An Empirical Investigation of the Role of Formulaic Sequences in Upgrading EAP Students' Academic Writing Skills

by

Lina AlHassan

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Applied Linguistics and Discourse Studies

Carleton University

Ottawa, Ontario

Abstract

Since the early 1990s, the body of literature on second language (L2) composition instruction has been profoundly influenced by corpus linguistics research, which empirically demonstrated the recurrence of formulaic sequences—defined as highly frequent register-, genre-, and/or discipline-bound multiword strings (Hyland, 2008a; Simpson-Vlach & Ellis, 2010)—in academic writing. These findings led several researchers to advocate the importance of integrating formulaic sequences into L2 writing pedagogy and the effectiveness of focused instruction, salient input, and repeated exposure in promoting their acquisition and improving L2 learners' writing proficiency (e.g. Cowie, 1992; Lewis, 1997; Willis, 2003; Wood, 2009, 2015).

The growing theoretical accounts of the importance of integrating formulaic sequences into L2 composition instruction has been associated with very few research studies on formulaic sequences and L2 learners' academic writing skills, most of which focused on one pedagogical approach and yielded inconsistent results. Accordingly, there is, as yet, no ample evidence to support the effectiveness of formulaic sequences in improving L2 learners' academic writing skills; nor is their sufficient evidence to demonstrate the pedagogical efficacy of the above-mentioned approaches.

To address these gaps, this quasi-experimental research study investigated the extent to which focused instruction and salient input can help English for Academic Purposes (EAP) students acquire and manipulate formulaic sequences to produce better quality graphical summaries and essays. The study was conducted in an EAP program, different pedagogical interventions were delivered, and 382 timed written texts were collected at three time points from 67 upper-intermediate and advanced students. The

textual data were manually quantified, statistically analyzed, and blindly evaluated. The study results were generally in agreement with the literature on the effectiveness of focused instruction and salient input but at odds with the literature on the positive effects of formulaic sequences on L2 academic achievement. The results also highlighted the higher learning gains that a focused instructional approach resulted in than a saliency approach and revealed the effects of the writing task and proficiency level on the participants' utilization of formulaic sequences. These results formed the basis for suggesting implications for EAP contexts and highlighting yet-to-explore areas for future research.

Keywords: formulaic sequences, focused instruction, saliency, EAP students' writing proficiency

With the support of the Social Sciences and Humanities Research Council of Canada

Acknowledgements

I would first like to express my sincere gratitude to the Social Sciences and Humanities Research Council of Canada for supporting this research study. I would also like to express my immense gratitude and deep appreciation to my supervisor, Dr. David Wood, for his invaluable guidance and support throughout the development and completion of this dissertation. Without his valuable advice, constructive feedback, and continuous support, the completion of this dissertation would not have been possible. In addition, I am deeply grateful to my advisory committee members, Dr. Stuart Webb and Dr. Scott Douglas, for taking the time to read and provide insightful comments and valuable suggestions at different stages of completing this research project. Special thanks go to the Center for Quantitative Analysis and Decision Support at Carleton University for their assistance and advice.

I would like to express my thanks to the EAP program manager for granting me the permission to conduct this research study and to the teachers who allowed me to access their classrooms. Special thanks to all the participants whose willingness to volunteer was essential for completing this dissertation. I would also like to thank my friends and colleagues for supporting me during the arduous journey of completing my PhD.

I am especially grateful to my husband, Hassan Al-Mohammad, for his selfless love, devotion, and support, which encouraged me to complete my PhD despite the tough and stressful time we were both going through. I am also greatly indebted to my mother and siblings for their endless emotional support. Finally, I would like to dedicate this

dissertation to my dearest father, Fathi AlHassan, whose soul has illuminated my path since his death and will do for many years to come.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter One: Introduction

Research on teaching academic writing skills has been a growing area of interest in second language acquisition (SLA) research and pedagogy since the 1940s (Ferris, 2012; Silva, 1990). This theoretical interest emerged as a result of the increasing implementation of English academic writing as an assessment tool in postsecondary education, the syntactic and structural complexity of English academic writing, and the growing status of English as an academic lingua franca (Celce-Murcia & Olshtain, 2000; Hinkel, 2004; Hüttner, 2010; Llach, 2011; Murray, 1985; Swales & Feak, 2012; Zhu, 2006). The paramount importance and inherent complexity of academic writing have led to a burgeoning of several approaches to teaching this communicative skill to L2 users, some of which were informed by approaches to teaching academic writing to English native (L1) speakers and most of which marked a departure from the long prevailing practices in L2 classrooms (Hinkel, 2004; Hyland, 2007; Johns, 1990; Shih, 1986; Silva, 1990).

Among the widely cited approaches in the literature are the product-based approach, the process-based approach, the genre-based approach, and the content-based approach. These approaches respectively focus on the importance of upgrading L2 learners' linguistic competence, the utility of immersing L2 learners in the multiple stages of the writing process, the effectiveness of familiarizing L2 learners with the rhetorical features of different types of academic texts, and the efficacy of focusing on writing tasks peculiar to university requirements (Celce-Murcia & Olshtain, 2000; Hyland, 2007; Kasper, 2000; Shih, 1986; Silva, 1990).

While each of the above-mentioned approaches has had its own merits and eminence in SLA research and pedagogy, the tendency to overemphasize certain writing aspects at the expense of others within each approach has attracted criticism, resulting in the emergence of a new approach or calls for integrating aspects from multiple approaches in order to fully equip L2 learners with the linguistic knowledge required to conform with the conventions of academic writing (Badger & White, 2000; Hinkel, 2004; Hyland, 2003a; Matsuda, 2003a, 2003b; Shah, 2003; Silva, 1990). Such conformity was mainly attributed to three linguistic dimensions: (a) a mastery over advanced syntactic rules, (b) a command over a wide range of vocabulary words, and (c) a grasp of rhetorical conventions, to which knowledge of formulaic sequences was added as a fourth dimension (Cowie, 1992; Hinkel, 2004; Howarth, 1998a; Jones & Haywood, 2004).

The fourth dimension was theoretically proposed in the 1990s by several SLA and corpus linguistics researchers (e.g. Cowie, 1992; Granger, 1998; Howarth, 1998a; Verstraten, 1992). This theoretical proposition was opposed but then gained ground because of the revolutionary results of corpus linguistics research, which empirically demonstrated that academic writing is marked by the frequent occurrence of different types of formulaic sequences (Byrd & Coxhead, 2010; Ellis & Simpson-Vlach, 2009; Howarth, 1998a). The empirical evidence in support of the formulaic nature of academic writing, together with the complexity of formulaic sequences and L2 learners' failure to properly utilize them in their writing production, formed the foundation for a vast array of literature that advocated the importance of integrating formulaic sequences into any approach to teaching academic writing skills (Cortes, 2004; Cowie, 1992; Lewis, 2000a, 2000b, 2000c, 2002; Oakey, 2002).

Three major approaches (i.e., repeated exposure, salient input, and focused instruction) have been proposed in this body of literature in order to address formulaic sequences in the language classroom. Each of these three approaches has its distinct methodology and pedagogical utility which, according to its proponents, help L2 learners notice, acquire, and utilize formulaic sequences in their writing production and, hence, promote their academic wiring proficiency (Ädel & Erman, 2012; Boers & Lindstromberg, 2009; Byrd & Coxhead, 2010; Jones & Haywood, 2004; Lewis, 2002; Staples, Egbert, Biber, & McClair, 2013).

The rapidly growing literature on the importance of integrating formulaic sequences into teaching academic writing and on the efficacy of different approaches to presenting this language phenomenon has been associated with a few attempts to empirically investigate the effects of a single presentation method of formulaic sequences on L2 academic writing proficiency, but these investigations yielded inconsistent results (e.g. AlHassan & Wood, 2015; Čolović-Marković, 2012; Jones & Haywood, 2004; Li & Schmitt, 2009). Accordingly, there is, as yet, no solid empirical evidence for the effectiveness of integrating formulaic sequences into teaching academic writing, nor is there ample evidence to support the pedagogical utility of the three above-mentioned presentation methods of formulaic sequences (i.e., repeated exposure, saliency, and focused instruction).

To address the existing gaps between the extensive theoretical accounts and the dearth of empirical evidence, this research study was designed by adapting a commonly used quasi-experimental design in educational research, that is, the pretest-posttest-nonrandomized-control-group design (Campbell & Stanley, 1963; Dörnyei, 2007; Leedy,

1997; Vogt, 2007). The designed study set out to empirically investigate the effects of two presentation methods of formulaic sequences (i.e., focused instruction and saliency) on EAP students' acquisition, internalization, and utilization of formulaic sequences when completing two different writing tasks (i.e., graphical summaries and essay responses). To this end, the pretest-posttest-nonrandomized-control-group design was expanded to include three groups (i.e., focused instruction, saliency, and control groups) at two different proficiency levels, three measurement levels (a pretest at the beginning of the study, a posttest at the end of the pedagogical intervention, and a delayed posttest one month after collecting the posttest), and writing materials for the control group. Different types of instruments were designed, partially piloted in AlHassan's (2016) research study, and used during the multiple stages of this research study, which took place in an EAP program at a large Canadian university during regular class hours.

At the beginning of the study, the EAP students (i.e., 115 upper-intermediate and advanced EAP students) in all the groups to which the researcher was granted access (a total of seven) were required to complete two timed writing tasks, that is, a graphical summary for 20 minutes and an essay response for 40 minutes. Each randomly assigned pedagogical intervention to each of the seven groups was then delivered based on a number of major concepts proposed in several approaches to L2 instruction. Each target pedagogical intervention in each group lasted for five hours and was divided into presentation and practice stages. The students in each group were required to complete two writing tasks at the end of and one month after the target pedagogical intervention following the same timing procedure. The three stages of data collection resulted in eliciting 382 graphical summaries and essay responses from 67 upper-intermediate and

advanced EAP students, the total number of students recruited in this study. In addition to the 382 texts, textual evaluation was obtained from a judge, an EAP instructor, based on the two analytical rubrics. The collected textual data were quantified, and the elicited quantitative variables along with the EAP judge's evaluation were quantitatively analyzed in order to highlight any differences within and/or between groups and associations between different variables.

This dissertation is divided into seven chapters, each of which, except for the current chapter, opens with an introduction that outlines its major component parts. The first chapter serves as an introduction to the dissertation. It provides a brief account of the theoretical background, the purpose, and the design of this research study along with an overview of the key components of each chapter. The second chapter, Literature Review, falls into nine major sections. The first two sections cast light on the importance and complexity of academic writing as well as the strengths and weaknesses of four major approaches to teaching academic writing skills to L2 learners. The third section of the Literature Review tackles the major definitions and the key characteristics of formulaic sequences. The major findings of corpus linguistics research on academic writing are discussed in the fourth and fifth sections. The sixth and seventh sections respectively discuss the importance and pedagogical utility of integrating formulaic sequences into teaching L2 academic writing skills. Suggested pedagogical approaches and classroom activities are reviewed in the eighth section and are followed by a summary of the major points tackled in the literature review in the ninth section.

The third chapter of this dissertation recapitulates the major points discussed in the Literature Review, reveals the gap between the growing theoretical account of the

importance of teaching formulaic sequences and the lack of solid empirical evidence, lists the objectives of this research study, and concludes with the five research questions that guided this study. The fourth chapter, Methods, is divided into nine major sections, most of which consist of several subsections that focus on different aspects of the design of this study. The first section provides an overview of the current study and AlHassan's (2016) study, which helped partially pilot some of the instruments. The second section describes the adapted design for this research study as well as the methodological and pedagogical rationale for adapting the current design. Both the participants and the context are described in the third section.

The fourth and fifth sections of the Methods chapter discuss the procedures followed to extract formulaic sequences from two major sources and the instruments developed for pedagogical interventions and data collection. The sixth section of the methods chapter provides an account of four L2 instructional approaches that informed delivering the three distinct types of pedagogical interventions. The multiple stages of pedagogical intervention and data collection are described in the seventh section and are followed, in the last two sections, by a description of the coding methods applied to derive quantitative variables from the textual data and the imputation, screening, descriptive, and inferential statistical techniques and tests that were carried out on the elicited variables. The major findings obtained from the statistical analyses are discussed in the fifth chapter and are integrated with relevant literature in the sixth chapter. The last chapter of this dissertation, Conclusion, reiterates the major findings of this research study, highlights its implications, discusses its limitations, provides some suggestions for future research, and concludes the entire dissertation.

Chapter Two: Literature Review

This chapter presents a synthesis of research on academic writing and formulaic sequences in an attempt to highlight the theoretical and empirical underpinnings of placing formulaic sequences at the center of writing pedagogy. With this goal as a guide, the chapter is divided into nine sections. The chapter, in the first section, tackles the various functions that writing, as an expressive skill, executes in different contexts, discusses the increasing importance of English academic writing in postsecondary education, and highlights the rationale for teaching English academic writing to L1 and L2 speakers of English. The second section explores the strengths and weaknesses of four major approaches to teaching English academic writing to L2 learners and concludes with a reference to the calls proposed in SLA research in the early 1990s to integrate formulaic sequences into writing pedagogy. The major definitions and defining characteristics of formulaic sequences are reviewed in the third section and are followed in the fourth section by a theoretical account of the shift in the view of academic writing as a result of the advances in corpus linguistics research and the key findings of corpus linguistics research on academic writing.

The fifth section explores some empirically derived lists of recurrent formulaic sequences in academic prose. The sixth and seventh sections respectively highlight the calls for assigning formulaic sequences a prominent position in second language pedagogy and the rationale for placing formulaic sequences at the center of writing pedagogy. The main approaches to addressing formulaic sequences in the language classroom along with some suggested classroom activities are tackled in the eighth section. The chapter concludes, in the ninth section, with a reiteration of the major

concepts that provide an incentive for conducting this quasi-experimental research study.

Writing Skills: Importance and Complexity

Writing is as an efficacious tool that has been used throughout history to conserve information and transfer knowledge to a target audience in a specific context; it is one of the most disciplined mediating tools that language users utilize to convey innovative ideas, parse complex meanings, and execute discursive actions in the form of readable texts (Elbow, 2012; Hyland, 2003a, 2016; Kress, 1994; Mayher, Lester, & Pradl, 1983; Murray, 1985; Paré, 2009; Raimes, 2002). As a meaning-making process,

> writing is the production of the written word that results in a text but the text must be read and comprehended in order for communication to take place. The writer, in other words, communicates his/her ideas in the form of a written text from which a known or unknown reader will eventually extract the ideas and their meanings. (Celce-Murcia & Olshtain, 2000, p. 142)

Writing is, thus, intimately bound up with a writer who shares general and scientific information in and a reader who ponders complex meanings from a written text (Celce-Murcia & Olshtain, 2000; Elbow, 2012; Hyland, 2016). It is a *goal-oriented* process that transcends the "features of the composing situation to [encompass] the purposes, goals and uses that the completed text may eventually fulfill" (Hyland, 2016, p. 21).

Given its communicative efficacy, writing has increasingly served as the primary medium for communicating and assessing academic knowledge worldwide. It has long been used as a powerful means of disseminating academic knowledge in published research (Hüttner, 2010; Jenkins, Cogo, & Dewey, 2011; Römer, 2009). Since the 1940s, this knowledge-dissemination form has mainly been produced in English due to the

growing status of English as an *academic lingua franca* (Björkman, 2013; Brick, 2012; Cogo & Dewey, 2012; Hüttner, 2010; Jenkins et al., 2011; Smit, 2010). Writing has also held a dominant status in English postsecondary institutions. In these institutions, university students are mainly assessed based on their ability to demonstrate their comprehension of academic materials in an impersonal and objective written form (i.e., academic writing) rather than their ability to deliver a persuasive speech, a form of assessment that prevailed in English-medium universities for long (Hewings & Hewings, 2002; Hinkel, 2004; Mitchell, 2004).

This shift in university assessment, which first appeared in English-dominant countries in the early 19th century, has placed English academic writing at the core of academic curricula and has been increasingly adopted in non-English medium universities as a result of the growing tendency to globalize education (Björkman, 2013; Brick, 2012; Hinkel, 2004; Hüttner, 2010; Kazemi, Katiraei, & Rasekh, 2014; Mitchell, 2004; Römer, 2009). University students in many non-English speaking countries partake in different writing tasks (e.g. reflections, research papers, etc.) in order to exhibit their in-depth understanding of core academic materials, an understanding that is often determined by their ability to effectively analyze and eloquently synthesize information from multiple academic sources in written texts (Hewings & Hewings, 2002; Hinkel, 2004; Llach, 2011; Paltridge, 2004; Swales & Feak, 2012; Zhu, 2006).

The use of English academic writing as a means of communicating and assessing academic knowledge worldwide was associated with a burst of new literature on the importance of proposing some effective pedagogical approaches to teaching academic writing (Brick, 2012; Celce-Murcia & Olshtain, 2000; Ferris, 2012; Hinkel, 2004;

Hüttner, 2010; Mayher et al., 1983; Murray, 1985). The emerging literature was partly due to the view of academic writing as a distinct skill that has its own characteristics: academic writing is "thought of as a unique variety of English which, although no one actually speaks it, has distinct properties that need mastering if one hopes to be able to read and write well" (Mayher et al., 1983, p. 4). These *properties* include structural complexity, rigorous precision, high density, and increased objectivity (Biber, 2006; Elbow, 2012; Mitchell, 2004; Murray, 1985; Williams, 2005). The emerging literature was also partly the result of the complex nature of academic writing, which was mainly ascribed to the advanced linguistic proficiency required for and the complex cognitive processes involved in the composition process (Celce-Murcia & Olshtain, 2000; Hinkel, 2004; Murray, 1985).

The unique and complex nature of academic writing was viewed by many scholars (e.g. Ferris, 2012; Hairston, 1982; Mitchell, 2004; Murray, 1985) as a compelling reason to depart from the long prevailing "*vitalist* attitude" towards writing instruction: "no one can really teach anyone else how to write because writing is a mysterious creative activity that cannot be categorized or analyzed" (Hairston, 1982, p. 78). As Murray (1985) so eloquently puts it:

> meaning is not thought up and then written down. The act of writing is an act of thought. This is the principal reason writing should be taught in the academy, yet, ironically, it is this concept that is most often misunderstood by academicians. (p. 3)

This departure resulted in the emergence of several pedagogical approaches to teaching L1 academic writing skills and was paralleled by reevaluating and revolutionizing L2

composition instruction in the mid 20th century (Ferris, 2012; Ferris & Hedgcock, 2014; Matsuda, 2003b; Silva, 1990).

Approaches to Teaching L2 Academic Writing Skills

The 1940s marked a turning point in L2 writing pedagogy (Ferris, 2012; Ferris & Hedgcock, 2014; Matsuda, 2003b; Silva, 1990). After occupying a marginal position, composition instruction started to receive theoretical attention because of (a) the increasing number of L2 students who study in English-medium universities, (b) the view of academic writing as a challenging task that L2 university students need to master in order to meet the complex demands of the academy, and (c) the awareness of the differences between general writing skills which L2 learners master in the language classroom and what L2 university students are required to do in academic contexts (Celce-Murcia & Olshtain, 2000; Ferris, 2012; Ferris & Hedgcock, 2014; Hinkel, 2004; Murray, 1985; Richard & Renandya, 2002; Williams, 2005). This theoretical attention resulted in proposing several approaches to teaching L2 academic writing skills, most of which were based on theories of L1 academic writing pedagogy and each of which held a dominant status for a few years, but it was, then, criticized and replaced, at least partially, by a new approach (Johns, 1990; Silva, 1990; Williams, 2005). Among the widely cited approaches in the literature are the current-traditional rhetoric, the process-based approach, the genre-based approach, and the content-based approach.

