5*(3), 3–15. doi:10.5539/elt.v5n3p3

Mutta, M. (2006). The impact of intercultural context on lexical development - a case study of Finnish Erasmus students learning French. In F. Dervin & E. Suomela-Salmi (Eds.), *Intercultural communication and education: Finnish perspectives*, (pp.153–181). Berne: Peter Lang.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet rasch measurement: Part I. Journal of Applied Measurement, *4*(4), 386–422.

Naidu, S. (2013). Instructional design models for optimal learning. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 268–281) New York, USA: Routledge Taylor & Francis Group.

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, *14*(4), 397–419.

Navés, T., Torras, M. R., & Celaya, M. L. (2003). Long-term effects of an earlier start: An analysis of EFL written production. In S. H. Foster-Cohen & S. P. Doehler (Eds.),

EUROSLA yearbook (Vol. 3, pp. 103–129). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Niazi, F., & Pourgharib, B. (2013). The effect of using email on enhancing Iranian intermediate EFL learners writing proficiency. *International Research Journal of Applied and Basic Sciences (IRJABS)*, *4*(11), 3531–3539.

Nikou, F. R., & Eskandarsefat, Z. (2012). The simultaneous effects of task complexity and task types on EFL learners' written performance. *Australian Journal of Basic and Applied Sciences*, *6*(7),137–143.

Nordin, S. M., & Mohammad, N. B. (2006). The best of two approaches: Process/genre-based approach to teaching writing. *The English Teacher*, *XXXV*, 75–85.

Norris, J. M. (2006). Assessing advanced foreign language learning and learners: From measurement constructs to educational uses. In H. Byrnes, H. Weger-Guntharp, & K. A. Sprang (Eds.), *Educating for advanced foreign language capacities: Constructs, curriculum, instruction, assessment* (pp. 167–187).Washington, DC: Georgetown University Press.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578.

Nunberg, G., Briscoe, T., & Huddleston, R. (2002). Punctuation. In R. Huddleston & G. K. Pullum (Eds.), *The Cambridge grammar of the English language* (pp. 1724-1764). Cambridge: Cambridge University Press.

Obari, H. (2012). *The effect of blended learning in EFL*. Paper presented at the PC Conference. Retrieved from http://gakkai.univcoop.or.jp/pcc/2012/papers/pdf/pcc048.pdf

Obari, H., & Lambacher, S. (2014). Impact of a blended environment with m-learning on EFL skills. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds), *CALL Design: Principles and Practice*; *Proceedings of the 2014 EUROCALL Conference, Groningen, the Netherlands* (pp. 267–272). doi:10.14705/rpnet.2014.000229

Ocak, M. A. (2010). Blend or not to blend: A study investigating faculty members perceptions of blended teaching. *World Journal on Educational Technology*, *2*(3), 196–210.

Oh, S. (2006). *Investigating the relationship between fluency measures and second language writing placement test decisions* (Master's thesis, The University of Hawai'i). Retrieved from http://scholarspace.manoa.hawaii.edu/bitstream/handle/10125/20203/Oh2006.pdf?sequence=1

Oliver, M., & Trigwell, K. (2005). Can 'blended learning' be redeemed? *E–Learning*, *2*(1), 17–26.

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218–233.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*, 492-518.

Osman, W. H., & Abu Bakar, A. L. (2009). Learning to writing an academic paper among medical students of Universiti Malaysia Sabah. *Proceedings of the 2nd International Conference of Teaching and Learning (ICTL) INTI University College, Malaysia*, 1–21. Retrieved from https://my.laureate.net/Faculty/docs/Faculty%20Documents/INTI%20Conferences/Parallel%20Sessions%205/5C/5C-03-P48%20(Malaysia).pdf

Pae, J. -K. (2011). Collaborative writing versus individual writing: Fluency, accuracy, complexity, and essay score. *Multimedia-Assisted Language Learning*, *14*(1), 121–148.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590–601.

Palpanadan, S. A., Salam, A. R. B., & Ismail, F. B. (2014). Comparative analysis of process versus product approach of teaching writing in Malaysian schools: Review of literature. *Middle-East Journal of Scientific Research*, *22*(6), 789–795. doi: 10.5829/idosi.mejsr.2014.22.06.21943

Paltridge, B. (2002). Genre, text type, and the English for academic purposes (EAP) classroom. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp. 27–90). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Paltridge, B. (2014). Genre and second-language academic writing . *Language Teaching*, *47*(3), 303–318. doi:10.1017/S0261444814000068

Palvianen, Å., Kalaja, P., & Mäntylä, K. (2012) Development of L2 writing: Fluency and proficiency. Retrieved from http://scholar.google.com/scholar_url?url=http://ojs.tsv.fi/index.php/afinla/article/download/7037/5598&hl=en&sa=X&scisig=AAGBfm3auG8431fpVrq9zlQ1KRq9c12bwQ&nossl=1&oi=scholarr

Pandya, N. B. (2013). Chatting - an effective tool to improve second language learning. *Scholarly Research Journal of Humanity Science & English Language (SRJIS)*, *I*(VI), 1015–1022.

Pardo-Gonzalez, J. (2013). Incorporating blended learning in an undergraduate English course in Colombia. In B. Tomlinson & C. Whittaker (Eds.), *Blended learning in English language teaching: Course design and implementation* (pp. 51–60). London, UK: British Council 2013 Brand and Design.

Park, N. C., & Son, J. -B. (2009). Implementing computer-assisted language learning in the EFL classroom: Teachers' perceptions and perspectives. *International Journal of Pedagogies and Learning*, *5*(2), 80–101.

Park, T. (2008). Scoring procedures for assessing writing. Retrieved from http://tesol-dev.journals.cdrs.columbia.edu/wp-content/uploads/sites/12/2015/05/3.2-Park-2003.pdf

Peachey, N. (2013). A blended learning teacher development course for the development of blended learning in English language teaching. In B. Tomlinson & and C. Whittaker (Eds.), *Blended learning in English language teaching: Course design and implementation* (pp. 65–74). London, UK: British Council Brand and Design.

Pérez-Vidal, C., & Juan-Garau, M. (2009). The effect of study abroad (SA) on written performance. In L. Roberts, G. D. Veronique, A.-C. Nilsson, & M. Tellier (Eds.), *EUROSLA yearbook* (Vol. 9, pp. 269–295). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Peterson, S. S. (2010). Improving student writing: Using feedback as a teaching tool. *Literacy and Numeracy Secretariat Monograph*, *29*, 1–4. Retrieved from http://www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/WW_Improving_Student_Writing.pdf

Phakiti, A., & Paltridge, B. (2015). Approaches and methods in applied linguistics research. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (pp. 5–25). London, UK: Bloomsbury Academic.

Piri, F., Barati, H., & Ketabi, S. (2012). The effects of pre-task, on-line, and both pre-task and on-line planning on fluency, complexity, and accuracy - the case of Iranian EFL learners' written production. *English Language Teaching*, *5*(6), 158–167. doi:10.5539/elt.v5n6p158

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47(1), 101–143.

Polio, C. (2001). Research methodology in second language writing research: The case of text-based studies. In T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 91–115). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Pourdana, N., & Behbahani, S. M. K. (2011). Task types in EFL context: Accuracy, fluency, and complexity in assessing writing performance. *International Journal of Social Science and Humanity*, *2*(1), 80–84.

Pourdana, N., Behbahani, S. M. K., & Safdari, M. (2011). The impact of task types on aspects of Iranian EFL learners' writing performance: Accuracy, fluency, and complexity. *Proceedings of International Conference on Humanities, Society and Culture, Singapore*, *20,* 261–265.

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. P. (2008). Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology*, *17*, 107–120.

Qi, D. (2014). *Syntactic complexity of EFL, ESL and ENL: Evidence from the international corpus network of Asian learners of English* (Master's thesis, National University of Singapore). Retrieved from
http://scholarbank.nus.sg/bitstream/handle/10635/79491/DongQ.pdf?sequence=1

Rahimi, A., & Dastjerdi, H. V. (2012). Impact of immediate and delayed error correction on EFL learners' oral production: CAF. *Mediterranean Journal of Social Sciences*, *3*(1), 45–54. doi:10.5901/mjss.2012.03.01.45

Rahimpour, M. (2007). Task Complexity and Variation in L2 Learner's Oral Discourse. Retrieved from
https://www.library.uq.edu.au/ojs/index.php/uqwpl/article/viewFile/14/15

Rahimpour, M., & Hosseini, P. (2010). The impact of task complexity on L2 learners' written narratives. *English Language Teaching*, *3*(3), 198–205.

Rahman, M. M. U., & Alhaisoni, E. (2013). Teaching English in Saudi Arabia: Prospects and challenges. *Academic Research International*, *4*(1), 112–118.

Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. *TESOL Quarterly*, *25*(3), 407–430.

Rajab, H. (2013). Developing speaking and writing skills of L1 Arabic EFL learners through teaching of IPA phonetic codes. *Theory and Practice in Language Studies*, *3*(4), 653–659. doi:10.4304/tpls.3.4.653-659

Rana, A. M. K., & Perveen, U. (2013). Motivating students through self correction. *Educational Research International*, *2*(2), 192–196.

Ransdell, S., Arecco, R. M., & Levy, M. C. (2001). Bilingual long-term working memory: The effects of working memory loads on writing quality and fluency. *Applied Psycholinguistics*, *22*, 113–128

Rausch, A. S. (2012) Complexity, accuracy and fluency: Toward a conceptual model of communicative and sociolinguistic frameworks towards a conceptual model of communicative and sociolinguistic frameworks. *Bull. Fac. Educ. Hirosaki University*, *108*, 19–30.