The current-traditional rhetoric. The current-traditional rhetoric, also known as the product-based approach, emerged in the 1960s in reaction to controlled writing—a teaching practice which limited writing instruction to the formation of grammatical sentences and reduced the writing skill to a mere reinforcement for oral skills—and

placed expressive skills that allow L2 learners to construct long texts in which they execute different discursive functions at the core of writing instruction (Badger & White, 2000; Curry & Lillis, 2003; Matsuda, 2003a, 2003b; Pratt, 2011; Raimes, 1991; Silva, 1990). The current-traditional rhetoric was adapted from the traditional approach that dominated teaching academic writing to L1 speakers in the 19th and early 20th century and viewed writing as a linear process that leads to the construction of modeled five-paragraph essays (Matsuda, 2003a; Mitchell, 2004; Silva, 1990). It was also partly influenced by the theory of contrastive rhetoric, which suggests that L2 writers' rhetorical knowledge needs to be explicitly addressed because L2 writers' rhetorical skills often deviate from the expectations of L1 readers (Kaplan, 1967; Leki, 1991; Silva, 1990).

Proponents of the current-traditional rhetoric advocate the importance of linguistic accuracy and organizational skills (Badger & White, 2000; Curry & Lillis, 2003; Hairston, 1982; Matsuda, 2003a; Silva, 1990). That is, classroom instruction should focus on the appropriate use of linguistic units to construct a well-written paragraph (i.e., one that includes a topic sentence, supporting ideas/details, and a concluding sentence) and execute such discursive functions as exemplification, comparison, and exposition (Curry & Lillis, 2003; Hyland, 2003a; Silva, 1990). Instruction should also revolve around L2 learners' ability to combine and rearrange different types of discourse elements (i.e., sentences and paragraphs), which can be either provided by the instructor or composed by students, to construct a textual product (Ferris & Hedgcock, 2014; Silva, 1990).

Teachers who adopt the current-traditional rhetoric view writing instruction as a process of '*assisted*' or '*guided imitation*' that is composed of four stages: *familiarization*, *controlled writing*, *guided writing*, and *free writing* (Badger & While,

2000; Hyland, 2003a, 2016). At the *familiarization* stage, whose major objective is to highlight certain distinctive features of a text, L2 learners are usually provided with a model or exemplar that is appropriate for university-level students, are asked to decompose it into linguistic segments (i.e., words, phrases, etc.), and are required to examine its style and organization (Badger & White, 2000; Curry & Lillis, 2003; Hyland, 2003a; Silva, 1990). L2 learners, then, practice the learned structures in both controlled and guided writing, for example, completing a text by using provided sentences and writing a text based on an outline, respectively (Badger & White, 2000; Hyland, 2003a; Silva, 1990). At the *free writing* stage, L2 learners utilize the linguistic knowledge and organizational skills they have already mastered to compose a genuine text (Badger & White, 2000; Curry & Lillis, 2003; Hyland, 2003a; Silva, 1990).

Although the current-traditional rhetoric was applied to teaching academic writing skills for several years and is still present in some textbooks, it has been criticized by many scholars for (a) its overemphasis on language forms and organization skills and (b) its understatement of L2 writer's contrastive rhetorical context as well as the conscious and unconscious psychological processes involved in generating ideas and composing a text; these critical views paved the way for the emergence of a '*process-oriented*' and '*student-centered*' approach to L2 academic writing instruction, that is, the process-based approach (Curry & Lillis, 2003; Ferris & Hedgcock, 2014; Hairston, 1982; Leki, 1991; Matsuda, 2003a; Raimes, 1991; Silva, 1990).

The process-based approach. This approach to teaching academic writing developed between the late 1960s and 1970s as a counteraction to focusing on the perfection of a writing product and the view of composition as a linear process, and it has

been prevalent in teaching general and academic writing skills since the 1980s (Curry & Lillis, 2003; Ferris & Hedgcock, 2014; Hyland, 2003a; Matsuda, 2003a, 2003b; Mora-Flores, 2009; Silva, 1990). Similar to the current-traditional rhetoric, the process-based approach is based on theories of teaching L1 academic writing skills; unlike the former, however, it views writing as a nonlinear process and focuses on the top-down processes involved in the composition of a written text (Celce-Murcia & Olshtain, 2000; Hinkel, 2004; Matsuda, 2003a; Mayher et al., 1983; Mora-Flores, 2009; Murray, 1985; Silva, 1990).

Proponents of the process-based approach advocate the pedagogical efficacy of involving L2 learners in the different stages of the writing process in order to promote their writing fluency and proficiency. In other words, classroom activities should be geared towards leading L2 learners through the four major stages that are usually involved in the writing process—*planning*, *drafting*, *revising*, and *editing*—and the other three stages that are imposed by L2 writing instructors, namely, *responding*, *evaluation*, and *publishing* (Celce-Murcia & Olshtain, 2000; Matsuda, 2003a, 2003b; Mayher et al., 1983; Murray, 1985; Raimes, 1991; Seow, 2002).

Planning. Taking into consideration that most writers utilize ideas from their surroundings, classroom procedures embracing the process-based approach start with a *planning* stage during which ideas that relate to a discussion topic are generated in groups or as a class (Celce-Murcia & Olshtain, 2000; Mayher et al., 1983; Murray, 1985; Seow, 2002). This stage is of prime importance because it helps reduce L2 learners' anxiety, which is often elevated by not knowing what to write (Celce-Murcia & Olshtain, 2000; Murray, 1985; Seow, 2002). According to Seow (2002), the planning stage "moves

students away from having to face a blank page toward generating tentative ideas and gathering information for writing" (p. 316). Among the strategies that can be implemented to generate ideas at this stage are brainstorming, outlining, free writing, and discussion questions (Celce-Murcia & Olshtain, 2000; Hyland, 2003a; Seow, 2002).

Drafting. Once sufficient information is elicited at the planning stage, L2 learners partially or fully utilize the generated information in order to produce a draft—an unevaluated rough version that is neither complete in meaning nor edited for form (Celce-Murcia & Olshtain, 2000; Mayher et al., 1983; Seow, 2002). When writing their drafts, L2 learners may include such writing elements as an introduction, a rhetorical question, and a summary of the points of discussion without much attention to linguistic accuracy, organization, or style (Seow, 2002). Although the *drafting* stage focuses on writing fluency with a very minimal attention to grammatical and structural accuracy, the writing elements included during this stage should form the basis for their final versions (Mayher et al., 1983; Seow, 2002).

Responding. *Responding* is a major stage that distinguishes the process-based approach from traditional ones. Instead of passively receiving feedback from instructors on their final written products, L2 learners receive constructive feedback that relates to both content and form on their drafts (Hyland, 2003a; Matsuda, 2003a; Mayher et al., 1983). In other words, L2 learners receive some recommended changes on their drafts to help them improve their texts at the revision stage and to correct their mistakes at the editing stage (Mayher et al., 1983; Mora-Flores, 2009; Seow, 2002). This feedback, oral or written, can be provided by writing instructors or peers in either the form of *retrospective structuring* (i.e., focusing on the linguistic systems used to convey

meaning) or *projective structuring*, that is, commenting on the appropriateness of organization and style (Mayher et al., 1983; Seow, 2002).

Revising. L2 learners, at the revision stage, reassess the content of their paper on the basis of the feedback received from their writing instructors or their peers; they may rewrite sections, add or eliminate information, or change the organization of their texts to ensure that the intended message is clearly conveyed to the target audience (Celce-Murcia & Olshtain, 2000; Murray, 1985; Seow, 2002). The revision stage is, as Murray (1985) explains, a cornerstone in the composition process because "for most writers, especially inexperienced writers or experienced writers dealing with a new subject, a new audience or a new genre, writing is indeed rewriting" (p. 56).

Editing. After revising their writing for content, L2 learners carefully proofread their own or their peers' texts to identify and correct lexical, grammatical, or structural errors (Celce-Murcia & Olshtain, 2000; Matsuda, 2003a; Mayher et al., 1983; Seow, 2002). The *editing* stage takes place prior to submitting the final written text rather than during the drafting stage because a focus on form at the earlier stages of writing may impede writing fluency (Mayher et al., 1983; Seow, 2002).

Evaluation and publishing. After assigning holistic or analytical grades during the *evaluation* stage, writing instructors can implement such post-writing activities as reading aloud, posting a text on a bulletin board, or publishing a text in a student magazine to mark the end of the writing process (Mayher et al., 1983; Seow, 2002). The *publishing* stage has a dual function as it both motivates L2 learners to perfect their writing before they hand in their final products to their writing instructors and recognizes L2 learners' efforts as valuable and rewarding (Mayher et al., 1983; Seow, 2002). The

publishing stage is of special importance because "no one will work hard at a task unless it is toward a worthwhile goal" (Mayher et al., 1983, p. 328).

Despite its popularity in both general and academic English courses, the effectiveness of the process-based approach in adequately preparing L2 students for university writing has been questioned by several researchers (Hinkel, 2004; Silva, 1990). Placing the writing process rather than the written product at the core of writing pedagogy indeed helps improve L2 learners' expressive skills, yet issues that relate to lexical, grammatical, rhetorical, and structural accuracy persist (Ferris, 2002; Hinkel, 2004; Hyland, 2016; Reppen, 2002). This shortcoming in the process-based approach is predominantly ascribed to its heavy reliance on the same approach applied to L1 settings, and, hence, it fails to take into consideration the fact that L2 learners might lack in the linguistic knowledge required for such a complex skill as academic writing (Celce-Murcia & Olshtain, 2000; Ferris, 2002; Hinkel, 2004).

Moreover, evaluating students based on the processes involved in constructing a written text rather than the final written product might be misleading for many L2 learners because it does not realistically reflect evaluation in academic contexts, where "the faculty in the disciplines are not particularly concerned about the writing process that affects (or does not affect) the quality of the writing product" (Hinkel, 2004, p. 6). Furthermore, having L2 learners write about their personal experiences is radically different from what university students are required to do in most of their written assignments, that is, synthesizing information from different academic sources (Ferris & Hedgcock, 2014; Hewings & Hewings, 2002; Hinkel, 2004; Kasper, 2000; Reppen, 2002; Shih, 1986). Such critiques have contributed to the simultaneous emergence of the genre-

based approach and the content-based approach.

The genre-based approach. The genre-based approach to teaching academic writing was greatly influenced by the view of writing as a social communicative act that should conform with the social context in which it is produced and the expectations of the intended readers to whom it is directed (Badger & White, 2000; Curry & Lillis, 2003; Hyland, 2003a, 2003b, 2004, 2007; Johns, 2002). This view of writing has sprung from the awareness that similar communicative acts are executed in texts that share several characteristics and belong to a particular genre, defined as "abstract, socially recognized ways of using language" (Hyland, 2016, p. 9). The genre-based approach, with its primary focus on both language and context, is considered a radical departure from the long prevailing process-informed pedagogy and has gained considerable popularity since the mid 1980s (Henry & Roseberry, 1998; Hyland, 2004, 2007).

As a context-based, reader-focused, and goal-oriented approach, the genre-based approach stresses the importance of effectively manipulating linguistic knowledge to communicate information to readers in a particular social context rather than focusing on the writing process that leads to the construction of a written text (Badger & White, 2000; Henry & Roseberry, 1998; Hyland, 2003a, 2003b, 2004, 2007, 2016; Marco, 2000). Successful communication is contingent on the writer's ability to construct a written text that complies with the target reader's expectations (Hyland, 2007). Without an understanding of how words and syntactic structures are utilized to construct and organize written texts in a particular context to a target audience, L2 writers "will continue to find their writing practices regarded merely as failed attempts to approximate prestigious forms" (Hyland, 2003b, p. 24).

The genre-based approach can be executed in the writing classroom following different techniques, of which the most renowned two are Systemic Functional Linguistics (SFL) and English for Specific Purposes (ESP) (Hyland, 2003a, 2007). While SFL and ESP view genres differently—that is, respectively as "broad rhetorical patterns" and as "purposive actions routinely used by community members to achieve a particular purpose"(Hyland, 2007, pp. 153-154) and have different areas of focus (i.e., with the former focusing on language and the latter focusing on context), they both place at their core the key concepts of the genre-based approach (Hyland, 2003b, 2007). These key concepts include the importance of identifying L2 writers' needs, exposing them to various types of academic texts, and raising their awareness of the commonalities as well as the differences between these texts (Badger & White, 2000; Henry & Roseberry, 1998; Hyland, 2003b, 2007; More-Flores, 2009; Reppen, 2002).

Writing instructors who opt for the genre-based approach introduce different texts that relate to the same genre and have L2 learners identify the common defining features of these texts, an activity that is likely to familiarize L2 learners with the common identifying characteristics and the rhetorical organization of the genre they will, later on, produce in writing (Henry & Roseberry, 1998; Hyland, 2003a, 2003b, 2007, 2016; Mora-Flores, 2009; Reppen, 2002). Writing teachers, then, discuss the lexis and structure of the genre under investigation with their students and provide them with extensive constructive feedback on both language and organization to guide them through the process of writing in a particular genre (Badger & White, 2000; Hyland, 2003b; Knapp & Watkins, 1994). The support offered by writing instructors decreases as L2 learners' writing skills improve (Hyland, 2003b; Reppen, 2002).

Similar to the product-oriented and the process-oriented approaches, the genre-based approach has faced criticism. The first major concern about the genre-based approach is the contradiction between its views of a communicative text as contextually grounded and goal oriented and the unauthentic context in which it is adopted, that is, the language classroom (Dias & Paré, 2000; Hyland, 2004, 2007; Johns, 2002). The second major concern about the genre-based approach is power relations (Hyland, 2003b, 2004; Johns, 2002; Luke, 1996). Focusing on *dominant discourses* and rigid structural templates in the language classroom has led critics to raise "questions about ideology and about the fact that genres embody the values, attitudes, and 'ways of doing' of the dominant culture" (Hyland, 2004, p. 18). Other concerns about this approach relate to overemphasizing organizational skills at the expense of linguistic knowledge, viewing L2 learners as passive recipients, and inhibiting L2 learners' creativity (Badger & White, 2000; Hyland, 2004).

The content-based approach. This approach to teaching academic writing skills, which appeared in the mid 1980s and has increasingly gained adherence since the 1990s, amalgamates second language acquisition theories (e.g. *Krashen's comprehensible input* and *Cummins's two-tiered skill model*) with cognitive learning theories that stress the importance of pedagogical support (i.e., scaffolding) and practice in upgrading L2 learners' language skills (Amiri & Fatemi, 2014; Brown, 2004; Kasper, 2000; Raimes, 1991). With its focus on improving the skills crucial to L2 learners' success in English-medium education systems, its view of language as a medium of communication rather than an object of instruction, and its manipulation of academic materials, the content-based approach has been considered a radical departure from most prevailing traditional

approaches in which instruction mainly revolved around language forms and structure and L2 learners wrote about their personal experiences (Brown, 2004; Kasper, 2000; Shih, 1986; Warrington, 2008).

The content-based approach simulates academic contexts by requiring L2 learners to use the target language in order to accomplish academic tasks (Amiri & Fatemi, 2014; Brown, 2004; Kasper, 2000; Raimes, 1991; Shih, 1986; Williams, 2005). In other words, writing instructors who are oriented toward this approach present academic materials that relate to various content areas in one or more disciplines (Kasper, 2000; Shih, 1986; Silva, 1990). The presented materials have two functions: providing an increasingly challenging linguistic input that may improve L2 learners' language proficiency and developing the academic skills (e.g. researching, analyzing, synthesizing, etc.) that are prerequisites for writing in academic contexts (Kasper, 2000; Shih, 1986). L2 learners, in turn, are required to read, scrutinize, and discuss the presented materials in the classroom. They, then, synthesize the analyzed information in a variety of written texts to demonstrate their understanding of the subject matter (Kasper, 2000; Shih, 1986; Silva, 1990). These activities have a dual function: they prepare students for tasks they may encounter in postsecondary institutions and pressure them to use their linguistic knowledge in tasks that transcend the mere presentation of general information in a second language (Kasper, 2000; Shih, 1986; Silva, 1990).

Despite its increasing popularity in both English and non-English speaking countries, the content-based approach to teaching L2 academic writing has been criticized by many scholars for four major reasons: (a) L2 learners' linguistic proficiency takes a back seat as a result of overemphasizing content and not addressing linguistic forms

explicitly; (b) utilizing authentic academic texts without adapting them to suit L2 learners' proficiency level may not always be feasible; (c) it requires a lot of preparation on the part of language instructors who may also need training; and (d) the content areas covered in the language classroom may be extremely challenging for both L2 learners and language instructors (Brown, 2004; Shah, 2003; Warrington, 2008; Williams, 2005).

In brief, each of the above-mentioned approaches to teaching academic writing skills has its own methodology and merits which, according to its proponents, are likely to provide L2 learners with the linguistic skills crucial to complying with academic conventions and meeting the demands of the academy. These linguistic skills mainly revolved around three major aspects (i.e., syntactic complexity, lexical accuracy, and rhetorical precision) until, in the early 1990s, a fourth recurrent linguistic aspect in academic writing came to the fore, that is, formulaic sequences.

Formulaic Sequences: Definitions and Defining Characteristics

Since the early 1990s, formulaic sequences have received considerable attention from SLA and corpus linguistics researchers (Biber, 2010; Flowerdew, 2015; O'Donnell, Römer, & Ellis, 2013; Pecorari, 2009; Read & Nation, 2004). Great effort has been devoted to defining and scrutinizing the sematic and syntactic characteristics of formulaic sequences, resulting in numerous overlapping types that have been labeled in the literature with as many as 40 different terms—formulaic language/sequences, lexical bundles, collocations, phrasal expressions, multiword sequences, recurring word combinations, compositional units, and routines, to mention but a few (Biber & Conrad, 1999; Chen & Baker, 2010; Howarth, 1996; Kennedy, 2008; O'Keefe, McCarthy, & Carter, 2007; Pecorari, 2009; Schmitt & Carter, 2004; Wood, 2010a, 2015; Wray, 2002).

Definitions. Because of researchers' tendency to use the same term to refer to syntactically and semantically distinct types of multiword sequences or to use different terms to label identical ones, there is, as yet, no single comprehensive definition of this language phenomenon (Chen & Baker, 2010; Cortes, 2002; Coxhead & Byrd, 2007; Hyland, 2008a; Schmitt & Carter, 2004; Simpson-Vlach & Ellis, 2010; Weinert, 1995). Therefore, providing a thorough definition of formulaic sequences entails scrutinizing different definitions under different terms in the literature, of which the most recurrent ones are *formulaic sequences*, *lexical bundles*, and *collocations*.

Formulaic sequences. The term *formulaic sequences* is among the most common terms in the literature and is frequently used to refer to multiword sequences that are recurrent in both spoken and written language production (Jones & Haywood, 2004; Schmitt & Carter, 2004; Simpson-Vlach & Ellis, 2010; Wray, 2000). Among the earliest and most comprehensive definitions of formulaic sequences is Weinert's (1995), in which formulaic sequences are defined as "multi-word (How do you do?) or multi-form strings (rain-ed, can-'t) which are produced or recalled as a whole chunk, much like an individual lexical item, rather than being generated from individual lexical items/forms with linguistic rules" (p. 182). A very similar definition to that of Weinert (1995) is Wray's (2002), in which she describes a formulaic sequence as

> a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 9)

Based on these two definitions, the term *formulaic sequences* seems to refer to expressions that are orthographically composed of two or more words, but are stored in and retrieved from the long-term memory as monolexical units.

Lexical bundles. *Lexical bundles* is another ubiquitous term that was coined by Biber, Johansson, Leech, Conrad, and Finegan (1999) and has been extensively used in the literature to refer to what is believed to be a distinct and highly register-bound type of multiword sequences (Biber, 2006, 2009; Biber & Conrad, 1999; Cortes, 2004, 2006; Coxhead & Byrd, 2007; Hyland, 2008a; Lee, 2013; Liu, 2012). This term is associated with recurrent multiword sequences that "are usually not structurally complete and not idiomatic in meaning, but… serve important discourse functions in both spoken and written texts" (Biber & Barbieri, 2007, p. 264). Their statistical tendency to co-occur in a register is "consistently functional, indicating that high frequency is a reflection of pre-fabricated or formulaic status" (Biber & Barbieri, 2007, p. 265). In this respect, the term *lexical bundles* seems to encompass frequently recurring multiword sequences that are structurally incomplete and semantically transparent, and whose high frequency contributes to viewing them as important building blocks of fluent linguistic production.