Read, J. (2000). *Assessing vocabulary*. Oxford, England: Oxford University Press.

Read, J., & Nation, P. (2002). *An investigation of the lexical dimension of the IELTS Speaking Test* (IELTS Research Reports Vol. 6). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.5271&rep=rep1&type=pdf

Reid, J. M. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, *21*(1), 87–111.

Reppen, R. (2002). A genre-based approach to content writing instruction. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 321–327). Cambridge, UK: Cambridge University Press.

Rezazadeh, M., Tavakoli, M., & Eslami Rasekh, A. (2011). The role of task type in foreign language written production: Focusing on fluency, complexity, and accuracy. *International Education Studies*, *4*(2), 169–176.

Richards, B., & Malvern, D. (2000, September). *Measuring vocabulary richness in teenage learners of French*. Paper presented at the British Educational Research Association Conference, Cardiff University, Wales, UK.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, *20*(1), 83–96.

Robinson, P. (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27–57.

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

Robinson, P. (2003). The cognition hypothesis, 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. *IRAL*, *43*,1–32.

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. D. P. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–26). Clevedon, UK: Multilingual Matters LTD.

Robinson, P. (2011a). Task-based language learning: A review issues. In P. Robinson (Ed.), *Task-based language learning* (pp. 1-36). Sussex, UK: Wiley-Blackwell Publishing.

Robinson, P. (2011b). 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. 3–37). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Robinson, P. (2015). The cognition hypothesis, second language task demands, and the SSARC model of pedagogic sequencing. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. 87–121). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Robinson, P. Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *IRAL*, *45*, 161–176. doi:10.1515/IRAL.2007.007

Roux, R., Guzmán, N. P. T., & González, E. F. (2014). Distance education for EFL teachers: Perceptions of learner support. *GIST Education and Learning Research Journal*, (9 JUL-DEC), 157–178.

Roy, D. (2012). Website analysis as a tool for task-based language learning and higher order thinking in an EFL context. *Computer Assisted Language Learning*, *27*(5), 1–27. doi:10.1080/09588221.2012.751549

Ruiz-Funes, M. (2014). An exploration of task complexity in foreign language writing at the intermediate level and measures of linguistic production. *Journal of Linguistics and Language Teaching*, *5*(2),181–206.

Saba, F. (2013). Building the future: A theoretical perspective. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 49–65) New York, USA: Routledge Taylor & Francis Group.

Saba, M. S. (2013). *Writing in a new environment: Saudi ESL students learning academic writing* (Doctoral dissertation, the Virginia Polytechnic Institute and State University). Retrieved from https://vtechworks.lib.vt.edu/bitstream/handle/10919/54012/Saba_MS_D_2014.pdf?sequence=1&isAllowed=y

Saddler, B., & Asaro-Saddler, K. (2013). Response to Intervention in writing: A suggested framework for screening, intervention, and progress monitoring. *Reading and Writing Quarterly*, *29*(1), 20–43. doi:10.1080/10573569.2013.741945

Şahin, A. (2010). Effects of jigsaw II technique on academic achievement and attitudes to written expression course. *Educational Research and Reviews*, *5*(12), 777–787.

Saito, H. (1994). Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal*, *11*(2), 46–70.

Sakuragi, T. (2011). The construct validity of the measures of complexity, accuracy, and fluency: Analyzing the speaking performance of learners of Japanese. *JALT Journal*, *33*(2), 157–174.

Salebi, M. Y. (2004). Saudi college students' perception of their errors in written English. *Scientific Journal of King Faisal University (Humanities and Management Sciences)*, *5*(2), 209–228.

Saleh, A. A., El-Bakry, H. M., Asfour, T. T., & Mastorakis, N. (2010, March). Adaptive e-learning tools for numbering systems. *Proceeding of 9th WSEAS International Conference on Applications of Computer Engineering (ACE'10), Penang, Malaysia.* Retrieved from https://www.researchgate.net/profile/Nikos_Mastorakis/publication/234816897_Adaptive_e-learning_tools_for_numbering_systems/links/54495a1b0cf244fe9ea33b2f.pdf

Salimi, A., & Dadashpour, S. (2012). Task complexity and language production dilemmas (Robinson's cognition hypothesis vs. Skehan's trade-off model). *Procedia-Social and Behavioral Sciences*, *46*, 643–652. doi:10.1016/j.sbspro.2012.05.177

Sample, E., & Michel, M. (2014). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, *31*(8), 23–46.

Sapp, D. A., & Simon, J. (2005). Comparing grades in online and face-to-face writing courses: Interpersonal accountability and institutional commitment. *Computers and Composition*, *22*, 471–489. doi:10.1016/j.compcom.2005.08.005

Sari, P. P., & Saun, S. (2013). Teaching writing by using the process-genre approach at junior high schools. *Journal of English Language Teaching*, *2*(1), 255–264.

Sasaki, M. (2004). A multiple-data analysis of the 3.5-year development of EFL student writers. *Language Learning*, *54*(3), 525–582.

Sasaki, M. (2007). Effects of study-abroad experiences on EFL writers: A multiple-data analysis. *The Modern Language Journal*, *91*(iv), 602–620.

Sasayama, S. (2011). Cognition hypothesis and second language performance composition of written and oral task performance. *Second Language Studies*, *29*(2), 107–129.

Sasayama, S. (2015). *Validating the assumed relationship between task design, cognitive complexity, and second language task performance* (Doctoral dissertation). Retrieved from

https://repository.library.georgetown.edu/bitstream/handle/10822/1029904/Sasayama_g eorgetown_0076D_13147.pdf?sequence=1

Sasayama, S. (2016). Is a 'complex' task really complex? validating the assumption of cognitive task complexity. *The Modern Language Journal*, *100*(1), 231–245. doi:10.1111/modl.12313

Saudi Electronic University. (2011). Establishment. Retrieved from https://www.seu.edu.sa/sites/en/Pages/main.aspx

Sawalmeh, M. H. M. (2013). Error analysis of written English essays: The case of students of the preparatory year program in Saudi Arabia. *English for Specific Purposes World*, *14*(40), 1–17.

Sayadi, F, & Khoshsima, H. (2016). The effect of virtual language learning method on writing ability of Iranian intermediate EFL learners. *Advances in Language and Literary Studies*, *7*(2), 192–202. doi:10.7575/aiac.alls.v.7n.2p.192

Schenker, T. (2015). Telecollaboration for novice language learners—negotiation of meaning in text chats between nonnative and native speakers. In E. Dixon & M. Thomas (Eds.), *Researching language learner interactions online: From social media to MOOCs* (Vol. 13, pp. 237–257). Texas, USA: the Computer Assisted Language Instruction Consortium (CALICO).

Schilderman, H. (2012). Quantitative method. In B. J. Miller-McLemore (Ed.), *Tile wiley-blackwell companion to practical theology* (pp. 123–132). Sussex, UK: Blackwell Publishing Limited.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *26*(02), 173–199.

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, *11*(2), 133–158.

Seidlhofer, B. (2005). Key concepts in ELT: English as a lingua franca. *ELT Journal*, *59*(4), 339–341. doi: 10.1093/elt/cci064

Selfe, C. L. (1999). Technology and literacy: A story about the perils of not paying attention. *College Composition and Communication*, *50*(3), 411–436.

Sentence. (2010). In *Oxford Advanced Learner's Dictionary* (8th ed., International Student's Edition, p. 1345). Oxford: Oxford University Press.

Seow, A. (2002). The writing process and process writing. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 315–320). Cambridge, UK: Cambridge University Press.

Serrano, R. S. (2011). The effect of program type and proficiency level on learners' written production. *Revista Española de Lingüística Aplicada (RESLA)*, (24), 211–226.

Seyyedi, K., Ismail, S. A. M. M., Orang, M., & Nejad, M. S. (2013). The effect of pre-task planning time on L2 learners' narrative writing performance. *English Language Teaching*, *6*(12), 1–10. doi:10.5539/elt.v6n12p1

Shang, H. -F. (2007). An exploratory study of e-mail application on FL writing performance. *Computer Assisted Language Learning*, *20*(1), 79–96.

Sharma, V. K. (2015).How do productive skills of Saudi students affect EFL learning and teaching. *Asian Journal of Humanities and Social Sciences (AJHSS)*, *3*(2), 91–99.

Shaykina, O. I. (2015). Blended learning in English language teaching: Open educational resources used for academic purposes in Tomsk Polytechnic University. *Mediterranean Journal of Social Sciences*, *6*(3 S5), 255–260. doi:10.5901/mjss.2015.v6n3s5p255

Shearer, R. L. (2013). Theory to practice in instructional design. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 251–267) New York, USA: Routledge Taylor & Francis Group.

Shen, L. -B., Tseng, C. -Y., Kuo, S. -W., Su, Y. -J., & Chen, M. -Y. (2005). A preliminary study of college students' out-of-class English learning activities. *Chia-Nan Annual Bulletin*, *31*, 464–475.

Shih, R. C. (2011). Can web 2.0 technology assist college students in learning English writing? Integrating Facebook and peer assessment with blended learning. *Australasian Journal of Educational Technology*, *27*(5), 829–845.

Shukri, N. A. (2014). Second language writing and culture: Issues and challenges from the Saudi learners' perspective. *Arab World English Journal (AWEJ)*, *5*(3) 190–207.

Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 11–23). Cambridge, UK: Cambridge University Press.

Simsim, M. T. (2011). Internet usage and user preferences in Saudi Arabia. *Journal of King Saud University-Engineering Sciences*, *23*(2), 101–107. doi:10.1016/j.jksues.2011.03.006

Singh, H. (2003). Building effective blended learning programs. *Educational Technology*, *43*(6), 51–54.

Singhal, M. (1997). The internet and foreign language education: Benefits and challenges. *The Internet TESL Journal*, *III*(6). Retrieved from http://iteslj.org/Articles/Singhal-Internet

Šišková, Z. (2012). Lexical richness in EFL students' narratives. *Language Studies Working Papers*, *4*, 26–36.

Skehan, P. (1996a). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp.17–30). Oxford: Heinemann.

Skehan, P. (1996b). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*(1), 38–62.

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

Skehan, P. (2003a). Task-based instruction. Lang. Teach., *36*, 1–14. doi:10.1017/S026144480200188X

Skehan, P. (2003b). Focus on from, tasks, and technology. *Computer Assisted Language Learning*, *16*(5), 391–411.

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

Skehan, P. (2013). Tasks and language performance assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp.167–185). Oxon, USA: Routledge Taylor and Francis Group.

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

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*(1), 93–120.

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

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp.193–216). Amsterdam, the Netherlands: John Benjamins Publishing Company.

Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency, and lexis in task-based performance: A synthesis of the earling research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 199–220). Amsterdam, The Netherlands: John Benjamins Publishing Company.

So, H. -J., & Bonk, C. J. (2010). Examining the roles of blended learning approaches in computer-supported collaborative learning (CSCL) environments: A Delphi study. *Educational Technology & Society*, *13*(3), 189–200.

Spelkova, I., & Hurst, N. (2008). Teachers' attitudes to skills and the writing process in Latvia and Portugal. *The APPI Journal*, *8*(1). Retrieved from https://scholar.google.com/scholar?q=Teachers%E2%80%99+Attitudes+to+Skills+and+the+Writing+Process+in+Latvia+and+Portugal.&btnG=&hl=en&as_sdt=0%2C5

Sponseller, A. C., & Wilkins, M. (2015). Investigating the impact of topic selection control on writing fluency. *Hiroshima Studies in Language and Language Education*, (18), 141–152.

Stadie, N., Schro, A., Postler, J. D. Lorenz, A., Swoboda-Moll, M., Burchert, F., & De Blese, R. (2008). Unambiguous generalization effects after treatment of non-canonical sentence production in German agrammatism. *Brain and Language*, *104*, 211–229. doi:10.1016/j.bandl.2007.08.006

Steele, G. E., & Thurmond, K. C. (2009). Academic advising in a virtual university. *New Directions for Higher Education*, *2009*(146), 85–95.

Stockwell, G., & Harrington, M. (2003). The incidental development of L2 proficiency in NS-NNS email interactions. *CALICO Journal*, *20*(2), 337–359.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of second language writing*, *14*(3), 153–173.

Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes*, *8*, 207–223. doi:10.1016/j.jeap.2009.03.001

Strömqvist, S., Johansson, V., Kriz, S., Ragnarsdóttir, H., Aisenman, R., & Ravid, D. (2002). Toward a cross-linguistic comparison of lexical quantain speech and writing. *Written Language and Literacy*, *5*(1), 45–67.

Sukamolson, S. (2012). Fundamentals of quantitative research. Retrieved from http://www.culi.chula.ac.th/Research/e-Journal/bod/Suphat%20Sukamolson.pdf

Sun, C., & Feng, G. (2009). Process approach to teaching writing applied in different teaching models. *English Language Teaching*, *2*(1), 150–155.

Šuster, S. (2011). Measuring lexical and syntactic complexity of the language production of the World-of-Warcraft gaming community. Retrieved from http://www.let.rug.nl/suster/unpublished/wow_complexity.pdf

Swarts, J., and L. Odell. (2001, October). *Rethinking the evaluation of writing in engineering courses*. Paper presented at the meeting of 31st ASEE/IEEE Frontiers in Education Conference, Reno, NV.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295–312.

Syed, Z. (2003). The sociocultural context of English language teaching in the Gulf. *TESOL in the Gulf*, *37*(2), 337–340.

Tahaineh, Y. S. Q. (2009). *A cross-sectional investigation of interlingual and intralingual errors made by EFL Arab Jordanian university students in the use of prepositions in their writing* (Doctoral dissertation, University of Malaysia). Retrieved from http://eprints.usm.my/28819/1/A_cross-sectional_investigation_of_interlingual_%26_intralingual_errors_made_by_efl_arab_jordanian_university_students_in_the_use_of_prepositions_in_their_writing.pdf

Tahaineh, Y. S. (2010). Arab EFL university students' errors in the use of prepositions. *Modern Journal of Applied Linguistics (MJAL)*, *2*(1), 76–112.