Collocations. The third widely used term in the literature is *collocations*, which first appeared in the *Oxford English Dictionary* in the 19th century and is used to refer to successions of two or more words that co-occur more than what is expected by chance (Biber & Conrad, 1999; Biber et al., 1999; Cortes, 2004; Durrant & Schmitt, 2009; Gledhill, 2000; Handl, 2008; Howarth, 1996; Laufer & Waldman, 2011; Lewis, 2002; Stubbs, 1995, 2002). Among the most widely cited and accepted definitions of collocations is Van Roey's (Erman, Lewis, & Fant, 2013). According to Van Roey

(1990), a collocation refers to "the linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than that of its <<synonyms>> because of constraints which are not on the level of syntax or conceptual meaning but on that of usage" (p. 48). The term *collocations*, therefore, seems to refer to pairs of words whose frequent co-occurrence reflects the arbitrary constraints of using English vocabulary.

Defining characteristics. In addition to being orthographically composed of two or more words, multiword sequences, irrespective of their type or length, should possess one or more of the following defining characteristics in order to have a formulaic status.

Holistic nature. One of the major defining characteristics that is recurrent in almost all the definitions is the holistic nature of multiword sequences; that is, although they are composed of multiple words, their holistic retention and retrieval from the long-memory at the time of language production demonstrate that they are '*glued together*' and treated in the mental lexicon as individual lexical items (Boers & Lindstromberg, 2012; Conklin & Schmitt, 2008; Lewis, 2002; Pawley & Syder, 1983; Willis, 2003; Wood, 2006; Wray, 2000). This defining characteristic has been empirically demonstrated in psycholinguistics research with reference to the processing advantage formulaic sequences offer to language users (Conklin & Schmitt, 2012; Schmitt & Carter, 2004; Schmitt, Grandage, and Adolphs, 2004; Wood, 2010b).

From a psycholinguistic viewpoint, language users have cognitive limitations which allow them to process a maximum of 8-10 words at the time of speech production without hesitation, yet they produce stretches of discourse that go beyond this processing limitation without pauses or hesitation (Conklin & Schmitt, 2012; Schmitt & Carter, 2004; Willis, 2003). One way the mind copes with such limitations is by storing

formulaic sequences in the long-term memory and retrieving them as chunks of language at the time of language production (Conklin & Schmitt, 2012; Lewis, 2002; Liu & Huo, 2011; Schmitt & Carter, 2004; Willis, 2003). As Conklin and Schmitt (2012) put it, "there are compelling reasons to think that the brain represents formulaic sequences in the long-term memory, bypassing the need to compose them online through word selection and grammatical sequencing in capacity-limited working memory" (p. 45).

Relative fixedness. Unlike stretches of discourse that are creatively constructed based on syntactic rules, formulaic sequences have a relative degree of complete or partial fixedness (Biber et al., 1999; Schmitt & Carter, 2004; Simpson & Mendis, 2003; Simpson-Vlach & Ellis, 2010; Weinert, 1995; Willis, 2003). Among the entirely fixed multiword sequences are idioms, defined as conspicuous sequences of words whose meaning cannot be derived based on the literal meaning of their constituent parts (Birch, 2014; Coxhead & Byrd, 2007; Howarth, 1996; Omazic, 2008; Simpson-Vlach & Ellis, 2010). For example, an idiom like *by and large* or an idiomatic expression like *on the other hand* does not allow any lexical or syntactic variations (Weinert, 1995; Willis, 2003). Other entirely fixed ones include phrases subject to syntactic constraints that restrict either their exchangeability (e.g. *up here* but not **here up*; *blind alley* but not **the alley is blind*) or their usage such as *I guess*, which cannot be negated **I don't guess* (Erman & Warren, 2000; Howarth, 1996).

Formulaic sequences can also be partially fixed in that they contain open slots that can be filled with semantically related words (Schmitt & Carter, 2004; Weinert, 1995; Willis, 2003). For example, they can be composed of a continuous string of words (e.g. *Would you mind* ____), a discontinuous string of words (e.g. *carry the* ____

responsibility), or a fixed stem with a variable slot, as in *single mother* in which *single* can be substituted with *queen, his,* etc. (Howarth, 1996; Nattinger & DeCarrico, 1992; Philip, 2008; Schmitt & Carter, 2004; Willis, 2003). This total or partial fixedness is believed to be a major defining characteristic that distinguishes formulaic sequences from other creatively constructed expressions: "for anything to be a prefab the choice of one word must determine or, at least definitely restrict, the choice of at least one other, normally adjacent, word" (Erman & Warren, 2000, p. 32).

Frequent occurrence and native-speaker preference. A third defining characteristic of formulaic sequences is frequency of occurrence (Simpson-Vlach & Ellis, 2010; Weinert, 1995). This characteristic is of crucial importance in the case of lexical bundles; that is, a sequence of words has to pass a minimum frequency cutoff—often between 10 and 40 occurrences per one million words—in a particular register to be classified as a lexical bundle (Biber & Barbieri, 2007; Chen & Baker, 2010; Cortes, 2002, 2004; Hyland, 2008a; Liu, 2012). *On the other hand*, for example, has 726 occurrences in the 3.5 million-word corpus analyzed in Hyland (2008a) and 353 occurrences in Byrd and Coxhead's (2010) OWL corpus. Frequency of occurrence has also been used as a defining characteristic in some studies that aimed at creating lists of formulaic sequences, such as Simpson-Vlach and Ellis's (2010) Academic Formulas List (AFL).

However, a string of words may not pass a frequency threshold and may still be considered formulaic (Durrant & Mathews-Aydınlı, 2011; Schmitt et al., 2004; Wray, 2000, 2002). The frequency of such formulaic sequences as *kick the bucket*, *long live the king*, *I want to marry you*, etc. may be very low even in large corpora because of their limited use (Durrant & Mathews-Aydınlı, 2011; Pawley & Syder, 1983; Wray, 2002).

Therefore, native-speaker preference of certain sequences over other equally possible grammatical constructions is viewed as another defining characteristic that helps distinguish a formulaic sequence from other creatively constructed sequences of words (Durrant & Mathews-Aydınlı, 2011; Weinert, 1995; Wray, 2002).

Being register-bound. Register sensitivity is another characteristic that determines the formulaic status of a multiword sequence; multiword sequences are highly register-bound in that the ones used in speech differ in structure and function from those used in writing (Biber, 2009; Biber et al., 1999; Chen & Baker, 2010; Cortes, 2008; Liu, 2012). This defining characteristic has been highlighted in research studies that have attempted to analyze the syntactic structure of the recurring lexical bundles in speech and/or writing (e.g. Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Biber et al., 1999; Chen & Baker, 2010; Hyland, 2008a). The lexical bundles used in speech are for the most part fixed clauses that execute stance functions, and the ones utilized in writing are mainly discontinuous phrases with variable slots and serve referential functions (Biber, 2006, 2009, 2010; Biber & Barbieri, 2007; Biber et al., 2004; Cortes, 2004, 2008).

Based on the above-listed definitions and defining characteristics, multiword sequences refer to a language phenomenon that has several defining characteristics other than being orthographically composed of multiple words (Ellis, Simpson-Vlach, & Maynard, 2008; Erman & Warren, 2000; Millar, 2011). This language phenomenon has been at the center of research on writing production, which has empirically demonstrated its recurrence in academic prose and, hence, has revolutionized the long prevailing view of the nature of academic writing as being entirely creative.

The Formulaic Nature of Academic Writing: Overview and Evidence

Until recently, a considerable number of linguists argued in favor of *the creative aspect of language use*, also known as *the Chomskyan view of language*, which perceived creative language production as a major distinctive and striking feature of human language (Chomsky, 1982, 2006; Cook, 1997; Fromkin, Rodman, & Hyams, 2003; Howarth, 1998a; Stubbs & Barth, 2003). This view of language was mainly grounded in a human faculty that enables language users, in principle, to generate and comprehend an infinite number of novel utterances based on a finite set of syntactic rules (Chomsky, 1982, 2006; Cook, 1997; Fromkin et al., 2003; Howarth, 1998a; Stubbs & Barth, 2003). *The creative aspect of language use* was originally proposed to describe *linguistic competence*, that is, the system of linguistic rules that enables language users to produce and comprehend language (Chomsky, 1982; Fromkin et al., 2003) and was, then, extended to describe speech production:

> much of what we say in the course of normal language is entirely new, not a repetition of anything we have heard before and not even similar in pattern… to sentences or discourse that we have heard in the past. This is a truism, but an important one, often overlooked and not infrequently denied in the behaviorist period of linguistics…, when it was almost universally claimed that a person's knowledge of a language is representable as a stored set of patterns, overlearned through constant repetition and detailed training, with innovation being at most a matter of analogy. (Chomsky, 2006, p. 10)

The creative view of language production was also extended by its proponents to describe academic writing. To them, academic writing includes a very limited number of

conventional forms, if any, because academics are '*cliché-phobes*,' and, hence, they exert a lot of effort writing and polishing their work to produce a genuine piece that is void of heavily used and/or common expressions in academic discourse (Howarth, 1998a).

However, after reestablishing the nature of speech production as being at least partially formulaic, a considerable number of linguists attempted to theoretically challenge the creative view of academic writing (Barlow, 2000; Cowie, 1992; Howarth, 1998a; Lewis, 1997; Liu, 2012; Pawley & Syder, 1983; Weinert, 1995). Among the early underlying arguments against the absolute creativity of academic writing was that of Howarth (1998a): "the expression of fresh or individual ideas does not entail the generation of novel word-combinations from scratch; originality of thought is not incompatible with familiarity of expression" (p. 165). Such arguments were, however, challenged based on the assumption that they imply reducing the intricate process of language learning "to phrase-book memorization and repetition" (Howarth, 1998a, p. 165).

The formulaic-creative dichotomy remained unsolved until the creative view of academic writing was empirically challenged by a new body of empirical research, which emerged in the late 20th century and early 21st century as a result of the advent of corpus linguistics (Coxhead, 2008; Ellis, 2012; Erman, 2007; Erman & Warren, 2000; Flowerdew, 2015; Stubbs, 2002; Stubbs & Barth, 2003).

Corpus linguistics. Corpus linguistics is a cornerstone in applied linguistics research that focuses on "the study of language based on examples of real life language use" (McEnery & Wilson, 1996, as cited in Baker, 2006, p. 1). In Biber's (2000) words, corpus linguistics is mainly "concerned with actual practice, and the extent to which

linguistic patterns are common or rare, rather than focusing exclusively on potential grammaticality" (p. 287). Since it is a methodological approach that is based on qualitative and quantitative analyses of large collections of electronic spoken and/or written texts, corpus linguistics research, whose starting point of analysis is the lexicon, has informed the field of applied linguistics by revealing the distinctive linguistic features which characterize language use (Biber, 2000, 2010; Biber & Conrad, 1999, 2001; Flowerdew, 2015; Gilquin, Granger, & Paquot, 2007; Hüttner, 2010; Schmitt & Carter, 2004). In other words, corpus linguistics researchers set out to empirically uncover surprising linguistic features that may be partially or fully obscure if applied linguists were to solely rely on intuition (Biber, 2010; Hüttner, 2010).

Corpus linguistics and formulaic sequences. At the core of corpus linguistics research is the recurrence of formulaic sequences in academic writing (Biber, 2010; Flowerdew, 2015; Gries, 2008; Howarth, 1996; Pecorari, 2009; Read & Nation, 2004). Formulaic sequences, which received only minimal attention for decades, were brought to the forefront with (a) the advances in corpus linguistics that were associated with a burst of new linguistic research studies, (b) the development of computer programs that facilitated rigorous analyses of large corpora, (c) applied linguists' belief that intuition alone does not suffice to reveal the linguistic features that characterize language use, and (d) the need for empirical evidence that thoroughly describes the frequently occurring patterns in language production, especially academic writing (Biber, 2010; Cortes, 2002, 2008). Most corpus research studies on formulaic sequences and academic writing have adopted one of the two major approaches to textual analysis: the corpus-driven approach or the corpus-based approach (Biber, 2009; Cortes, 2004).

Corpus-driven research is, in itself, a methodology in which a corpus serves as the basis on which applied linguists empirically develop new linguistic constructs that have not been previously identified in linguistic theories (Biber, 2009, 2010; Tognini-Bonelli, 2001). This approach to analysis is inductive in that it is neither driven by nor founded on a priori assumptions (Biber, 2009, 2010; Moon, 1998; Tognini-Bonelli, 2001). Rather, it views texts as composed of individual lexical words, and new linguistic categories, patterns, and constructs are inductively identified based on corpus observations (Biber, 2009; Tognini-Bonelli, 2001). The difference between corpus-driven research and other forms of analysis is inherent in three major defining characteristics: corpus-driven research focuses on the analysis of word forms; its major objective is to empirically identify recurrent multiword sequences; and it places very limited, if any, focus on the syntactic structures of words during the analysis stage (Biber, 2009).

The corpus-based approach, in turn, is typical of studies that mainly focus on the grammatical features of language or languages (Biber, 2009, 2010). It is deductive in that the analysis of a corpus or corpora is guided by structures that have been previously defined in linguistic theory in order to confirm or refute a linguistic construct (Biber, 2009, 2010; Biber & Conrad, 1999, 2001). In other words, corpus-based linguists analyze a corpus to empirically identify the frequency, variation, and use of predefined language features in order to validate a linguistic theory without focusing on any other language features that may emerge inductively (Biber, 2009, 2010; Tognini-Bonelli, 2001).

In research on recurrent multiword sequences in academic writing, many corpus linguists have adopted the corpus-driven approach to analysis—having words as their starting point of analysis to empirically identify recurrent word strings and inductively

document new linguistic constructs (Biber, 2009, 2010). Other linguists have opted for the corpus-based approach to validate the frequency and usage of predefined lists of formulaic sequences and reveal their variations in a particular register (Biber, 2009, 2010; Moon, 1998).

Major findings of corpus linguistics research. Corpus linguistics research on formulaic sequences and academic writing has put forward three major findings: the ubiquity of formulaic sequences in academic writing, their role in distinguishing expert writers from novice ones, and the constraints of different registers on their use.

The pervasiveness of formulaic sequences in academic writing. Corpus linguistics research has empirically demonstrated that academic writing is far from being composed afresh on the basis of a finite set of syntactic rules as claimed by proponents of *the creative aspect of language use*; instead, between one third and one half of academic writing consists of different types of formulaic sequences (Byrd & Coxhead, 2010; Cortes, 2004; Erman & Warren, 2000; Hyland, 2008a, 2008b; Martinez & Schmitt, 2012; O'Keefe et al., 2007; Römer, 2009). Moreover, the pervasiveness of formulaic sequences is neither accidental nor confined to corpus frequency, but it is rather a by-product of the important functions they execute in academic writing (Biber & Barbieri, 2007; Cortes, 2004; Hyland, 2008a, 2008b; Römer, 2009). These findings were reported in corpus linguistics research studies that adopted the corpus-driven and the corpus-based approaches to textual analysis.

Among the corpus-driven research studies that empirically demonstrate the formulaic nature of academic writing is that of Hyland (2008a), in which he analyzed the forms, functions, and structures of recurrent four-word lexical bundles in a 3.5 million-

word corpus of published research, M.A. theses, and doctoral dissertations in four different disciplines. Following a minimum frequency cut-off of 20 occurrences per one million words and an occurrence in at least 10% of the collected texts, Hyland (2008a) empirically demonstrates that four-word lexical bundles are highly frequent in academic prose and have important discourse functions that are peculiar to each target discipline.

Similar results were proposed in Cortes's (2008) study, in which she explored the recurrence of four-word lexical bundles in published history research in English and Argentinian Spanish. Both corpora, about one million words each, were analyzed to identify the most frequent lexical bundles that occurred 20 times in at least five texts, and the functions of these lexical bundles in both languages were, then, scrutinized. According to Cortes (2008), although published history research in Argentinian Spanish includes more lexical bundles than English published history research, the majority of the lexical bundles in both corpora serve very similar functions that are typical of history as a discipline. Not only does this study, as Cortes (2008) suggests, reveal the ubiquity of lexical bundles in academic prose, but it also demonstrates that lexical bundles are highly discipline-bound.

Similar results were proposed by Liu (2012) and Oakey (2002) despite the differences in the number of formulaic sequences investigated in their corpus-based research studies. Oakey's (2002) study attempted to empirically investigate the frequency of one formulaic sequence identified in Nattinger and DeCarrico's (1992) list of frequent phrases (i.e., *it is/has been (often) asserted/believed/noted that X*) in a mega corpus. Oakey (2002) analyzed the frequency of the target phrase in a sub-set of the British National Corpus (BNC) and concluded that the target phrase frequently occurred in three

different disciplines and executed important pragmatic functions. Liu (2012), in turn, examined the occurrence and distribution of formulaic sequences in two general academic writing corpora—the academic writing sub-corpora of both the Corpus of Contemporary American English (COCA) and the BNC—using predefined lists of formulaic sequences in the literature. One of the major findings yielded in Liu's (2012) study is that most of the formulaic sequences identified in previous research have high frequency in the sub-corpora of general academic writing.

The findings obtained from corpus-based and corpus-driven research, therefore, buttress arguments of opponents of absolute creativity—referred to by Sinclair (1991) as the open-choice principle (i.e., using individual words and syntactic rules to construct phrases and sentences afresh each time language is produced). These findings empirically prove that academic writing is formulaic and, thus, is inherently different from other forms of written communication (e.g. creative writing), in which a writer's success is often contingent on his/her ability to express well-known and newly proposed concepts in a novel and unique manner (Ädel & Erman, 2012; Biber, 2006; Conklin & Schmitt, 2008; Gledhill, 2000; O'Keefe et al., 2007; Wray, 2002).

The role of formulaic sequences in identifying expert writers. Given their ubiquity in academic writing, the proper use of formulaic sequences in academic writing is part and parcel of identifying language users as proficient writers (Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a; Kazemi et al., 2014; Lewis, 2000a; Wood, 2015). Such a role is corroborated in corpus-driven comparative research studies that have contrasted expert and novice writers' use of formulaic sequences. These studies (e.g. Chen & Baker, 2010; Cortes, 2004; Durrant & Schmitt, 2009; O'Donnell et al., 2013)

have revealed key differences between the writing production of expert and novice writers (both L1 and L2 speakers of English). When compared to expert writers who use various types of formulaic sequences, novice writers tend to overuse a limited number of highly frequent formulaic sequences, underuse a wide range of formulaic sequences utilized by expert writers, or misuse this language phenomenon in their writing production (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2002, 2004; Durrant & Mathews-Aydinli, 2011; Gilquin et al., 2007; Granger, 1998; O'Donnell et al., 2013).

Cortes (2004), for example, compared the use and functions of lexical bundles in published biology and history research to those in the writing of undergraduate and graduate biology and history students. To this end, she compiled three corpora from published research and university students' writing, extracted the four-word lexical bundles that occurred 20 times per one million words, classified the elicited lexical bundles structurally and functionally, and, finally, compared the functions of lexical bundles used by university students to those utilized by expert writers. Cortes (2004) clarified that the majority of the recurrent lexical bundles in published research were hardly used by university students who repeatedly drew on a limited number of lexical bundles, and the functions of the lexical bundles used by students were overtly different from those identified in published research. Based on these major findings, Cortes (2004) concludes that biology and history student writing could be generally described as repetitive and lacking proficiency and idiomaticity.

Very similar results were reported in Chen and Baker (2010) despite the differences in their corpora and methods. Chen and Baker (2010) identified the use and functions of lexical bundles in a corpus composed of expert and novice writing (both L1

and L2). After a detailed analysis of the three corpora, Chen and Baker (2010) explained that the highest number of lexical bundles was detected in published research followed by L1 student writing and then L2 student writing, suggesting that "the use of formulaic expressions grows with writing proficiency" (p. 43). Moreover, neither L1 nor L2 student writers used the lexical bundles that were employed by expert writers. The student writers, instead, overused other bundles that were not used by expert writers, which the researchers considered as signs of lack of proficiency. Chen and Baker (2010) concluded that although the L1 students in their study exhibited higher levels of writing proficiency than the L2 students, "there was a gap, in terms of the use of lexical bundles, between native expert academic writing and student writing" (p. 44).