Tahaineh, Y., & Daana, H. (2013). Jordanian undergraduates' motivations and attitudes towards learning English in EFL context. *International Review of Social Sciences and Humanities*, *4*(2), 159–180.

Tangkiengsirisin, S. (2006). Approaches to teaching second language writing. *Language Institute Journal*, *3*, 1–26.

Tangpermpoon, T. (2008). Integrated approaches to improve students writing skills for English major students. *ABAC Journal*, *28*(2), 1–9.

Tanni, Z. A. I. (2015). Attitudes toward English among AL-Quds Open University students in Tulkarm Branch. *World Journal of Education*, *5*(3), 139–147. doi:10.5430/wje.v5n3p139

Tavakoli, M., & Rezazadeh, M. (2014). Individual and collaborative planning conditions: Effects on fluency, complexity and accuracy in L2 argumentative writing. *Journal of Teaching Language Skills*, *5*(4), 85–110.

Tavakoli, M., Dastjerdi, H. V., & Esteki, M. (2011). The effect of explicit strategy instruction on L2 oral production of Iranian intermediate EFL learners: Focusing on accuracy, fluency and complexity. *Journal of Language Teaching and Research*, *2*(5), 989–997.

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

Teng, D. C. -E., Chen, N. -S., Kinshuk, & Leo, T. (2012). Exploring students' learning experience in an international online research seminar in the synchronous cyber classroom. *Computers & Education*, *58*, 918–930. doi:10.1016/j.compedu.2011.10.018

The College of Languages and Translation at Al-Imam Muhammad Ibn Saud Islamic University. (2016). كلية اللغات والترجمة: لمحة عامة عن الكلية، كلمة العميد، الرؤية، الرسالة، الاهداف [The college of languages and translation: General background information about the college, speech of its dean, its vision, message, and objectives]. Retrieved from https://units.imamu.edu.sa/colleges/LanguageAndTranslation/profile/Pages/default.aspx

Thompson, J., & Hills, J. (2005). Online learning support services for distance education students: Responding to and maintaining the momentum. *Proceedings of the ASCILITE 2005: Balance, Fidelity, Mobility: Maintaining the Momentum? Brisbane*, *Australia*, 661–665.

Thompson, M. M., & Wrigglesworth, J. W. (2013). Students and teachers as ethical actors. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 403–418). New York, USA: Routledge Taylor & Francis Group.

Thornbury, S. (2000) Accuracy, fluency and complexity. *English Teaching Professional*, *16*, 3–6.

Tidball, F., & Treffers-Daller, J. (2008). Analyzing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies*, *18*(3), 299–314. doi:10.1017/S0959269508003463

Tong, W. (2007). Teach writing as an ongoing process: Tips for EFL learners on reviewing EFL composition. *US-China Foreign Language*, *5*(11), 53–56.

Torras, M. R., Navés, T., Celaya, M. L., & Pérez-Vidal, C. (2006). Age and IL development in writing. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 156–182). Clevedon, UK: Multilingual Matters LTD.

Torras, M. R., & Celaya, M. L. (2001). Age-related differences in the development of written production. An empirical study of EFL school learners'. *International Journal of English Studies*, *1*(2), 103–126.

Towell, R. (2012). Complexity, accuracy and fluency from the perspective of psycholinguistic second language acquisition research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 47–70). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: An exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Javis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp.79–104).Amsterdam, The Netherlands: John Benjamins Publishing Company.

Tsai, S. -C. (2013). EFL business writing with task-based learning approach: A case study of student strategies to overcome difficulties. *K.U.A.S. Journal of Humanities and Social* Sciences, *10*(2), 217–238.

Tweedie, F.J., & Baayen, R.H. (1998). How variable may a constant be? Measures in lexical richness in perspective. *Computers and the Humanities 32*, 323–352.

University of Southern California. (2016). Organizing your social sciences research paper: Academic writing style. Retrieved from http://libguides.usc.edu/writingguide/academicwriting

University of Technology Sydney. (2013). Academic writing. Retrieved from https://www.lib.uts.edu.au/sites/default/files/attachments/page/Academic%20Writing%20Guide%20Part%201%20-%20Academic%20Writing.pdf

Van der Zee, K., & Van Oudenhoven, J. P. (2000). The multicultural personality questionnaire: A multidimensional instrument of multicultural effectiveness. *European Journal of Personality*, *14*, 291–309.

Van der Zee, K., & Van Oudenhoven, J. P. (2001). The multicultural personality questionnaire: Reliability and validity of self- and other ratings of multicultural effectiveness. *Journal of Research in Personality*, *35*, 278–288.

Van der Zee, K., Van Oudenhoven, J. P., Ponterotto, J. G., & Fietzer, A. W. (2013). Multicultural personality questionnaire: Development of a short form. *Journal of Personality Assessment*, *95*(1), 118–124. doi:10.1080/00223891.2012.718302

Vann, R. J. (1979). Oral and written syntactic relationships in second language learning. In C. Yorio, K. Perkins, & J. Schachter (Eds.), *On TESOL '79: The learner in focus* (pp. 322–329). Washington, DC: TESOL.

Vaughan, N., & Garrison, R. (2006). A blended faculty community of inquiry: Linking leadership, course redesign, and evaluation. *Canadian Journal of University Continuing Education*, *32*(2), 67–92.

Vercellotti, M. L. (2012). *Complexity, accuracy, and fluency as properties of language performance: The development of the multiple subsystems over time and in relation to each other* (Doctoral dissertation, University of Pittsburgh). Retrieved from https://core.ac.uk/download/files/457/12211043.pdf

Verkroost, M. -J., Meijerink, L., Lintsen, H., & Veen, W. (2008). Finding a balance in dimensions of blended learning. *International Journal on E-Learning*, *7*(3), 499–522.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, *17*, 65–83.

Vermeer, A. (2004). The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (Vol, 10, pp. 173–189). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, *24*(2), 203–220. doi:10.1007/s11145-010-9266-7.

Wang, C. X., Calandra, B., & Yi, Y. (2010). Technical, pedagogical, and cultural considerations for language learning in MUVEs. In G. Veletsianos (Ed.), *Emerging technologies in distance education* (pp. 285–299). Alberta, Canada: AU Press, Athabasca University.

Wang, C. X., Lefaiver, M. L. M., Wang, Q., & Hunt, C. (2011). Teaching in an EFL program in second life: Student teachers' perspectives and implications. *Journal of Educational Technology Development and Exchange*, *4*(1), 27–40.

Wang, S., & Slater, T. (2016). Syntactic complexity of EFL Chinese students' writing. *English Language and Literature Studies*, *6*(1), 81–86. doi:10.5539/ells.v6n1p81

Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, *16*, 194–209. doi:10.1016/j.jslw.2007.07.004

West, M. (1953). *A general service list of English words: With semantic frequencies and a supplementary word-list for the writing of popular science and technology.* London: UK, Longman, Green & Co.

White, E. M. (1993). Assessing higher-order thinking and communication skills in college graduates writing. *The Journal of General Education*, *42*(2),105–122.

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing*, *26*(3), 445–466.

Wind, A. M. (2013). Second language writing development from a dynamic systems theory perspective. In B. -P. Olmos-López, J. Huang, & J. Almeida (Eds.), *Papers from the 8th Lancaster University Postgraduate Conference in Linguistics & Language Teaching 2012: Vol. 8. Papers from LAEL PG 2013* (pp. 90–123).

Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics to assess L2 writing. *Iranian Journal of Language Testing*, *2*(1), 59–92.

Wolcott, W. (2004). Brutal choices in curricular design. *Perspectives: Teaching Legal Research and Writing*, *13*(1), 5–9.

Wolfe, M. B. W., & Woodwyk, J. M. (2010). Processing and memory of information presented in narrative or expository texts. *British Journal of Educational Psychology*, *80*, 341–362. doi:10.1348/000709910X485700

Wolfe-Quintero, K., Inagaki, S., & Kim, H. -Y. (1988). *Second language development in writing: Measures of fluency, accuracy & complexity*. Hawai'i, USA: Second Language Teaching & Curriculum Center.

Writing Development Centre of Newcastle University. (2016). Features of academic writing. Retrieved from http://www.ncl.ac.uk/students/wdc/learning/academic/

Wu, W.-C. V., Yen, L. L., & Marek, M. (2011). Using online EFL interaction to increase confidence, motivation, and ability. *Educational Technology & Society*, *14*(3), 118–129.

Xiao, J. (2008). Paired peer review in a distance-taught EFL writing course. *The Journal of Asia TEFL*, *5*(4), 85–115.

Yamada, M. (2009). The role of social presence in learner-centered communicative language learning using synchronous computer-mediated communication: Experimental study. *Computers & Education*, *52*, 820–833. doi:10.1016/j.compedu.2008.12.007

Yamani, H. A. (2014). E-learning in Saudi Arabia challenges and opportunities. *Journal of Information Technology and Application in Education*, *3*(4), 169–172. doi: 10.14355/jitae.2014.0304.10

Yang, J. (2015). Teaching writing to EFL college students in the big data era. In P. Perry, M. Tlougan, M. Hylind, I. Solovjeva, A. Moore, M. Roemer, & Y. Zhang (Eds.), *International Conference on Arts, Design and Contemporary Education 2015: Vol. 23. Advances in Social Science, Education and Humanities Research* (pp. 776–779).