Durrant and Schmitt (2009), in turn, investigated the use of collocations by L1 and L2 university writers. The corpora for their study included research papers submitted by both undergraduate and graduate L2 students, argumentative essays culled from the International Corpus of Learner English, L1 graduate students' assignments, and L1 students' argumentative essays. After extracting the most frequent collocations from their corpora, Durrant and Schmitt (2009) checked the frequency of the identified collocations in the BNC. The results of this research study denoted that L1 writers successfully utilized low frequency collocations typical of academic writing. On the other hand, L2 writers repeatedly used high frequency collocations but underused a whole set of collocations that function as important building blocks in academic writing. Durrant and Schmitt (2009) concluded that the lack of idiomaticity in L2 writing could be ascribed to L2 users' attempt to cling onto a limited set of collocations they had already mastered rather than to their tendency to avoid them altogether.

O'Donnell et al. (2013) compared the use of formulaic sequences by L1 and L2 students at different university levels. They identified three-, four-, and five-word formulaic sequences in both L1 and L2 writing of university students at different levels and compared these sequences to the ones used in the 61 expert writing texts investigated in another research study. The results of the study suggested that expert and graduate student writing was dense with formulaic sequences. Moreover, when compared to expert writers, both L1 and L2 novice writers used fewer formulaic sequences, and some of which were not typical of academic writing.

Taken together, the findings of the above-reviewed studies—along with other corpus-based research that analyzed expert or novice writing alone or synthesized the findings of such research (e.g. Cortes, 2002, 2008; Flowerdew, 2015; Gilquin et al., 2007; Hyland, 2008a)—indicate that the appropriate use of a wide variety of multiword sequences is a defining characteristic of expert writing; however, such usage is a major stumbling block for novice writers, especially L2 novice writers.

The restricted use of multiword sequences. Corpus linguistics research has also demonstrated that the use of formulaic sequences is confined to the target register, discipline, and genre (Biber et al., 1999; Durrant, 2017; Howarth, 1996; O'Keefe et al., 2007; Pérez-Llantada, 2014). That is, corpus-driven research studies, which have investigated the use of multiword sequences in different registers, have identified two major differences between the ones used in writing and those used in speech. The first major difference relates to the syntactic structure of multiword sequences based on register: while the ones employed in speech are clausal in their structure, the ones utilized

in academic writing are mainly phrasal (Biber & Barbieri, 2007; Biber et al., 1999; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a).

Another major register-based difference is the relative creativity of the formulaic sequences used in academic writing (Liu, 2012; Oakey, 2002). That is, fixed expressions were very rare in general academic writing, indicating that most of the ones used in academic writing have slots that could be creatively filled with semantically related words (Liu, 2012; Oakey, 2002). Other corpus studies which focus on the use and functions of formulaic sequences in different disciplines have suggested that this language phenomenon has different distributions in different disciplines and, thus, is discipline- and/or genre-bound (Cortes, 2004, 2008; Durrant, 2017; Hyland, 2008a; O'Keefe et al., 2007; Pérez-Llantada, 2014). In other words, different writers draw on a shared communicative repertoire of multiword sequences when they write within the same discipline and/or the same genre (Cortes, 2004, 2008; Durrant, 2017; Hyland, 2008a; Oakey, 2002; Pérez-Llantada, 2014). In this respect, multiword sequences "are not only central to the creation of academic discourse, but they offer an important means of differentiating written texts by discipline" (Hyland, 2008a, p. 4).

In brief, corpus linguistics research has established the ubiquity of formulaic sequences in academic writing, especially by expert writers, has highlighted their restricted use as being highly discipline-, genre-, and register-bound, and has revealed the fallacy of the notion of their syntactic and/or lexical fixedness (Durrant, 2017; Granger & Paquot, 2008; Moon, 1998; Oakey, 2002; Philip, 2008). This body of research has also examined the syntactic structures and discursive functions of recurrent formulaic

sequences in academic writing and has drawn up multiple lists that encompass what is believed to be distinct types of this language phenomenon.

Types of Multiword Sequences in Academic Writing

Experts in the field of discourse analysis have categorized formulaic sequences based on different dimensions. These categorizations have brought about various lists, of which four major lists are discussed in this section: formulaic sequences, lexical bundles, collocations, and lexicogrammatical sequences.

Formulaic sequences. A major list that categorizes frequently occurring formulaic sequences in academic discourse is Simpson-Vlach and Ellis's (2010) AFL—which comprises 500 empirically derived formulaic sequences. The AFL categorizes formulaic sequences into three major functional groups based on their predominant pragmalinguistic functions in both spoken and written academic registers (Ellis et al., 2008; Simpson-Vlach & Ellis, 2010). It is worth noting that only the formulaic sequences that frequently occur in academic writing are tackled in this section.

Referential expressions. The first and largest group of formulaic sequences in Simpson-Vlach and Ellis's (2010) AFL encompasses *referential expressions*. This group includes five categories that execute various functions in academic writing (Simpson-Vlach & Ellis, 2010). *Specification of attributes* formulas, the first category, include (a) *intangible framing attributes* which constitute the largest subcategory of specification of attributes and are used either to present the defining characteristics of the following noun phrase (e.g. *the notion of*) or clause (e.g. *the way in which*) or to link two elements of discourse, typically a verb with the following clause (e.g. *in such a way*); (b) *tangible framing attributes* that present a tangible characteristic (i.e., size, amount, etc.) of the

following noun (e.g. *the amount of*); and (c) *quantity specification formulas* which refer to the quantity of a noun and can be either anaphoric in that they describe the preceding noun (e.g. *both of these*) or cataphoric and, hence, describe the quantity of the following noun, as in *a wide range of* (Simpson-Vlach & Ellis, 2010).

Identification and focus is the second category of *referential expressions* and includes formulaic sequences that help writers exemplify or explain a proposed concept in writing, as in *as an example* and *this means that* (Simpson-Vlach & Ellis, 2010). *Contrast and comparison*, the third category of *referential expressions*, includes formulaic sequences that help compare and contrast different discourse elements in writing as in *similar to those, on the other hand,* etc. (Simpson-Vlach & Ellis, 2010). *Deictics and Locatives* and *vagueness markers* are the last two categories of *referential expressions*. While the former is used to refer to 'physical,' 'temporal,' or 'spatial' elements (e.g. *at the time of*, *at this stage*, *b and c*, etc.), the later is utilized to make a vague reference in written texts, as in *and so on* (Simpson-Vlach & Ellis, 2010).

Stance expressions. The second group of formulaic sequences in the AFL, *stance expressions*, includes six categories that carry out different pragmatic functions in academic writing (Simpson-Vlach & Ellis, 2010). These functional categories encompass (a) *hedges* which include formulaic sequences that present a degree of tentativeness (e.g. *it is likely that*); (b) *epistemic stance* formulaic sequences which help the writer objectively express his/her attitude toward a proposed notion (e.g. *be argued that*); (c) *obligation and directive* formulaic sequences which help the writer propose recommendations for the reader or reiterate an already proposed concept (e.g. *to ensure that the*); (d) *ability and possibility* formulaic sequences which are utilized to present

actions and propositions that range in their degree of possibility (e.g. *be achieved by*, *be used as a*, etc.); (e) e*valuation* formulaic sequences whose main function is to provide a non-directive evaluation of a proposed concept in writing (e.g. *it is difficult*); and (f) *intention/volition* formulaic sequences which are often used to present the writer's intention, as in *to do so* (Simpson-Vlach & Ellis, 2010).

Discourse organizing expressions. The last group of formulaic sequences consists of *discourse organizing expressions* that include four major categories (Simpson-Vlach & Ellis, 2010). The first category, *metadiscourse and textual reference*, includes formulaic sequences that academic writers use to make reference to other elements in a text (*as shown in*, *in this article*, etc.); the second category, *topic introduction and focus*, is composed of formulaic sequences whose function is very similar to *identification and focus* formulas but differ from the latter in that "the global discourse organizing function of introducing a topic is primary here, with the phrase often framing an entire clause or upcoming segment of discourse," as in *what are the* (Simpson-Vlach & Ellis, 2010, p. 507). *Topic elaboration* formulaic sequences constitute the third category and include formulaic sequences that are utilized in academic writing to elaborate on a proposed concept (e.g. *are as follows*) or to reveal the causal/resultative relationships between different discourse elements as in and *as a result of the* (Simpson-Vlach & Ellis, 2010). *Discourse markers*, the last category of *discourse organizing expressions*, help smooth transition between discourse elements, as in *in other words, as well as*, etc. (Simpson-Vlach & Ellis, 2010).

Lexical bundle. In addition to the major classification of recurrent formulaic sequences in academic writing, several scholars have attempted to classify lexical

bundles based on their distinct syntactic structures and various discursive functions in different types of academic texts, resulting in three major taxonomies, namely, structural, functional, and research-based.

The structural taxonomy of lexical bundles. The structural classification of lexical bundles was originally proposed by Biber et al. (1999), and it has been adopted in studies attempting to structurally classify recurrent lexical bundles in academic writing (e.g. Biber & Conrad, 1999; Chen & Baker, 2010; Cortes, 2002, 2004; Hyland, 2008a, 2008b). Biber et al. (1999) propose 12 structural types of lexical bundles that are typical of academic writing, which are also broadly grouped into three major categories (Biber, 2009; Chen & Baker, 2010).

Lexical bundles that incorporate verb phrase fragments. The first structural category comprises lexical bundles that are mainly composed of verb phrase fragments (Biber et al., 1999). Among the lexical bundles that belong to this category are the ones that consist of a verb phrase and an anticipatory *it* (e.g. *it can be seen*, *it should be noted*, etc.) or a passive verb phrase with a prepositional phrase (e.g. *is shown in figure*, *is referred to as*, etc.); other lexical bundles that are composed of *verb to be* and a noun or an adjective phrase (e.g. *is part of the* and *may be due to* respectively) also belong to this category (Biber & Conrad, 1999; Biber et al., 1999; Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2002, 2004; Hyland, 2008a).

Lexical bundles that incorporate dependent clause fragments. The second category in the structural taxonomy includes lexical bundles that begin or end with *complementizers* (Byrd & Coxhead, 2010; Cortes, 2002; Hyland, 2008b). This category, for example, includes lexical bundles that are composed of a verb/adjective + to-clause

fragment, as in *are likely to be*, a prepositional phrase + to-clause fragment, as in *in order to make*, or a that-clause fragment, as in *that there is a* (Biber, 2009; Biber et al., 2004; Biber et al., 1999; Byrd & Coxhead, 2010; Chen & Baker, 2010; Hyland, 2008a).

Lexical bundles that incorporate noun and prepositional phrase fragments. The third category of the structural taxonomy, which is also referred to as NP-based bundles, includes lexical bundles that are composed of phrasal rather than clausal elements (Biber et al., 1999; Cortes, 2004). This category incorporates lexical bundles that are composed of a noun phrase (e.g. *the fact that the*), a prepositional phrase (e.g. *by the fact that*), a noun phrase and an of-phrase fragment (e.g. *the end of the*, *the base of the*, etc.), or a noun phrase with a post-modifier fragment (e.g. *the way in which*, *the extent to which*, etc.); other types of lexical bundles that belong to this structural category link two prepositional phrases or two elements of discourse, as in *as a result of* and *as well as the* (Biber, 2009; Biber & Conrad, 1999; Biber et al., 1999; Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2002, 2004; Hyland, 2008a).

The functional taxonomy of lexical bundles. Lexical bundles have also been classified based on their discourse functions in academic writing. The function-based classification of lexical bundles, which echoes the Hallidayan metadiscourse functions of language (Ädel & Erman, 2012; Biber, 2009; Cortes, 2004; Hyland, 2008b), includes three main categories under which several subcategories have been grouped.

Referential bundles. The first and most widespread type of lexical bundles in academic writing encompasses *referential bundles* that help writers make reference to physical or abstract entities, refer to other discourse elements, or signal important characteristics of discourse elements in a text (Ädel & Erman, 2012; Biber, 2009; Byrd &

Coxhead, 2010; Cortes, 2004). This type of lexical bundles includes (a) *framing bundles* which highlight the characteristics of the following noun (e.g. *the nature of the*, *the existence of a*, etc.); (b) *quantifying bundles* which help writers specify measurable attributes of a discourse element (e.g. *a number of the*, *a wide range of*, *the extent to which*, etc.); and (c) *place/time/text deictic bundles* which refer to place, time, and other textual elements in a text, as in *the center of the*, *at the beginning of*, and *as shown in fig.*, respectively (Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2004).

Stance bundles. The second category in the functional taxonomy of lexical bundles includes *stance bundles*, also referred to as *expressions of attitude* (Ädel & Erman, 2012; Biber, 2009; Byrd & Coxhead, 2010; Cortes, 2004). They include three subcategories and are used to present epistemic evaluations of different discourse elements (Ädel & Erman, 2012; Biber, 2009; Cortes, 2004). *Epistemic bundles*, the first subcategory of *stance bundles*, are used to objectively evaluate different discourse elements with various degrees of certainty, as in *seems to have* and *the fact that the* (Chen & Baker, 2010; Cortes, 2004). *Obligatory directive* bundles constitute the second subcategory; they help evince the writer's attitude towards or recommendations about a proposition (e.g. *it is necessary to* and *it has to be*); *ability bundles*, the last subcategory of *stance bundles*, convey the writer's evaluation of the possibility of performing an action, as in *will be able to*, *it is difficult to*, etc. (Chen & Baker, 2010).

Discourse organizers. The last category of the functional taxonomy includes lexical bundles that are chiefly used to structure academic texts by linking prior and upcoming discourse elements (Ädel & Erman, 2012; Biber, 2009; Chen & Baker, 2010; Cortes, 2004). These lexical bundles are divided into three categories. *Topic introduction*

and *topic elaboration* bundles are essentially used to introduce and elaborate on discourse elements, as in *last but not least* and *be taken into account* (Chen & Baker, 2010). *Inferential bundles*, in turn, are used to make inferences, as in *as a result of*, *in the sense that*, etc. (Chen & Baker, 2010; Cortes, 2004). *Identification/focus* bundles, the last subcategory of *discourse organizers*, are employed to direct the reader's attention to the following noun phrase or signal the elements which the writer believes should receive special focus, as in *one of the most* and *bear in mind that* (Chen & Baker, 2010; Cortes, 2004).

The research-based taxonomy of lexical bundles. The third taxonomy of recurring lexical bundles in academic writing is Hyland's (2008a, 2008b) research-based taxonomy. This taxonomy is similar to the functional taxonomy in that they both propose three major categories based on the broad Hallidayan metadiscourse functions of language (Hyland, 2008a, 2008b). However, they differ in that the former focuses on the discursive functions of lexical bundles in research and the latter investigates the functions that lexical bundles execute in different registers (Hyland, 2008a).

Research-oriented bundles. The first category in Hyland's (2008a, 2008b) taxonomy includes *research-oriented bundles* whose main function is ideational; they are used to "help writers to structure their activities and experiences of the real world" (Hyland, 2008a, p. 13). Hyland (2008a, 2008b) subcategorizes these lexical bundles into (a) *location bundles* which specify time or place (e.g. *at the beginning of*), (b) *procedure bundles* which present a research procedure (e.g. *the role of the*), (c) *quantification bundles* which identify the quantity of the following head noun (e.g. *the magnitude of*), (d) *description bundles* which present the characteristics of the following noun (e.g. *the*

structure of the), and (e) *topic bundles* which refer to a research environment (e.g. *the currency board system*).

Text-oriented bundles. The second category of Hyland's (2008a, 2008b) taxonomy includes *text-oriented bundles* that help writers present different discourse elements in a text. These bundles comprise (a) *transition signals* that link discourse elements either by adding new information or by contrasting two discourse elements in a text, as in *in addition to* and *on the other hand*; (b) *resultative signals* which help writers present inferred information (e.g. *these results suggest that*) and establish causative relations between discourse elements in a text (e.g. *as a result of*); (c) *structuring signals* which are used to organize or direct readers to textual elements (e.g. *as shown in fig*., *in the next section*); and (d) *framing signals* which help situate specific points of discussions, as in *with respect to the*, *on the basis of*, etc. (Hyland, 2008a, 2008b).

Participant-oriented bundles. The lexical bundles that belong to the last category in Hyland's (2008a, 2008b) taxonomy focus either on the writer or on the reader and are divided into *stance features* and *engagement features*. While *stance features* convey the writer's objective evaluation of or attitude towards existing discourse elements (e.g. *are likely to be*, *it is possible that*, etc.), *engagement features* are utilized in academic writing as means of addressing a reader in an impersonal way, as in *it should be noted* and *as can be seen* (Hyland, 2008a, 2008b).

Collocations. The third classification of recurrent multiword sequences in academic writing is that of collocations. Collocations are categorized in the literature into lexical collocations and grammatical collocations (Bahns, 1993; Coxhead & Byrd, 2007; Gledhill, 2000; Howarth, 1996; Laufer & Waldman, 2011; Lewis, 2000b; Martin, 2008).

Lexical collocations include pairs of words that are constructed of two content words, that is, nouns, verbs, adjectives, and adverbs (Bahns, 1993; Durrant, 2009; Howarth, 1996). A *lexical collocation*, for example, can be constructed of a verb and a noun (e.g. *suggest an alternative*), an adjective and a noun (e.g. *a crushing defeat*), two nouns (e.g. *a pride of lions*) or an adverb and an adjective (e.g. *deeply absorbed*), among other possible combinations (Bahns, 1993; Lewis, 2000b). *Lexical collocations* can also have variable positions in that they may be separated by other lexical or grammatical words (Durrant, 2009; Howarth, 1996). For instance, a *lexical collocation* that is composed of an adjective and a noun (e.g. *a powerful argument*) can be separated by content and grammatical words, as in *a powerful but ultimately convincing argument* (Durrant, 2009).

Grammatical collocations, in turn, are composed of a lexical word and a grammatical word, such as a preposition, a determiner, a clause, etc. (Bahns, 1993; Durrant, 2009; Howarth, 1996). Among the *grammatical collocations* are those that consist of a verb and a preposition (e.g. *account for*), an adjective and a preposition (e.g. *adjacent to*), an infinitive and an adjective (e.g. *to be afraid*), a verb and a that clause (e.g. *assume that…*), etc. (Bahns, 1993; Durrant, 2009; Lewis, 2000b).

Lexicogrammatical sequences. The last list of recurrent multiword sequences in academic writing includes eight distinct types that are identified by different researchers based on their syntactic constituents and are proposed as a list by Coxhead and Byrd (2007). It is worth mentioning that only six types are discussed in this section because two of the proposed types in Coxhead and Byrd's (2007) list (i.e., lexical bundles and collocations) have been discussed above**.**

Phrasal verbs and prepositional verbs. The first type in Coxhead and Byrd's (2007) list includes *phrasal verbs* and *prepositional verbs*. These sequences include semantically idiosyncratic structural units whose meaning is independent of the individual words that constitute them (Biber et al., 1999; Coxhead & Byrd, 2007; Fromkin et al., 2003; Lewis, 2000b; Willis, 2003). In other words, what distinguishes *phrasal verbs* and *prepositional phrases* from free combinations is that "with free combinations, each word has an independent meaning, while the meaning of multi-word verbs often cannot be predicted from the individual parts" (Biber et al., 1999, p. 404). While *phrasal verbs* are composed of a lexical verb and an adverbial particle (e.g. *look up, knock over*, set off, carry out, etc.), *prepositional verbs* consist of a syntactically related lexical verb and a preposition such as *approve of, deal with, listen to,* etc. (Biber et al., 1999; Carter & McCarthy, 2006; Coxhead & Byrd, 2007; Howarth, 1998b; Hyland, 2008a; Willis, 2003).

Idioms. The second type in Coxhead and Byrd's (2007) list encompasses the most conspicuous and the least recurring sequences of words and/or expressions whose use is indicative of advanced proficiency, that is, *idioms* (Bahns, 1993; Coxhead & Byrd, 2007; Hyland, 2008b; Lewis, 2002; Omazic, 2008; Simpson-Vlach & Ellis, 2010). *Idioms* can be constituted of different syntactic structures; for example, an idiom can be composed of a wh-question (e.g. *How do you do?* and *What on earth?*), a noun phrase as in *a piece of cake* and *a slap in the face*, a verb phrase (e.g. *to scream blue murder* and *beat around the bush*), or a prepositional phrase such as *in a nutshell* and *as a matter of fact* (Bahns, 1993; Biber et al., 1999; Hyland, 2008b).

Verb features and valency patterns. *Verb features* and *valency patterns* comprise sequences of words whose usage is constrained by certain characteristics of a lexical verb (Biber et al., 1999; Coxhead & Byrd, 2007). Since the choice of other clausal elements in a sentence is hemmed in by the potentials of a lexical verb, this type of patterned language is believed to reveal the strong ties between syntax and vocabulary (Biber et al., 1999; Coxhead & Byrd, 2007). *Verb features* refer to a characteristic feature in a verb that determines the grammatical class of its syntactically possible complement (Coxhead & Byrd, 2007). The majority of English lexical verbs, for example, can take either a that-clause (e.g. *speculate*, *argue*, etc.) or an infinitive (e.g. *fail*, *refuse*, etc.) as their complements (Biber et al., 1999; Coxhead & Byrd, 2007). *Valency patterns*, in turn, refer to the characteristic features of lexical verbs that determine the occurrence of other clause elements; that is, lexical verbs can be transitive (e.g. *help*), intransitive (e.g. *come*), or copular (e.g. *feel*), and, thus, they should be respectively used with an object, no object, and an adjective (Biber et al., 1999; Coxhead & Byrd, 2007).

Semi-fixed sequences. *Semi-fixed expressions* refer to sequences of words that are similar to *verb features* and *valency patterns* in that they reveal the strong links between words and grammar (Coxhead & Byrd, 2007). They, however, differ from the latter in that *semi-fixed expressions* are also subject to semantic constraints (Coxhead & Byrd, 2007). An example of *semi-fixed sequences* is the use of *may* with the modifier *well* (e.g. *may well be*), a combination that is determined by the semantic boundaries of the logical possibility modal *may* (Biber et al., 1999; Coxhead & Byrd, 2007).

Sentence frames and heads. The fifth type in Coxhead and Byrd's (2007) list includes *sentence frames* and *sentence heads*, which are also referred to in the literature

as *sentence builders*. *Sentence frames* and *heads* include both continuous and discontinuous strings of words that are utilized in academic writing as frames to structure phrases, clauses, and paragraphs (Coxhead & Byrd, 2007; Lewis, 1997; Willis, 2003). Among the short discontinuous sentence frames is *the___ of___*, which can be filled with different lexical words that belong to one grammatical category (i.e., nouns); long continuous ones include such sentence frames as *there is a growing body of evidence that* or *it has been asserted,* which are used to structure clauses and sentences (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Coxhead & Byrd, 2007; Lewis, 1997; Willis, 2003).

Binominal phrases. *Binomial phrases* incorporate patterned sets of grammatically parallel words that are coordinated by '*and*' or '*or*' (Biber et al., 1999; Coxhead & Byrd, 2007). These patterned sets can be composed of two nouns (e.g. *health and safety*), two verbs (e.g. *wait and see*), two adjectives as in *black and white*, or two adverbs as in *back and forth* (Biber et al., 1999; Coxhead & Byrd, 2007).

The overlap among the major classifications in the literature. A closer look at the terms, categories, subcategories, and examples reviewed above reveals a clear overlap among the proposed lists. Not only is such an overlap evident in the terms used by different scholars who claim to have adopted different approaches to categorizing and subcategorizing recurrent multiword strings in academic writing, but it is also obvious in the examples included in different lists that have been labeled differently.

The first overlap in the above-reviewed literature relates to the categories and subcategories used in the frequency-based, research-based, and pragmatic-based classification of multiword sequences in academic writing. While *stance bundles*,

discourse organizers, and *referential bundles* are introduced as the three main categories derived from the frequency-based classification of lexical bundles (e.g. Ädel & Erman, 2012; Biber, 2009; Chen & Baker, 2010; Cortes, 2004), Simpson-Vlach and Ellis's (2010) AFL introduces *stance expressions*, *discourse organizing expressions*, and *referential expressions* as the three major categories of the pragmatic-based classification of recurrent formulaic sequences in written academic discourse. In addition, most of the subcategories are recurrent in at least two lists (e.g. *specification of attributes, identification and focus, epistemic stance,* etc.), with one subcategory (i.e., *topic introduction*) co-occurring in the three approaches to classification (e.g. Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a, 2008b; Simpson-Vlach & Ellis, 2010).

The overlap in the literature can also be seen in the identical examples that are proposed in different lists. *(A) and (B)*, for instance, is referred to as a *binominal phrase* in Biber et al. (1999) and Coxhead and Byrd (2007) but as a *deictics and locatives* formulaic sequence in Simpson-Vlach and Ellis (2010). Moreover, one of the highly frequent *epistemic stance* formulas in Simpson-Vlach and Ellis's (2010) AFL (i.e., *assume that*) is listed as a collocation in Durrant's (2009) list. Other phrases have been identified in the literature with three different terms. *The (a) of (b)*, for example, is referred to as a *referential bundle* (e.g. Cortes, 2004), *sentence frame* (e.g. Coxhead & Byrd, 2007), and an *intangible framing attribute* (e.g. Simpson-Vlach & Ellis, 2010). The terms *discourse organizer*, *text-oriented bundle*, and *contrast and comparison*, in turn, are respectively used by Chen and Baker (2010), Hyland (2008a), and Simpson-Vlach and Ellis (2010) in order to categorize *on the other hand*.

Despite the differences in the terms and approaches used in the literature, which can be partly the result of researchers' tendency to build on the work of others, the overlap among the different categorizations reviewed above indicates that the different proposed lists are likely to represent the same pervasive language phenomenon that executes various discursive functions in academic writing. Because of this overlap and some researchers' tendency to adopt *formulaic sequences* as an umbrella term (e.g. Jones & Haywood, 2004; Schmitt & Carter, 2004; Wray, 2000), the term *formulaic sequences* will henceforth be used to collectively refer to different types of multiword sequences.

Formulaic Sequences and L2 Instruction: Rationale

The revolutionary results of corpus linguistics research—which have empirically demonstrated the ubiquity and pervasiveness of different types of formulaic sequences in language production—have had robust implications for SLA research and pedagogy, signaling a departure from already existing beliefs in the utility of several L2 teaching practices (Arnaud & Savignon, 1997; Boers & Lindstromberg, 2009; Zimmerman, 1997).

A cornerstone of this shift is *the Lexical Approach*. *The Lexical Approach* emerged in the early 1990s as a counteraction to the prevalent pedagogical approaches that solely focused on teaching syntactic rules and were influenced by the theoretical distinction between *linguistic competence* and *linguistic performance* (Hill, 2000; Lewis, 1997, 2002, 2008; Zimmerman, 1997). This theoretical distinction, often associated with Chomsky, differentiates the required linguistic knowledge to produce and comprehend language (i.e., *linguistic competence*) from the actual use of linguistic rules to produce language, that is, *linguistic performance* (Chomsky, 2006; Fromkin et al., 2003; Lewis, 2002). The prevalence of this theoretical distinction contributed to viewing language

learning as "synonymous with mastering the structures of the language, that is, achieving competence" (Lewis, 2002, p. 11) and, hence, placed syntactic structures at the core of L2 pedagogy (Hill, 2000; Lewis, 1997, 2002, 2008; Zimmerman, 1997).

The firm belief in the value and utility of structure-based instruction (a) prioritized grammar instruction as a result of considering it the backbone of language, the basis of creative language production, and the mainstay of native-like proficiency, (b) overlooked the actual usage of language by its native speakers, and (c) marginalized the role of vocabulary instruction in SLA research and methodology (Lewis, 2002, 2008; Willis, 2003; Zimmerman, 1997). Vocabulary instruction, especially formulaic sequences, took a back seat to other forms of instructions (e.g. structure-based, incidental learning, etc.) until the emergence of *the Lexical Approach*, which is compatible with *Communicative Language Teaching* (CLT): a meaning-focused and learner-centered approach that arouse in the late 1970s and aims at promoting L2 learners' communicative competence through engaging them in communicative activities (Ellis, 2003; Lewis, 2002, 2008; Spada, 2007; Zimmerman, 1997).

The Lexical Approach has several notions that revolve around three key concepts. The first key concept revolutionizes the long prevailing view of lexis as being solely composed of individual vocabulary words: lexical items "are **socially sanctioned independent units**. Many are words, but many consist of multi-word units" (Lewis, 2002, p. 90). The second concept invalidates the vocabulary/grammar dichotomy and the prevalent view of language production as being based on the construction of stretches of discourse by dint of laborious syntactic rules: "the basis of language is lexis. It has been, and remains, the central misunderstanding of language teaching to assume that grammar

is the basis of language communication" (Lewis, 2002, p. 133). The third major concept relates to the importance of assigning lexis, especially formulaic sequences, a more prominent status than structure because of its communicative power: "more meaning is carried by lexis than grammatical structure. Focus on communication necessarily implies increased emphasis on lexis, and decreased emphasis on structure" (Lewis, 2002, p. 33).

The Lexical Approach, however, has not denied the effectiveness of focusing on structure altogether; it has rather reformed the widespread and misguided belief in the pedagogical utility of solely focusing on syntactic rules: "Although it is possible to communicate quite complex messages through lexis alone it is self-evident that the ability to grammaticalise is important. Any approach to language teaching which emphasises lexis and de-emphasises grammar represents not a revolution, but a change of focus" (Lewis, 2002, p. 133). *The Lexical Approach* also suggests that L2 learners can utilize syntactic rules to creatively utilize the formulaic sequences they have mastered at the early stages of language learning:

> [t]he concept of a large vocabulary is extended from words to lexis, but the essential idea is that fluency is based on the acquisition of a large store of fixed and semi-fixed prefabricated items, which are available as the foundation for any linguistic novelty and creativity. Grammatical knowledge permits the creative re-combination of lexis in novel and imaginative ways, but it cannot begin to be useful in that role until the learner has a sufficiently large mental lexicon to which grammatical knowledge can be applied. (Lewis, 2008, p. 15)

In brief, *the Lexical approach* associates advanced knowledge of and proficiency in a language with L2 learners' mastery of a wide range of formulaic sequences and syntactic

rules (Lewis, 1997, 2002, 2008; Nattinger & DeCarrico, 1992; Willis, 2003).

Placing formulaic sequences at the core of classroom materials has also partly emerged in an attempt to overcome the limitation of incidental L2 vocabulary acquisition, which was influenced by *the default hypothesis* of L1 vocabulary acquisition; that is, L1 vocabulary acquisition is a byproduct of exposure rather than instruction (Coady, 1997; Laufer, 2005, 2006; Lewis, 2008; VanPatten & Williams, 2007; Xu, 2010). *The default hypothesis* constituted the seeds for what is known to be *the Monitor Theory*, also referred to as *the Natural Approach* (Laufer, 2005; Lewis, 2008; VanPatten & Williams, 2007). Proponents of *the Natural Approach* argue that L2 learners naturalistically acquire vocabulary without instruction and, hence, emphasize the importance of extensive reading and cognitive activities to enhance naturalistic learning (Conzett, 2000; Gass, 1999; Laufer, 2005; Zimmerman, 1997).

Opponents of *the Natural Approach* have, however, argued that this approach suffers from many limitations: (a) multiple exposures to vocabulary items entail providing L2 learners with '*floods of input*,' which may not always be feasible because of time constrains; (b) providing L2 learners with rich input does not necessarily lead to vocabulary acquisition because L2 learners tend to disregard new lexical items; (c) noticing and guessing the meaning of new lexical words do not always result in their successful acquisition and retention; and (d) newly learned lexical items through repeated exposure may promote receptive but not productive skills (Alali & Schmitt, 2012; Bishop, 2004a; Conklin & Schmitt, 2012; Laufer, 2005; Nation, 2002; Schmitt, 2000).

The limitations of *the Natural Approach* are even more pronounced in the case of formulaic sequences, which lack perceptual saliency because of their partial transparency

or total opacity (Biber, 2010; Boers & Lindstromberg, 2009; Cortes, 2004; Granger, 1998; Moon, 1992; Schmitt & Carter, 2004). Many L2 learners disregard formulaic sequences in discourse because of the mental burden imposed by the complexity of this language phenomenon and L2 learners' tendency to see language as constructed of individual words rather than chunks (Boers & Lindstromberg, 2009, 2012; Conklin & Schmitt, 2012; Coxhead & Byrd, 2007; Durrant & Schmitt, 2009; Granger & Meunier, 2008; Hasselgren, 1994; Howarth, 1996; Wray, 2002). Therefore, exposing L2 learners to rich input lacks in both feasibility because of the constraints of time in the language classroom and effectiveness "since grammatically generated strings of words can appear identical to formulaic sequences" (Bishop, 2004a, p. 15). Unless formulaic sequences, whose acquisition is fraught with complexity, are explicitly addressed in the language classroom or at least made salient, L2 learners may fail to notice and acquire them (Bishop, 2004a; Cortes, 2004; Laufer, 2005; Moon, 1992; Wood, 2010b).

The importance of placing formulaic sequences at the core of classroom materials has emerged as a result of the crucial role formulaic sequences play in L2 learners' fluent and proficient language production (Coxhead & Byrd, 2007; Hill, 2000; Lewis, 1997, 2000c; Meunier, 2012; Siepmann, 2008). As Ellis (1996) points out, "lexical phrases are as basic to SLA as they are to the L1 ..., and so instruction relies as much on teaching useful stock phrases as it does on teaching vocabulary and grammar" (p. 97). Such an approach to language teaching is of crucial importance in academic contexts because, as several scholars suggest, academic lexis—including formulaic sequences—is "a *lexical bar* or barrier that students need to transcend in order to move successfully from

everyday ways of expressing meaning to the specialized, 'high-status' academic language" (Coxhead & Byrd, 2007, p. 134).

In short, the limitations of the structure-based and natural approaches, the complexity of formulaic sequences, and the crucial role of formulaic sequences in language production have resulted in a major shift in second language acquisition research and pedagogy, granting formulaic sequences a solid position in L2 pedagogy in general and L2 composition instruction in particular.

Formulaic Sequences and L2 Academic Writing Pedagogy

Placing formulaic sequences at the core of any approach to teaching L2 academic writing skills was proposed as early as the 1990s (e.g. Cowie, 1992; Ellis, 1996; Howarth, 1996; Lewis, 1996, 2008; Verstraten, 1992). Among the prominent scholars who advocated such importance in the early 1990s is Cowie (1992): "the sheer density of ready-made units in various types of written text is a fact that any approach to teaching writing to foreign students has to come to terms with" (p. 10). The reason behind such a call stems from the pivotal role of formulaic sequences in identifying L2 learners as proficient writers and improving L2 learners' linguistic accuracy (Ädel & Erman, 2012; Boers & Lindstromberg, 2012; Cortes, 2004; Ellis, 2005a; Hyland, 2008a; Kennedy, 2008; Paqout & Granger, 2012; Peters & Pauwels, 2015).

Since formulaic sequences are so pervasive a part in academic writing and function as important building blocks of academic prose, L2 learners need to build a rich repertoire of formulaic sequences in order to come across as proficient writers (Byrd & Coxhead, 2010; Coxhead & Byrd, 2007; Erman & Warren, 2000; Flowerdew, 2015; Howarth, 1996; Hyland, 2008a, 2008b). Such control is likely to help L2 learners (a)

utilize a wide range of formulaic sequences in their writing production to shape text meanings, (b) express concepts in ways typical of academic contexts, (c) abide by the expectations of an intended academic reader and (d) come across as full-fledged members of a discourse community (Ädel & Erman, 2012; Boers & Lindstromberg, 2012; Coxhead & Byrd, 2007; Ellis, 2005a; Haswell, 1991; Hyland, 2008b; Jones & Haywood, 2004; Kazemi et al., 2014). It can also help L2 learners write more fluently and rise above the characteristic behavior of apprentice writers, that is, producing texts that are devoid of formulaic sequences or clinging onto a very limited number of formulaic sequences in their writing production (Boers & Lindstromberg, 2012; Durrant & Mathews-Aydinli, 2011; Ellis & Simpson-Vlach, 2009; Gledhill, 2000; Howarth, 1996; Hyland, 2008b; Jones & Haywood, 2004; Laufer & Waldman, 2011).

Incorporating formulaic sequences into classroom materials is also believed to elevate L2 learners' linguistic accuracy, which is part and parcel of being identified as proficient writers (Ellis et al., 2008; Ellis, 2005a; Fan, 2009; Nesselhauf, 2003; Shin & Kim, 2017; Willis, 2003). One of the major sources of linguistic inaccuracy in writing production is L2 learners' underdeveloped knowledge of phraseology, also referred to as lack of *collocational competence*, because using a lexical item productively is a complex task that requires knowledge of its form, its meaning, and, more importantly, its collocates (Bahns, 1993; Coxhead, 2008; Coxhead & Byrd, 2007; Ellis, 2005a; Hill, 2000; Laufer & Waldman, 2011; Lewis, 2002; Verstraten, 1992; Ward, 2007). Without knowledge of L2 phraseology, L2 learners may construct awkward word combinations that are atypical of English (Bahns, 1993; Cortes, 2006; Durrant & Schmitt, 2009; Fan, 2009; Hill, 2000; Lewis, 2008). Such lexical errors are usually considered serious ones

because “a lexical mistake can hinder effective communication, unlike a grammar mistake which rarely does” (Lewis, 2008, p. 95).

The lack of *collocational competence* can also result in producing academic texts that are neither syntactically accurate nor semantically precise (Durrant & Schmitt, 2009; Hill, 2000; Jones & Haywood, 2004; Laufer & Waldman, 2011; O’Donnell et al., 2013; Willis, 2003). According to Hill (2000), lack of *collocational competence* “forces students into grammatical mistakes because they create longer utterances” that do not “express precisely what they want to say” (p. 49). It may also adversely affect academic achievement: “students with good ideas often lose marks because they do not know the four of five most important collocates of a key word that is central to what they are writing about” (Hill, 2000, p. 50).

As a solution to this problem, researchers have advocated focusing on formulaic sequences because it will help L2 learners become more aware of L2 word combinations and produce error-free stretches of discourse (Coxhead, 2008; Ellis, 2005a; Granger, 1998; Hill, 2000; Willis, 2003). It can also mitigate most of the errors L2 writers make in the use of prepositions and articles, especially when presenting formulaic sequences with their adjacent prepositions and articles (Nesselhauf, 2003; Shin & Kim, 2017). Focusing on formulaic sequences in writing pedagogy may as well help overcome the pedagogical limitations of teaching lists of individual words—for example, the Academic Word List (AWL)—an approach whose efficacy is believed to be limited to receptive skills (Coxhead, 2008; Durrant, 2009; Hinkel, 2004; Howarth, 1996; Nattinger & DeCarrico, 1992). In brief, addressing formulaic sequences in academic writing courses is believed to be an attainable solution to help L2 learners develop a rich repertoire of formulaic

sequences, promote their linguistic accuracy, and be ultimately identified as proficient writers (Boers & Lindstromberg, 2009; Ellis, 2012; Ellis et al., 2008; Fan, 2009; Millar, 2011; Oakey, 2002; Paquot & Granger, 2012; Wood, 2010a).

Formulaic Sequences in the Language Classroom

In order to integrate formulaic sequences into L2 pedagogy, experts in the field of SLA have proposed different approaches to addressing formulaic sequences in the language classroom as well as some classroom activities to effectively incorporate formulaic sequences into pedagogical materials.