Yang, S. C., & Chen, Y. -J. (2007). Technology-enhanced language learning: A case study. *Computers in Human Behavior*, *23*(1), 860–879.

Yang, W. (2014). *Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of L2 writing* (Doctoral dissertation, Georgia State University). Retrieved from http://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1029&context=alesl_diss

Yi, J. -Y. (2009). Defining writing ability for classroom writing assessment in high schools. *Pan-Pacific Association of Applied Linguistics*, *13*(1), 53–69.

Yoon, H. -J., & Polio, C. (2014). A longitudinal study of linguistic complexity in two genres [PowerPoint slides]. Retrieved from http://www.vub.ac.be/TALK/sites/default/tiles/H.%20Yoon%20-%20A%20longitudinal%20study%20of%20linguistic%20complexity%20in%20two%20genres.pdf

Yoon, S. Y. (2011). Students' reflection on feedback in L2 writing in blended learning. *Multimedia-Assisted Language Learning*, *14*(2), 235–261.

Yoon, S. Y., & Lee, C. -H. (2010). The perspectives and effectiveness of blended learning in L2 writing of Korean university students. *Multimedia-Assisted Language Learning*, *13*(2), 177–204.

Younes, M. I. A. (2016). *The impact of CLIL on Arabic, English and content learning of Arab high school students in the UAE* (Master's thesis). Retrieved from https://dspace.aus.edu/xmlui/handle/11073/8333

Yu, G. (2007). Lexical diversity in MELAB writing and speaking task performances. In J. S. Johnson (Ed.), *Spaan fellow working papers in second or foreign language assessment* (Vol, 5, pp. 79–116). Washington: English Language Institute, University of Michigan.

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

Yule, G. (1997). *Referential communication tasks*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Zabihi, R., Rezazadeh, M., & Dastjerdi, H. V. (2013). Creativity and narrative writing in L2 classrooms: Comparing individual and paired task performance. *Bellaterra: Journal of Teaching & Learning Language and Literature*, *6*(3), 29–46.

Zahedi, K., & Dorrimanesh, P. (2008). Metacognitive learning strategies and academic success of TEFL MA Students in distance education. *International Journal of Criminology and Sociological Theory*, *1*(2), 161–176.

Zaid, M. A. (2011). Effects of web-based pre-writing activities on college EFL students' writing performance and their writing apprehension. *Journal of King Saud University-Languages and Translation*, *23*(2), 77–85.

Zainuddin, H., Yahya, N., Morales-Jones, C., & Ariza, E.N. (2011). *Fundamentals of teaching English to speakers of other languages in K-12 mainstream classrooms* (3rd ed.). Dubuque, IA: Kendall/Hunt Publishing

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, *19*(1), 79–101.

Zarei, N., & Al-Shboul, Y. (2013). Participating in a blog: Jordanian EFL learners' voices. *English Linguistics Research*, *2*(2), 53–59. doi:10.5430/elr.v2n2p53

Zughoul, M.R. (2003). Globalization and EFL/ESL pedagogy in the Arab world. *Journal of Language and Learning*, *1*(2), 1–38.

Zyad, H. (2016). L2 writing development: The case of two high-achieving and two struggling college-level students. *International Journal for 21st Century Education*, *3*(1), 25–36.

Appendixes

Appendix A: Writing Skills Pre-Test

Al-Imam Muhammad

bin Saud Islamic University

Ministry of Higher Education

College of Languages & Translation

Time:………..

(Writing Skills Pre-Test)

Name:

Level:

Group:

Write on the following topics:

1. Are you for or against smoking in general and smoking in public places in particular? Discuss people's different opinions about this.

2. The different brands of cars in Saudi Arabia. Discuss the advantages and disadvantages of each brand.

•••
•••
•••
•••
•••
•••
•••
•••
•••
•••
•••
•••
•••
•••
•••
•••
•••

Appendix B: Writing Skills Mid-Term Test

Al-Imam Muhammad

bin Saud Islamic University

Ministry of Higher Education

College of Languages & Translation

Time:………..

(Writing Skills Mid-Term Test)

Name:

Level:

Group:

Write on the following topics:

1. Are you for or against bodily punishment at schools? Discuss people's different opinions about this issue.

2. Nowadays, people in Saudi Arabia and the Arab world can watch a variety of TV channels. Discuss the advantages and disadvantages of each type of these channels.

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

Appendix C: Writing Skills Post-Test

Al-Imam Muhammad   College of Languages & Translation

bin Saud Islamic University   Time:………..

Ministry of Higher Education

(Writing Skills Post-Test)

Name:

Level:

Group:

Write on the following topics:

1. Mobile phones have brought more good than harm to our life. Do you agree or disagree?

2. Saudi university students' different hobbies. Discuss the advantages and disadvantages of each hobby type.

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....