Approaches to addressing formulaic sequences. Different scholars have advocated the pedagogical efficacy of one of the three major approaches pinpointed in the literature, that is, frequency, saliency, and focused instruction.

Frequency. A considerable number of researchers consider frequency of exposure the basic factor influencing the acquisition of formulaic sequences (e.g. Boers & Lindstromberg, 2012; Conklin & Schmitt, 2012; Ellis et al., 2008). Such a consideration has its origin in the literature on vocabulary acquisition: "The broad consensus is that each time you meet a word in context and (at least partly) understand it, you understand more of its meaning, and gradually integrate it into your lexicon for immediate access" (Lewis, 2008, p. 51) and is extended to formulaic sequences: "[i]t is only in spontaneous communication that the immediate and flexible selection of formulaic sequences becomes apparent. Repeated exposure to such input over time would encourage learners to achieve a certain level of comfort with natural expression in English" (Wood, 2002, p. 10). With repeated exposure, L2 learners may notice the frequent occurrence of formulaic sequences in discourse, understand their various discourse functions, entrench them in

memory, and ultimately utilize them in their academic writing (Conklin & Schmitt, 2012; Conzett, 2000; Durrant & Schmitt, 2010; Hyland, 2008a; Lewis, 2008; Li & Schmitt, 2009; Schmitt, 2008).

Saliency. Other scholars suggest that repeated exposure alone may have very limited pedagogical utility and validity, especially when dealing with such a complex phenomenon as formulaic sequences (Alali & Schmitt, 2012; Byrd & Coxhead, 2010; Cortes, 2004, 2006; Granger, 1998). This limited effect can be ascribed to the fact that incidental learning of complex language units usually results in the improvement of receptive rather than both receptive and productive knowledge (Byrd & Coxhead, 2010; Cortes, 2004; Granger, 1998). Thus, L2 instructors should make formulaic sequences salient in input and provide L2 learners with practice activities in order to promote the chances of noticing and acquiring formulaic sequences and develop L2 learners' receptive and productive skills (Bishop, 2004a, 2004b; Conzett, 2000; Laufer & Waldman, 2011; Lewis, 2008; Schmitt, 2000; Wood, 2010a). On the pedagogical efficacy of saliency, Cortes (2004) suggests that providing students with salient input may help them *notice* lexical bundles and make them "more aware of the different contexts and discourse functions they perform in academic disciplines" (p. 420).

It is worth noting here that advocates of both frequency and saliency stress the importance of raising L2 learners' awareness of the frequent occurrence and important functions of formulaic sequences in input (Boers & Lindstromberg, 2012; Lewis, 1997, 2008; Willis, 2003). In so doing, teachers may foster independent learning skills in that L2 learners may start to identify formulaic sequences in the academic input they may encounter outside the language classroom (Boers & Lindstromberg, 2012; Granger &

Meunier, 2008; Laufer & Waldman, 2011). As Willis (2003) puts it, teachers need to "establish general recognition by making learners aware of the importance of lexical phrases and encouraging them to look for these phrases in future input" (p. 45).

Focused instruction. Other scholars, however, argue that neither frequency nor saliency can result in the successful acquisition of formulaic sequences, whose functions and use may not be fully intelligible to many L2 learners without explicit instruction (Boers & Lindstromberg, 2009; Coxhead & Byrd, 2007; Jones & Haywood, 2004; Staples et al., 2013). It is, thus, the responsibility of language instructors to extract the most useful formulaic sequences from the presented texts and teach them explicitly in the language classroom (Boers & Lindstromberg, 2009; Byrd & Coxhead, 2010; Laufer & Waldman, 2011; Staples et al., 2013; Woolard, 2000). Focused instruction, as proponents of this approach suggest, is likely to provide L2 learners with the instruction and practice needed for the accurate production of formulaic sequences (Byrd & Coxhead, 2010; Peters & Pauwels, 2015; Wood, 2009; Woolard, 2000).

Classroom Activities. Several scholars have suggested different classroom activities that can enhance the acquisition and internalization of formulaic sequences into L2 learners' linguistic repertoire and ultimately improve their academic writing skills.

Practicing sentence builders and frames. One of the major activities suggested in the literature is practicing the creative use of sentence builders and frames. These activities, as Lewis (1997), Nattinger and DeCarrico (1992), and Willis (2003) explicate, may help L2 learners practice and acquire stretches of discourse that execute important discursive functions and ultimately utilize them to express ideas in highly informational academic texts. Teachers, for example, can focus on topic nominating frames (e.g. *the*

(basic) emphasis/ proposal/ goal of this paper/ article is to) and topic organization frames (e.g. *this paper will show/ compare/contrast/describe/ demonstrate that*) in the language classroom to familiarize the students with academic frames that can be utilized in writing to express ideas and organize texts (Nattinger & DeCarrico, 1992).

Teachers can also have L2 learners practice frames that help them "make individual choices, 'with no serious risk of error,' and thereby give scope for creativity and innovation" (Willis, 2003, p. 45). Such an activity can be implemented in the language classroom by having L2 learners practice frames with slots to construct meaningful stretches of discourse (Lewis, 1997; Willis, 2003; Woolard, 2000). L2 learners, for example, can be required to complete *from a/an _____ point of view* with such syntactically related words as *financial*, *economical, political,* etc. to form grammatical and meaning phrases (Willis, 2003). L2 learners can also be instructed to use frames in order to compose meaningful sentences (Lewis, 1997; Woolard, 2000). For example, after introducing *she holds very strong views on marriage. She thinks everybody should be married in a church,* L2 learners complete *most people hold strong views on____* (Woolard, 2000). Teachers can also have L2 learners practice the creative use of such phrases as *very different from* by replacing the adverb *very* with other syntactically related words, such as *so*, *little*, and *fundamentally* (Nation, 2001). Alternatively, teachers can have their students match different sentence heads with syntactically and semantically appropriate endings to form meaningful sentences (Lewis, 1997, 2008).

Dictation. Different scholars have suggested different forms of dictation—defined as "a technique where the learners receive some spoken input, hold this in their memory

for a short time, and then write what they hear" (Nation, 1991, p. 12)—to promote the successful acquisition of formulaic sequences.

To start with, *dividing up texts* is a technique proposed by Nation (1991, 2001) to promote the acquisition and internalization of formulaic sequences into L2 learners' linguistic repertoire. It can be implemented in the language classroom using four different activities, two of which involve dictation, namely, delayed repetition and dictation (Nation, 1991, 2001). In delayed repetition and dictation activities, teachers dictate texts rich in formulaic sequences to L2 learners who, in turn, are required to reproduce the dictated texts orally, as in the case of delayed repetition, or in writing, as in the case of dictation (Nation, 1991, 2001). The listen and reproduce activities involved in delayed repetition and dictation promote the retention of the dictated language in L2 learners' short-term memory before producing it chunk by chunk and, thus, may create a chance for L2 learners to acquire formulaic sequences (Nation, 1991, 2001).

Dictogloss—a multi-stage, output-oriented task that requires L2 learners to compose a semantically approximate and syntactically accurate version of a dictated text (Jacobs & Small, 2003; Nabei, 1996; Nation, 1991; Qin, 2008; Wajnryb, 1990; Wood, 2001)—is another activity that may promote the acquisition of formulaic sequences (Meunier, 2012; Wood, 2001, 2002). Unlike classic dictation, this collaborative activity, which is also referred to in the literature as dicto-comp, involves both dictation and composition (Nation, 1991; Wajnryb, 1990). In dictogloss, L2 learners are first introduced to the target topic through discussions and/or the introduction of new words; they are then dictated a text replete with formulaic sequences at normal speed twice with very short pauses between sentences; at the dictation stage, L2 learners can jot down

some content words and/or phrases; they, then, compose a written text that is syntactically accurate and semantically similar to the dictated one in pairs or in small groups (Jacobs & Small, 2003; Nabei, 1996; Nation, 1991; Wajnryb, 1990; Wood, 2001, 2002). L2 learners' texts are either analyzed and corrected as a whole class or checked against the original text by the students themselves at the final stage of this activity (Jacobs & Small, 2003; Nabei, 1996; Wood, 2002). As a classroom activity, dictogloss may help L2 learners "to attend to formulaic sequences in text…. It can also help them to retain the sequences by having them focus on their constituent parts and see how they fit into the flow of discourse" (Wood, 2002, p. 11).

Another collaborative dictation activity that may facilitate acquiring formulaic sequences is chain dictation (Wood, 2010b, 2015). In this activity, L2 learners collaborate in small groups on constructing a text that is dictated to individual group members (Wood, 2010b, 2015). To do so, teachers should first divide the students into small groups and assign each group member a different number (e.g. 1, 2, 3, etc.); teachers will then stand in a quiet place, inside or outside the classroom, and dictate chunks of the target text to group members starting with number one, then two, and so on until the entire text is orally dictated (Wood, 2010b, 2015). During this process, each group member goes to the teacher, listens to the dictated sentence, goes back to his/her group, and retells the sentence to his/her group members (Wood, 2010b, 2015).

Alternatively, language teachers may have their students form a semi-circle and orally retell one another the dictated sentences while sitting; that is, the teacher dictates the student sitting at the far end of the semi-circle a sentence, who will in turn retell the sentence to the student sitting next to him/her until the sentence reaches the last student,

who is responsible for writing the dictated text on the board (Model Course 3.17: Maritime English, 2009). In either case, the same procedure is repeated until the whole text is dictated (Model Course 3.17: Maritime English, 2009; Wood, 2015). Chain dictation may promote the acquisition and automatization of formulaic sequences because the length of the dictated texts is likely to force L2 learners to focus on chunks of language rather than individual words to reduce processing load (Wood, 2010b, 2015).

Student dictation is another collaborative activity that may promote the successful internalization of formulaic sequences into L2 learners' linguistic repertoire (Wood, 2010b, 2015). This activity involves L2 learners in peer dictation; that is, language teachers give each student half a dictation text and have L2 learners work in pairs to construct the full target text by dictating their partners the missing half (Wood, 2010b, 2015). This gap-filling activity urges L2 learners to "notice and retain formulaic sequences in working memory to complete the task" (Wood, 2015, p. 151). L2 learners also need to "negotiate word-by-word in order to complete a formulaic sequence, which forces a focus on its structure and component parts" (Wood, 2010b, p. 203).

Creating lists. Other activities that can be utilized to incorporate formulaic sequences into classroom materials include using lists of formulaic sequences. Providing L2 learners with lists of formulaic sequences rather than single words can promote L2 learners' ability "to process language as patterns and phrases rather than individual words" (Willis, 2003, p. 151). Teachers, for example, can develop lists in which different formulaic sequences are categorized based on their meaning or structural functions in discourse (Howarth, 1996; Willis, 2003). Teachers can also create thematic lists of formulaic sequences along with visual representations to enhance the deep processing of

formulaic sequences, and, thus, facilitate their acquisition (Boers & Lindstromberg, 2009; Celce-Murcia & Olshtain, 2000; Hunt & Beglar, 2002). As well, teachers can compile and utilize lists of formulaic sequences typical of a specific topic as a prewriting activity in order to familiarize L2 learners with the frequently used ones to convey meaning in a specific context (Conzett, 2000). Teachers can also create lists of formulaic sequences along with prepositions (e.g. *pass judgment on* instead of *pass judgment*) to minimize the number of errors resulting from the inaccurate use of prepositions (Nesselhauf, 2003).

In addition to providing lists of formulaic sequences, teachers can actively use created lists in the language classroom. Teachers, for example, can involve their students in a matching activity, in which they match two lists of words, sentence parts, or dialogue components in order to respectively form formulaic sequences, meaningful sentences, and mini dialogues that are rich in formulaic sequences (Lewis, 1997, 2008; Nation & Newton, 1997). They can also provide their students with a list of formulaic sequences and ask them to come up with a new list that is opposite in meaning or a list of formulaic sequences together with a model dialogue and ask L2 learners to create a dialogue similar to the model one based on the provided list (Lewis, 1997, 2008). Other activities that involve the active use of lists include (a) requiring L2 learners to predict the words that collocate with a target stem, (b) having L2 learners use a list of formulaic sequences to complete sentences, texts, and/or dialogues, (c) instructing L2 learners to categorize lists of formulaic sequences based on their grammatical structures or functions in discourse, or (d) providing L2 learners with a list of stem words along with several possible collocates and asking them to omit the collocates that do not pair with the stem words (Lewis, 1997, 2002, 2008; Willis, 2003).

Lists of formulaic sequences can also be used to reveal the diverse usage and meanings of a single formulaic sequence. Language teachers, for example, can create a list of different sentences that include one formulaic sequence, change one part of the target formulaic sequence in each sentence, and have L2 learners identify the unique meaning each formulaic sequence conveys (Nation & Newton, 1997; Willis, 2003). Teachers can also create a list of different sentences that have one formulaic sequence with a variable slot (Lewis, 1997). In such an activity, L2 learners complete the target formulaic sequence (e.g. *out of*____ in such sentences as *I am unfit. If I climb the stairs I am out of*____) with semantically related words (Lewis, 1997). These lists can promote L2 learners' understanding of the meaning and usage of different formulaic sequences when communicating in the target language (Lewis, 1997; Willis, 2003).

Having L2 learners identify formulaic sequences in input. Language teachers can have L2 learners write down formulaic sequences they encounter in various input, typically in the media or their communities, and arrange them alphabetically, thematically, or functionally (Lewis, 1997; Wood, 2010b; Woolard, 2000). This technique might help L2 learners revisit the target formulaic sequences and may provide them with the multiple exposures needed for their acquisition (Woolard, 2000). Other proposed techniques are based on the use of reading texts. For instance, teachers may assign L2 learners a reading text and have them underline or write down the formulaic sequences used in it (Hill, Lewis, & Lewis, 2000; Lewis, 2008; Willis, 2003). Teachers can also have L2 learners outline an academic text, a task that is likely to make the target formulaic sequences more salient for L2 learners and raise their awareness of the forms and functions of this language phenomenon (Nattinger & DeCarrico, 1992). Having L2

learners identify chunks in input is particularly helpful because such an activity "encourages accurate recording in their lexical notebooks, and more importantly, storage in chunks in the mental lexicon" (Lewis, 2008, p. 89).

Communicative activities. Focusing on communicative activities can enhance L2 learners' acquisition of formulaic sequences and promote their tendency to use them when communicating in the target language: "It is only through repeated encounters and use in communicative activities that receptive knowledge of collocations will develop into productive knowledge and learners may gradually build up confidence in L2 collocational use" (Fan, 2009, p. 121). The pedagogical utility of communicative practice is paramount when in written practice because "[d]oing such practice in groups takes advantage of different cognitive styles, different students' individual knowledge, and the doing of practice can itself be a communicative activity" (Lewis, 2002, p. 161).

Among those communicative activities is the mingle jigsaw, a highly cooperative activity in which different parts of a text rich in formulaic sequences are assigned to different L2 learners who, in turn, are required to learn their assigned part by heart, orally share it with their peers, listen to the information delivered by their peers, and go back to their seats to write down the missing parts (Wood, 2001, 2002, 2009, 2010b). The mingle jigsaw activity is likely to promote the acquisition of formulaic sequences because it "incorporates the repetition necessary for automatization and encourages students to chunk words together in order to express and retain the pieces of text" (Wood, 2002, p. 12). Jigsaw activities can also be used to practice dialogues (Lewis, 2008). In jigsaw dialogues, students are required to first rearrange chunks of a dialogue dense with formulaic sequences and then repeatedly practice the target dialogue until they can

produce it fluently (Lewis, 2008).

Read-and-look-up is another communicative activity suggested in Nation's (1991, 2001) *dividing up texts*. In the read-and-look-up activity, L2 learners work in pairs and are assigned different roles (i.e., a reader and a listener); the reader gets a written text, typically a dialogue, and rehearses it before delivering it orally to the listener or independently, for example, at home in front of a mirror (Nation, 1991, 2001). Similar to the mingle jigsaw, L2 learners should not refer to the original text when sharing information with their peers (Nation, 1991, 2001). Therefore, the read-and-look-up activity is likely to urge L2 learners to chunk the target text and, thus, commit formulaic sequences to their long-term memory (Nation, 1991, 2001).

Memorization. Memorization is viewed as one of the effective techniques that can promote the likelihood of committing formulaic sequences to L2 learners' memory. At the early stages of language learning, teachers can utilize drills to help L2 learners memorize flexible formulaic sequences that can be used creatively when L2 learners achieve higher levels of proficiency (Lewis, 2002; Nattinger & DeCarrico, 1992). Teachers, for example, can introduce formulaic sequences with open slots, such as *I'm (very) sorry to (hear (about) X*, and have L2 learners complete, practice, and memorize them (Nattinger & DeCarrico, 1992). Memorizing such formulaic sequences can be highly effective in promoting L2 learners' linguistic accuracy as they can retrieve language as chunks from the long-term memory and, hence, bypass the need to construct language based on syntactic rules at the time of spoken or written language production (Lewis, 2002; Nattinger & DeCarrico, 1992; Nesselhauf, 2003).

In addition, language teachers can utilize memorization as a pedagogical

technique to promote L2 learners' writing proficiency; that is, language teachers can have their students memorize formulaic sequences that can be utilized to organize discourse segments and produce coherent and cohesive academic written texts (Conzett, 2000). Flash cards can also be used in the language classroom to help L2 learners' memorize formulaic sequences (Nation, 2001, 2002). Teachers, for example, can write formulaic sequences on flash cards, instruct L2 learners to repeat them with an appropriate time interval to promote deep processing, and, then, have L2 learners utilize the practiced formulaic sequences to construct meaningful sentences (Nation, 2001, 2002).

Using L1 in teaching L2 collocations. This technique is obviously limited to monolingual contexts. However, it has its own merits as it gives L2 learners the chance to compare the use of formulaic sequences, most often collocations, in their mother tongue with that in the target language, which might raise their awareness of the importance of this language phenomenon in language production and promote their linguistic accuracy (Fan, 2009; Lewis, 2008; Willis, 2003). For example, language teachers can ask L2 learners to think of some collocation examples and reflect on their importance in their mother tongue; in so doing, L2 learners can realize that collocations are as important in the target language as they are in their mother tongue (Fan, 2009; Nesselhauf, 2003; Willis, 2003). When introducing the functions of some collocations, teachers may also have L2 learners identify collocations in their L1 which have similar meanings and functions to those in the L2 (Willis, 2003). Alternatively, they can provide L2 learners with a list of inaccurate formulaic sequences and have L2 learners first correct them and then come up with equivalent ones in their L1 (Lewis, 2008).

Using L1 in teaching L2 collocations can also mitigate the negative effects of L1

transfer on L2 production (Fan, 2009). This can be done by drawing L2 learners' attention to how L2 collocational use is different from that in L1 whenever necessary (Fan, 2009). Language teachers may also create 'deviant' L1 collocations or literally translate some L2 collocations into L1 and have L2 learners comment on their acceptability (Fan, 2009). Having L2 learners regularly translate a number of collocations into their own language as single units is another technique that helps L2 learners view language as composed of multiword sequences rather than single vocabulary words (Hill et al., 2000). These techniques promote L2 learners' lexical accuracy and help them become more sensitive to collocations in L2 input and more ready to use them in their language production (Fan, 2009; Lewis, 2008; Nesselhauf, 2003; Willis, 2003).

Other activities. *Running words together* is a technique that Willis (2003) recommends for speech fluency. When designing pronunciation activities, teachers should include such formulaic sequences as *there is a___*, *have you got any___*, etc. and have L2 learners pronounce them as fast as possible to promote their acquisition as chunks rather than individual words (Willis, 2003). *Running words together* can also be designed as a listening activity in which L2 learners listen to the language as it is produced by its native speakers at a normal pace (Willis, 2003). In so doing, "the careful enunciation of the language, which learners hear in the early stages, reinforces the notion that English is made up of a series of isolated words" (Willis, 2003, p. 160). L2 learners may also be involved in such practice activities as role plays, which may provide them with the rehearsal required to acquire the formulaic sequences suitable for certain contexts and to promote speech fluency (Lewis, 1997; Wood, 2010b). Alternatively, L2 learners can be provided with a list of formulaic sequences along with intensifying and

softening adverbs and may be required to orally practice their use with a special focus on stress patterns in order to master expressing different degrees of emotions (Lewis, 2008).

In addition to fluency activities, designing writing activities that require L2 learners to utilize formulaic sequences repeatedly is another technique that can provide L2 learners with the multiple exposures required for their acquisition and ultimately upgrade L2 learners' writing proficiency (Conzett, 2000; Nation, 2001). L2 learners can also be assigned a text-copying task, which is referred to as delayed copying, the last form of Nation's (1991, 2001) *dividing up texts*. In this activity, L2 learners are assigned a reading text to copy individually (Nation, 1991, 2001). The effectiveness of this individual activity lies in an essential feature of the copying process; that is, L2 "learners try to hold as large a phrase as possible in their memory before writing it. So instead of copying word for word, the learners read a phrase, look away from the text, and then write it" (Nation, 1991, p. 13). Teachers can also engage their students in different types of sequencing activities (Lewis, 1997, 2008). For example, they can have L2 learners rearrange words to form meaningful sentences or dialogue segments to construct a meaningful dialogue (Lewis, 1997, 2008). These activities may raise L2 learners' awareness of word patterns and institutionalized language (Lewis, 1997, 2008).

Using concordance lines in the classroom can be an extremely useful pedagogical technique; it provides L2 learners with multiple exposures to formulaic sequences in various contexts and promotes "a deep and thoughtful level of mental processing as students become involved in investigating for themselves the typical patterns of use of the target items" (Jones & Haywood, 2004, p. 272). Students, for example, can be required to check and compare the use of a single formulaic sequence in multiple

concordance lines (Lewis, 2008; O'Keefe et al., 2007). They can also be asked to identify the commonalities and differences between confusable words by checking their usage in concordance lines (Lewis, 2008). Concordance lines can provide L2 learners with rich input, promote their creative use of formulaic sequences, and help them better understand the semantic and syntactic features of this language phenomenon (Jones & Haywood, 2004; Lewis, 2008; Nation, 2001; O'Keefe et al., 2007).

In brief, the above-reviewed pedagogical approaches and classroom activities can raise L2 learners' awareness of formulaic sequences as an essential part of language production, promote the deep processing required for the acquisition of formulaic sequences, foster L2 learners' ability to use this language phenomenon creatively in their language production, and ultimately upgrade L2 learners' fluency and proficiency (Conzett, 2000; Jones & Haywood, 2004; Lewis, 1997; Nation, 2001; Nattinger & DeCarrico, 1992; Nesselhauf, 2003; Willis, 2003; Wood, 2002, 2010b, 2015). It is worth noting here, as a final remark, that the pedagogical utility of the proposed presentation methods and classroom activities is highly contingent on language teachers' ability to correctly identify the formulaic sequences that best account for their students' needs (Boers & Lindstromberg, 2009; Ellis et al., 2008; Howarth, 1996; Lewis, 1997, 2002, 2008). That is, language teachers need to select the formulaic sequences that are typical of the target register (i.e., spoken or written) and context (e.g. academic) in order to boost the efficiency of such an approach to language teaching (Fan, 2009; Hyland, 2008a; Lewis, 2000a, 2008; Nesselhauf, 2003; Oakey, 2002).

Summary of Chapter Two

This chapter has highlighted the increasing use of English academic writing as a means of communicating and assessing knowledge worldwide, which has resulted in placing English academic writing at the core of SLA research and pedagogy. It has then explored the strengths and weaknesses of four major approaches that have been proposed by experts in the field of SLA in order to teach L2 academic writing skills. The weaknesses of each approach have mainly been ascribed to overemphasizing certain aspects at the expense of others and to paying no head to a frequently recurring language phenomenon in academic writing, that is, formulaic sequences.

The chapter has presented the most recurring definitions of three different types of multiword sequences—*formulaic sequences*, *lexical bundles*, and *collocations*—in an attempt to provide a comprehensive definition of this language phenomenon. It has also listed several defining characteristics that distinguish formulaic sequences from grammatically generated sequences of words. The chapter has then investigated the two major approaches adopted in corpus linguistics research studies on formulaic sequences and written academic discourse. These studies have empirically demonstrated that the formulaic nature of academic writing and have categorized formulaic sequences based on their syntactic structures and their semantic and pragmatic functions in academic discourse.

The chapter has also discussed the important implications of the findings of corpus linguistics research for teaching L2 academic writing pedagogy. These findings have contributed to considering formulaic sequences a prerequisite for improving L2 learners' linguistic proficiency and identifying them as proficient writers and, hence,

have paved the way for placing different types of formulaic sequences at the center of L2 pedagogy. The chapter has highlighted the three different presentation methods of formulaic sequences in the language classroom, namely, frequency, saliency, and explicit instruction. It has also tackled different activities proposed by experts in the field of SLA to provide language teachers with some insights into integrating formulaic sequences into classroom materials. In short, the review of literature in this chapter has laid the theoretical foundations for conducting this quasi-experimental research study, which is discussed in the following chapter.

Chapter Three: Motivation for the Study

The third chapter briefly reiterates the major theoretical accounts of formulaic sequences and academic writing skills and highlights the underlying rationale for conducting this quasi-experimental research study. It then concludes with the research questions that are theoretically rooted in the gap between the growing literature on the effectiveness of formulaic sequences in augmenting L2 academic writing skills and the limited empirical evidence in support of such a role.

Since the early 1990s, there has been rapidly growing research on formulaic sequences and academic writing (Biber, 2009, 2010; Flowerdew, 2010, 2015; Gries, 2008; Pecorari, 2009; Read & Nation, 2004). Several corpus linguistics researchers have attempted to empirically investigate the recurrence of this language phenomenon in different types of academic texts: published research (e.g. Cortes, 2008; Hyland, 2008a, 2008b), M.A. theses and doctoral dissertations (e.g. Hyland, 2008a, 2008b), freshman university students (e.g. Cortes, 2002), learner corpora (e.g. Siyanova & Schmitt, 2008), and mega corpora (e.g. Liu, 2012; Oakey, 2002). Other researchers have taken their analysis one step further by comparing written texts that have been produced by expert writers and university students (e.g. Cortes, 2004), expert writers and L1 as well as L2 university students (e.g. Chen & Baker, 2010), or L1 and L2 university students (e.g. Ädel & Erman, 2012; Durrant & Schmitt, 2009; O'Donnell et al., 2013).

The results of corpus linguistics research on formulaic sequences and academic writing skills—which have empirically demonstrated the recurrence of formulaic sequences in academic writing and novice writers' failure to effectively utilize them in their writing production—have been viewed as a compelling reason to argue in favor of

integrating this language phenomenon into teaching L2 academic writing skills, especially in EAP contexts (Byrd & Coxhead, 2010; Cowie, 1992; Ellis et al., 2008; Jones & Haywood, 2004; Martinez & Schmitt, 2012; Oakey, 2002; Wood, 2010a). Such an approach to teaching academic writing is believed to promote L2 learners' optimal utilization of this language phenomenon in their writing production, which is, in turn, likely to improve their writing proficiency (Boers & Lindstromberg, 2012; Durrant & Schmitt, 2009; Hill, 2000; Jones & Haywood, 2004; Laufer & Waldman, 2011; O'Donnell et al., 2013).

Placing formulaic sequences at the core of classroom materials has also inspired researchers to suggest different approaches to addressing them in the language classroom. While some researchers argue that providing L2 learners with multiple exposures to formulaic sequences helps increase the chances of their acquisition and internalization (e.g. Alali & Schmitt, 2012; Conklin & Schmitt, 2012; Conzett, 2000; Li & Schmitt, 2009), others have asserted that formulaic sequences need to be made salient in input in order for L2 learners to discern and, subsequently, acquire them (e.g. Bishop, 2004a, 2004b; Byrd & Coxhead, 2010; Cortes, 2004; Laufer & Waldman, 2011). Other scholars, however, have questioned the utility of both repeated exposure and saliency of input in prompting the acquisition of formulaic sequences and have, therefore, maintained that such a complex language phenomenon needs to be explicitly taught in order to ensure its successful acquisition and internalization into L2 learners' linguistic repertoire (e.g. Boers & Lindstromberg, 2009; Coxhead & Byrd, 2007; Wood, 2009; Woolard, 2000).

While the theoretical account of the importance of formulaic sequences in augmenting L2 learners' academic writing skills and the pedagogical utility of different

approaches to presenting them in the language classroom have been rapidly growing in depth and breadth since the 1990s, very few studies have empirically investigated the possible positive effects formulaic sequences may have on EAP students' writing skills. In addition, none of the studies that were conducted in EAP contexts has provided a full picture of the possible effects formulaic sequences may have on EAP students' writing skills or ample evidence in support of the pedagogical utility of the proposed presentation methods because these studies have yielded conflicting results.

AlHassan (2016), for example, explored, in a pilot study, the effects of a focused instructional approach to formulaic sequences on 12 EAP students' comprehension and production of graphical information in uncontrolled reading and writing tasks. A similar study was conducted by AlHassan and Wood (2015), in which they attempted to investigate the effects of focused instruction of formulaic sequences on 12 EAP students' ability to report graphical information in paragraphs. The effects of focused instruction of formulaic sequences on L2 learners' writing skills were also probed in Čolović-Marković's (2012) research study, in which 63 L2 university students registered in an academic writing course were engaged in both controlled and uncontrolled tasks (i.e., C-test and essays). Jones and Haywood (2004), in turn, examined the effects of repeated exposure to and noticing of formulaic sequences on 21 EAP students' writing skills in both controlled and uncontrolled tasks (i.e., underlining, C-test, and essay writing).

While the participants in AlHassan (2016) and AlHassan and Wood (2015) utilized more formulaic sequences in their writing production and exhibited increased writing proficiency after being explicitly taught a number of formulaic sequences, there was no significant difference in the uncontrolled writing production of Čolović-Marković's

(2012) participants after explicit instruction despite the significant improvement in the controlled task. Results similar to those of the latter were reported in Jones and Haywood's (2004) study despite the difference in their presentation method (i.e., saliency).

Other research studies that were conducted in contexts other than EAP—among which are public schools (e.g. Alali & Schmitt, 2012), intensive writing courses for native speakers of English (e.g. Cortes, 2006), and graduate courses (e.g. Kazemi et al., 2014; Li &Schmitt, 2009)—have also produced inconsistent results. For example, Li and Schmitt (2009) suggested that exposure to formulaic sequences in natural contexts mainly helped their university student participant acquire and accurately use 166 new lexical bundles in her academic writing. Repeated exposure to salient formulaic sequences along with consciousness raising, however, did not result in an increase in the use of formulaic sequences in the writing samples produced by 8 L1 university students who were registered in an intensive writing course for native speakers of English (Cortes, 2006). The presentation method used in Cortes's (2006) study was adopted by Kazemi et al. (2014) in their investigation of the effects of formulaic sequences on the writing production of 20 L2 graduate students, but it augmented their participants' writing proficiency.

The very limited number of research studies on formulaic sequences and EAP students' writing skills as well as the apparent inconsistency in the results of the above-reviewed research studies—which may be ascribed to the implemented presentation method, the designed instruments, the type of writing task, the participants' L1 background and/or L2 proficiency level, or any other factors—mean that there is, as yet,

no solid empirical evidence for the effectiveness of formulaic sequences in augmenting EAP students' writing proficiency or for the pedagogical utility of any of the suggested presentation methods. It is also interesting to note that while several scholars (e.g. Gledhill, 2000; Jones & Haywood, 2004; Laufer & Waldman, 2011; Schoonen, van Gelderen, Stoel, Hulstijn, & de Glooper, 2011) theoretically postulate that formulaic sequences help L2 writers communicate their ideas concisely and promote their writing pace, especially in timed writing tasks, whether formulaic sequences affect the length of written texts or not is, as Wood (2015) suggests, a question that requires further investigation.

To help address these gaps in the literature, the present research study attempts to empirically investigate the recurrent notions proposed in the literature using two presentation methods (i.e., saliency and explicit instruction) and two different writing tasks (i.e., graphical summaries and essay responses). More specifically, this research study is an attempt to investigate whether integrating formulaic sequences into teaching academic writing, by making them salient or teaching them explicitly, can help EAP students (a) acquire and utilize them in their writing production, (b) write more efficiently and/or fluently, (c) mitigate the number of linguistic errors they make in writing, and (d) receive better evaluation (Ädel & Erman, 2012; Cortes, 2004; Durrant & Schmitt, 2009; Martinez & Schmitt, 2012). To this end, the following research questions guided this study:

1. Will both explicit instruction and saliency of formulaic sequences help EAP students (a) acquire and properly utilize a wider range of formulaic sequences when

approaching different writing tasks and (b) outperform EAP students who practice other writing materials?

2. Will both saliency and explicit instruction of formulaic sequences (a) affect the number of words EAP students generate in different timed written texts as well as the evaluation they receive and (b) help them outdo EAP students who do not practice formulaic sequences?
3. Will both saliency and explicit instruction of formulaic sequences help EAP students (a) reduce the number of linguistic errors they make in different writing tasks and (b) write more accurately than their counterparts in control groups?
4. Will the increase, if any, in the use of formulaic sequences positively correlate with receiving better evaluation and producing longer timed written texts?
5. Will the increase, if any, in the number of the utilized formulaic sequences negatively correlate with the number of linguistic errors made in different writing tasks?

Chapter Four: Methods

The fourth chapter falls into nine sections, each of which discusses one aspect of this quantitative research study. The first section provides a brief description of the present research study and summarizes AlHassan's (2016) pilot study, which helped partly pilot some of the instruments. The second section explains the study's quasi-experimental design and clarifies the pedagogical and methodological rationale for adapting this design. In the third section, information about the context in which the study took place, the participants who volunteered in this study, and the sampling strategies used to recruit participants is provided. The fourth section, in turn, discusses the sources from which the target formulaic sequences were extracted. The instruments used and the procedures followed for both pedagogical intervention and data collection are discussed in the fifth, sixth, and seventh sections and are followed by a description of the methods applied to quantify the textual data in the eighth section. The last section of this chapter presents the statistical procedures that were followed in order to impute missing data, check statistical assumptions, and answer the five research questions that guided this research study.

Overview

As mentioned in the third chapter, this quantitative research study was designed to investigate the effectiveness of different presentation methods of formulaic sequences in augmenting EAP students' academic writing skills. The study was partly based, in its instruments, on AlHassan's (2016) pilot study, which aimed (1) to investigate the effects of an explicit instructional approach to formulaic sequences on EAP students' academic reading and writing proficiency, (2) to explore EAP teachers' perspective on the

effectiveness of such an instructional approach, and (3) to pilot some of the instruments for this research study.

AlHassan's (2016) pilot research study had a *mixed sequential explanatory design* with a major quantitative phase followed by a minor qualitative phase and was informed by Activity Theory (Creswell & Plano Clark, 2011; Dörnyei, 2007; Engeström, 1987). It was conducted in an EAP program at a large Canadian community college. The pool of participants included 12 intermediate EAP students who were registered in the same course in which the intervention took place and three judges who had experience teaching EAP at the same community college. The study was divided into several phrases during which pedagogical intervention (i.e., explicitly teaching 63 formulaic sequences) and data collection (i.e., eliciting three different reading and writing tasks, evaluating the elicited tasks, and interviewing the three EAP judges) took place. The study yielded two major findings: (1) a focused instructional approach promoted the 12 EAP students' acquisition of formulaic sequences, reduced their errors, and helped them receive better evaluation; (2) the three EAP teachers had mixed views regarding the pedagogical utility of explicitly teaching formulaic sequences in EAP contexts.

The list of formulaic sequences as well as the instruments used in AlHassan's (2016) pilot study were revised and modified to account for the objectives of the present study, that is, exploring the effects of different instructional approaches on EAP students' summaries of graphical information and responses to essay prompts. The list of formulaic sequences was revised and expanded to include 72 formulaic sequences, and the pedagogical instruments were revised and adapted to include the target formulaic sequences, fit the time allocated for the pedagogical intervention, and create instruments

for three different groups. The instruments used for data collection were also replaced with three graphical summary tasks and three essay prompts, and the analytical rubric for evaluating graphical summaries was used as the basis for creating another one for the evaluation of essay responses.

The current quantitative research study was conducted over two terms (Summer 2016 and Fall 2016) in an EAP program at a large Canadian university. Three different types of pedagogical intervention—that is, teaching formulaic sequences explicitly, making formulaic sequences salient in input, and explicitly teaching some materials that relate to linguistic aspects—were randomly assigned to different groups of upper-intermediate and advanced EAP students. The EAP students, in turn, were required to complete six different writing tasks, three graphical summaries and three essay responses, at three time points, that is, a pretest before delivering the target pedagogical intervention, a posttest at the end of the training period, and a delayed posttest one month after the end of the training period. The collected written texts were evaluated by an EAP judge and were converted into numerical variables. The missing scores were manually imputed, and the derived numerical variables were then analyzed using several statistical tests.

Design and Rationale

To answer the research questions listed in Chapter Three, a quasi-experimental research study was designed. A quasi-experiment is a special form of research design which resembles experimental research that usually takes place in laboratory-like settings and has three major characteristics: randomly assigning participants to different groups (i.e., control and experimental), manipulating the target experimental variable(s), and fully controlling for intervening and unrelated factors (Dörnyei, 2007; Reichardt, 2009;

Tabachnick & Fidell, 2007a; Vogt, 2007). Similar to true experiments, quasi-experiments have control and experimental groups and involve researchers' attempts to control for extraneous factors (Dörnyei, 2007; Mark & Reichardt, 2009; Vogt, 2007). Unlike the former, however, researchers may not be able to fully control for all intervening factors, and they assign preexisting groups to different treatments (Campbell & Stanley, 1963; Dörnyei, 2007; Spada, 2005; Vogt, 2007).

Quasi-experiments are very common in educational research, in which conducting research studies in laboratory settings is neither feasible nor realistic (Dörnyei, 2007; Reichardt, 2009; Spada, 2005; Vogt, 2007). Classroom researchers often need to work with predefined groups of students and, hence, assign target interventions to preexisting class groups. While randomized-group designs are essential to remove researchers' bias and average initial differences among subjects, working with intact class groups has its own advantages in SLA research as it usually produces results that apply to L2 classrooms more than those obtained in true experiments (Dörnyei, 2007; Mark & Reichardt, 2009; Spada, 2005; Tabachnick & Fidell, 2007a; Vogt, 2007).

One of the most widely used quasi-experimental designs in educational research is the pretest-posttest-nonrandomized-control-group design (Campbell & Stanley, 1963; Dörnyei, 2007; Leedy, 1997; Vogt, 2007). This design includes an experimental group that receives a target pedagogical intervention and a control group that shares in common every characteristic with the experimental group but does not receive any pedagogical intervention; the participants in both groups are measured before and after intervention to detect any differences that can be attributed to the training period (Campbell & Stanley, 1963; Dörnyei, 2007; Leedy, 1997; Reichardt, 2009). Similar to the above-mentioned

design, this quasi-experimental research study has multiple groups and levels of measurements. However, it differs from the former in that it includes (1) three major groups at two different proficiency levels (i.e., upper-intermediate and advanced), (2) three levels of measurements (i.e., pretest, posttest, and delayed posttest), and (3) compensation materials for the control groups (see Table 1 below).

Table 1 *Quasi-Experimental Design*
Experimental Focused-Instruction Groups: $O_{(1)} \rightarrow X \rightarrow O_{(2)} \rightarrow O_{(3)}$ Experimental Saliency Groups: $O_{(1)} \rightarrow X \rightarrow O_{(2)} \rightarrow O_{(3)}$ Control Groups: $O_{(1)} \rightarrow C \rightarrow O_{(2)} \rightarrow O_{(3)}$
Note. X indicates treatment; $O_{(1)}$, $O_{(2)}$, and $O_{(3)}$ refer to pretest, posttest, and delayed posttest respectively; C refers to treatment using compensation materials

Adapted from Leedy's (1997) Summary of Experimental Methodology Table (p. 232)

The rationale for adapting the pretest-posttest-nonrandomized-control-group design is pertinent to the objectives of this research study, the merits of this design, and the resemblance of this design to one of the most cited experimental research designs. The two major objectives of this research study were (a) to inquire into the effectiveness of two approaches to presenting formulaic sequences in promoting EAP students' learning gains of this language phenomenon and (b) to identify any possible differences that could be attributed to different pedagogical interventions. These two major objectives necessitated expanding the adapted design to include a second experimental group (i.e., saliency) and a third measurement (i.e., a delayed posttest) in order to explore the learning gains that transcend immediate learning.

Measuring the participants' performance multiple times is not only justifiable in research that focuses on "the lasting or residual effects of a treatment" (Stevens, 2002, p. 494), but it is also crucial because of the assumption that teaching may have limited

effects on learning gains (Alali & Schmitt, 2012). It is also a common practice in research on L2 vocabulary acquisition as it allows researchers to explore the extent to which the target vocabulary items are successfully acquired (Nation & Webb, 2011). As for the compensation materials, although in most quasi-experimental research studies the control group does not receive any pedagogical intervention (Dörnyei, 2007), compensation materials were developed for the control groups in an attempt to match the control groups to the experimental groups. That is, given the amount of time spent on the pedagogical intervention in the four experimental groups (i.e., a total of five hours), delivering extra worksheets to the control groups helped equate the control and the experimental groups on two aspects other than being registered in the same EAP program at the same proficiency levels; that is, all groups practiced extra materials and received instruction and feedback from the researcher who was not their instructor.

From a methodological point of view, the pretest-posttest-nonrandomized-control-group design was adapted because of its merits. Among the major merits of this design is working with intact class groups, which helps researchers provide implications that better inform pedagogy in L2 classrooms (Dörnyei, 2007; Spada, 2005). Working with intact groups also promoted the inclusiveness of this research study in that it helped the researcher provide equal opportunities to both participants and nonparticipants in the groups to which she was granted access and include all students in those groups who were willing to participate in this research study but might not have been available to meet with her outside class hours. Another methodological merit of this design is measuring the participants' performance before and after intervention. While measuring participants after intervention helps identify differences between different groups, the

pretest element in this design is particularly important in classroom research as it helps researchers identify differences within groups and control for initial differences between groups that may be attributed to individual differences and/or any instruction that has taken place prior to delivering intervention (Dörnyei, 2007; Spada, 2005; Tabachnick & Fidell, 2007a).

The other methodological merit of this design relates to including a control group. Several researchers (e.g. Dörnyei, 2007; Reichardt, 2009) warn against conducting a research study without the inclusion of a control group. While such basic quasi-experimental research has the advantage of being methodologically simple and straightforward, it "has a very important drawback; namely that the difference between the pretest and the posttest observations could be due to one or more other factors besides the treatment or intervention…. [which is] a threat to internal validity" (Reichardt, 2009, p. 48). The inclusion of a control group was crucial because of the context of this research study (i.e., an EAP program), for it functioned as a baseline for comparison and controlled for the above-stated threat to internal validity—which is referred to in the literature as the *maturation* effect and is defined as the participants' natural development and progression over time (Campbell & Stanley, 1963; Christ, 2007; Dörnyei, 2007).

The above-mentioned design was also adapted because it is very similar to a major experimental design that combines different randomized groups with repeated measurements, that is, the mixed between-within-subjects experimental design (e.g. Meyers, Gamst, & Guarino, 2006; Stevens, 2002; Tabachnick & Fidell, 2001, 2007a, 2007b, 2014). The mixed between-within-subjects design is very similar to the adapted design for this research study in that both designs include multiple groups and repeated

measurements. However, they differ in that the former includes random assignments of subjects to different treatments and the latter involves random assignment of different treatments to preexisting class groups (Spada, 2005; Tabachnick & Fidell, 2001).

Context and Participants

The selection of both the context and the participants was pertinent to the major objectives of this quasi-experimental research study.

Context. Since this research study focused on the effects of a language phenomenon on EAP students' academic writing skills, the context of the study was a language program with a special focus on EAP, defined as a language program that is mainly centered around upgrading L2 users' academic skills to help them successfully handle academic tasks in English-medium universities (Hyland & Hamp-Lyon, 2002; Kasper, 2000; Reid, 2001). One such context when the study was conducted was an EAP program at a large Canadian university. When the study took place in Summer 2016 and Fall 2016 terms, the EAP program in which the study took place consisted of several levels, ranging from beginner to advanced. At each level, instruction was delivered on a daily basis by different instructors over approximately 14 weeks and was tailored to address several learning outcomes essential to succeeding in English-medium universities: mastering a wide range of academic reading skills, learning how to write different academic texts, being proficient in a variety of academic listening and speaking skills, and promoting both lexical and syntactic competency.

The selection of the target EAP program as a context for this study was influenced by several factors. When the study was conducted, the program was geared to help L2 learners' acquire and master skills decisive for succeeding in English-medium

universities, which is a major attribute of an EAP program. Moreover, some of the learning outcomes for the intermediate and advanced levels, that is, writing summaries and essays, were in line with the two target writing tasks in this research study. Furthermore, the program in its assessment guidelines that are shared with its instructors, of whom the researcher was one, placed a special focus on the importance of one aspect of formulaic sequences (i.e., collocations) in both instruction and evaluation at the upper-intermediate and advanced levels. Therefore, the nature of the program (i.e., EAP), the special focus on some writing tasks (i.e., summaries and essays), and the attention paid to one aspect of formulaic language in both instruction and assessment (i.e., collocations) made the selected program a context that was in line with the objectives of this study.

As for the teaching approach in the selected program, although the program does not explicitly state a teaching approach, it seems to adapt the task-supported approach to language teaching. To help provide a rationale for describing the program teaching approach as task-supported, the characteristics that constitute a task and the major attributes that distinguish the task-supported approach from the task-based approach are briefly tackled. A task is defined as a meaning-focused activity that relates to situations L2 learners may encounter in contexts other than an L2 classroom; it may focus on one or several integrated language skills; it engages L2 learners in goal-oriented communicative activities that center on authentic language and demand cognitive processes in order to achieve an ultimate communicative outcome (Beglar & Hunt, 2002; Celce-Murcia & Olshtain, 2000; Ellis, 2003; Skehan, 1996; Van den Branden, 2016).

The task-supported approach to language teaching is a weak form of CLT that combines CLT (i.e., a focus on meaningful communication) with traditional instruction

(i.e., a focus on form); it utilizes tasks to provide L2 learners with communicative practice of what may have been explicitly taught using a traditional instructional approach (Ellis, 2003, 2005b; Skehan, 1996; Van den Branden, 2016). Unlike the task-based approach which is a strong form of CLT and in which instruction takes a back seat and "tasks provide the basis for an entire language curriculum" (Ellis, 2003, p. 30), tasks in the task-supported approach are "a necessary but not a sufficient basis for language curriculum" (Ellis, 2003, p. 28). They "are a vital part of language instruction, but they are embedded in a more complex pedagogical context" (Skehan, 1996, p. 39). Combining CLT with traditional approaches is what distinguishes task-based teaching, in which "teaching [is] based exclusively on meaning-focused tasks," from task-supported teaching, in which "teaching… uses tasks to practise pre-selected and presented linguistic forms" (Ellis, 2005b, p. 6).

Similar to the task-supported approach, the objectives, learning outcomes, curriculum, and assessment guidelines of the EAP program in which the study took place seem to combine aspects of CLT and traditional approaches. When the study took place, the objective of the selected EAP program focused on aspects that transcend the language classroom (i.e., academic tasks, meaningful communication, sociocultural awareness, etc.). These objectives were achieved through integrating different receptive and productive skills, the use of authentic materials, and the inclusion of some sociocultural activities. Such a focus on communication, authentic materials, and sociocultural activities can be linked to CLT, which "aims to develop the ability of learners to use language in real communication… [and] to function interactionally and transactionally in an L2" (Ellis, 2003, p. 27).

The learning outcomes of the selected program, in turn, mirror the above-mentioned characteristics of a task; that is, writing emails to professors, discussing academic interests, avoiding plagiarism, participating in debates, doing presentations, and so on can be labeled as tasks because they focus on one or integrated language skills and involve L2 users in cognitive processes to complete and master meaningful tasks they may encounter in contexts beyond the language classroom. These tasks, being the learning outcomes of the program, are placed at the core of the curriculum.

The program curriculum, however, has some aspects of traditional instructional approaches. When the study took place, the instructors were expected to cover several textbooks at each level along with any materials they wished to develop in order to address the students' needs. Of the assigned textbooks at some lower levels were grammar books with decontextualized activities, which can be seen as an aspect of a traditional instructional approach. Moreover, the assessment guidelines shared with instructors partly focused on structural aspects in writing assessment and multiple-choice tests in reading and listening assessment. These aspects do not fully mirror pure task-based instruction, which focuses primarily on meaning and fluency (Ellis, 2003; Skehan, 1996).

It is worth noting here that describing the instructional approach of the selected program as task-supported was solely based on the researcher's interpretation and understanding of some of the information posted on the program website and the materials to which she had access as an instructor. It was by no means intended to assign a teaching approach to the program or evaluate its effectiveness; nor did it aim to claim that the teachers adhered to one teaching approach or another as the researcher was

unaware of the teaching approaches implemented by other teachers in their classrooms. The sole aim of this description was to provide some information regarding the program to help the reader get a clear view of the context in which the study took place.

Participants. The pool of participants comprised both EAP students and an EAP judge. When the study took place, the EAP students were registered in the highest two levels of the selected program. These two levels, henceforth referred to as upper-intermediate and advanced levels, engaged L2 learners in a variety of academic tasks including writing reports, summaries, and essays, dealing with academic reading texts, giving academic presentations, to list but a few. In Summer 2016 and Fall 2016, there were several upper-intermediate and advanced groups, of which the researcher was granted access to 7 class groups (three upper-intermediate and four advanced groups) with 115 students. There were also intermediate groups in Summer 2016 and Fall 2017, but they were excluded because the researcher was not granted access to an intermediate group in the summer. The intermediate groups in Fall 2016 were excluded because no data were collected from the same level in the summer term and the researcher was teaching one of the intermediate groups during the fall term. The beginners and lower-intermediate levels, in turn, were excluded because having enough proficiency to produce summaries and essays was a major criterion to recruit participants.

The three target pedagogical interventions were randomly assigned to the first six groups to which the researcher was granted access based on the sequence of the emails the researcher received back from the instructors after sending them an invitation email (see Table 2 below). As for the seventh group, it was assigned salient formulaic sequences because, at the time the researcher was granted access to that group, the upper-

intermediate and the advanced saliency groups had the lowest number of participants. The three randomly assigned pedagogical interventions were delivered to the 115 students who, in turn, completed the practice materials and received feedback on the practice writing tasks they completed during the researcher's visits. Only those who were 18 years or older, completed at least two rounds of data collection, and accepted to participate in this study by signing a consent form were recruited as participants.

Table 2

Random Assignment of Pedagogical Interventions

	Upper-Intermediate	**Advanced**
First Email	Focused Inst.: Teacher_____	Focused Inst.: Teacher_____
Second Email	Saliency: Teacher_____	Saliency: Teacher_____
Third Email	Control: Teacher_____	Control: Teacher_____

Of the 115 EAP students, 67 students (39 males and 28 females) from six different first language backgrounds (i.e., Mandarin, Arabic, French, Spanish, Russian, and Persian) accepted to participate in this study. It is worth noting here that one potential participant was excluded because he only completed one round of data collection, that is, the pretest. Had he completed two rounds of data collection, the total number of participants would have been 68. Another point worth mentioning is that all the names that will appear in the subsequent sections and appendices are pseudonyms that the participants chose or the researcher assigned to the ones who did not choose a pseudonym. The pool of participants respectively included 13 and 19 students from the upper-intermediate and advanced focused instruction groups, five and 13 students from the upper-intermediate and advanced saliency groups, and 10 and seven students from the upper intermediate and advanced control groups.

In addition, one EAP judge, a female English native speaker with an approximately four-year experience teaching EAP in other programs, was recruited to

blindly evaluate the collected written texts. Recruiting an EAP judge from a different EAP program was intended to promote the validity of the evaluation and keep the identity of the EAP participants anonymous. That is, being an instructor in a different program, the evaluation of the EAP judge could not have been influenced by any of the instruments used in this study, which an instructor in the selected program may have incidentally had access to; nor could she accidently identify any of the participants based on their style, errors, etc.

Based on the recruitment procedure described above, the sampling strategies used in this study have characteristics of both *stratified random sampling* and *convenience sampling*. *Stratified random sampling* is one form of *probability sampling* which is viewed as the most scientific sampling strategy that grants all the subjects of a target population equal opportunities to volunteer in a research study (Dörnyei, 2007; Feild, Pruchno, Bewley, Lemay, & Levinsky, 2006; Tansey, 2007; Vogt, 2007). *Stratified random sampling* is most efficiently used in research studies that have a specific focus, and it is achieved through two stages: selecting groups based on stratifying characteristics (e.g. age, gender, etc.) that are relevant to the research objectives and randomly assigning participants to different groups (Dörnyei, 2007; Feild et al., 2006; Vogt, 2007). Similar to *random sampling*, *stratified random sampling* aims at eliminating bias by randomly assigning participants to different groups; however, it differs from the former in that it stratifies subjects based on several shared characteristics prior to randomly assigning them to different groups (Dörnyei, 2007; Tansey, 2007; Vogt, 2007).

Convenience sampling, also known as *opportunity* or *availability sampling*, is a practical non-probability sampling strategy that is commonly utilized in quasi-

experiments in general and in SLA research in particular (Dörnyei, 2007; Tansey, 2007; Vogt, 2007). As its name indicates, *convenience sampling* allows researchers to recruit participants from groups which researchers can approach and/or have access to without much effort or strict selection criteria (Dörnyei, 2007; Feild et al., 2006; Tansey, 2007; Vogt, 2007). In many cases, *convenience sampling* is combined with *purposive sampling*, another form of *non-probability sampling* strategies that require groups to possess some attributes pertinent to the research objectives (Dörnyei, 2007; Feild et al., 2006).

The description of the sampling strategy in this research study as *stratified random sampling* could be ascribed to the criteria that were used to stratify tentative groups (i.e., having enough language proficiency to produce summaries and essays) and to the procedure described above to randomly assign different pedagogical interventions to different groups. The characteristics of *convenience sampling* strategies, on the other hand, could be related to the fact that the researcher recruited groups from the program at which she was an instructor. Although the researcher was not the instructor of the groups in which she delivered the target pedagogical interventions, the fact that she taught in the same program may have facilitated her access to different groups. Moreover, recruiting any upper-intermediate and advanced EAP students who agreed to participate in this study and any EAP judge who accepted to evaluate the collected written texts could again be linked to *convenience sampling* that has some characteristics of *purposive sampling* in that the researcher recruited any participants who were willing to take part in this research study, but those participants were purposively selected from the target context, that is, EAP.

Formulaic Sequences: Sources of Selection

Selecting formulaic sequences for this research study constituted a pivotal and assiduous stage. The crucial nature of this stage could be attributed to the fact that this language phenomenon was at the center of the entire investigation. The assiduity with which the target formulaic sequences were chosen, in turn, could be mainly ascribed to (a) the nature of this language phenomenon as being highly register- and genre/discipline-bound (Cortes, 2004, 2008; Hyland, 2008a), (b) the assumption that teaching formulaic sequences atypical of a register or a genre may have adverse effects on the quality of EAP students' academic writing akin to those of not addressing them (Ellis et al., 2008; Granger, 1998), and (c) the need to extract formulaic sequences for the two target writing tasks in this study (i.e., graphical summaries and essay responses).

In order to select formulaic sequences, two academic sources typical of the target register and the target writing tasks were consulted. The first source comprised developed lists based on published and unpublished academic writing in four disciplines (i.e., Byrd & Coxhead, 2010; Hyland, 2008a), published and general academic writing (i.e., Chen & Baker, 2010; Nattinger & DeCarrico, 1992; Stubbs & Barth, 2003), published research and a mega corpus (i.e., Simpson-Vlach & Ellis, 2010), or mega corpora (i.e., Conrad, 1999; Liu, 2012). These lists were deemed highly reliable because they are compiled by experts in the field of discourse analysis, and an essential part of such research is to provide the reader with detailed information about the register under investigation, the corpus used in the study, and the method of selection. Such information was crucial during the selection stage in order to determine whether the selected formulaic sequences were recurrent in the spoken or the written register and whether they were peculiar to a

specific discipline, multiple disciplines, or general academic writing. All the selected formulaic sequences from the first source were typical of written academic discourse rather than a specific discipline, and the majority of the selected formulaic sequences were recurrent in two or more published lists (see Appendix A).

The second source was the COCA. The COCA is the largest available 'monitor' corpus of American English that consists of 520 million words and is equally divided between five registers—spoken, fiction, newspapers, popular magazines, and academic journals (Davies, 2008, 2009, 2010). It was first released in 2008 with 385 million words in order to overcome the limitations of other online corpora by annually including contemporary texts from a wide variety of genres (Davies, 2009, 2010). The dynamic nature, various functions, and advanced interface of the COCA make it an accessible source that can be used for various functions, such as checking raw and normalized frequencies of words or phrases in a specific genre and time period, identifying lists of collocates for stem words, comparing the usage of different words, checking concordance lines, and tracking recent changes in language use (Davies, 2009, 2010).

Of the various functions of the COCA, two functions (i.e., the *Chart* and the *Word and Phrase. Info*) were used to extract formulaic sequences. The *Chart* function is one of the basic frequency functions that allows the COCA users to check raw and normalized frequencies of words, phrases, or syntactic constructions in five genres, fifty sub-genres, and four time periods (Davies, 2008, 2009). The *Word and Phrase. Info* function, in turn, is a more advanced interface that is divided into *Frequency Lists* and *Analyze Texts* and can be customized by genres (i.e., *All Genres* or *Academic*). The *Frequency Lists* function allows the COCA users to check the frequency of a word in different genres or

sub-genres along with a number of collocates, synonyms, and concordance lines (Davies, 2008). The *Analyze Text* function, on the other hand, allows users to upload a text to the COCA in order to analyze the frequency of the words or phrases in a target register, genre, or sub-genre (Davies, 2008). It also provides users with definitions, collocates, and concordance lines for selected words or phrases (Davies, 2008).

The *Chart* and the *Word and Phrase. Info* functions were used to check the normalized frequencies of phases and stem words and to identify collocates for the latter, respectively. Prior to using these two COCA functions, a minimum frequency cut-off of 30 occurrences per one million words in the academic register was set for this study. The rationale for choosing this frequency cut-off is that it falls within the range of most corpus research studies on formulaic sequences—between 10 and 40 occurrences per one million words (Simpson-Vlach & Ellis, 2010). It is also a more conservative threshold than the one used in Cortes (2002, 2004), Hyland (2008a, 2008b), and Liu (2012), that is, 20 occurrences per one million words, or the one set for Chen and Baker's (2010) as well as Ädel and Erman's (2011) research studies (i.e., 25 occurrences per one million words). Focusing on normalized frequencies in the academic register, which comprises academic texts from approximately 100 peer-reviewed journals (Davies, 2008, 2009, 2010), was driven by the study context (EAP) and the genre of writing (academic). Using both sources helped extract 72 formulaic sequences that have different syntactic structures and execute various discursive functions in academic writing (see Appendix A).

Instruments

Three different sets of instruments were designed in order to collect textual data, inform textual evaluation, and deliver the three target pedagogical interventions.

Instruments for data collection. The repeated-measures design of this study along with its major objectives necessitated the development of three different sets of writing prompts with two different tasks. These three different sets included three line graphs that were generated using the graphing tools in Microsoft Excel and presented fictitious changing trends over a specific time period (i.e., Task 1) and three essay prompts which constituted Task 2 and focused on three different essay types (i.e., cause/effect, compare/contrast, and opinion essays). The three line graphs had the same format (i.e., a title, labels, scales, and four lines), and they all instructed the students to summarize hypothetical graphical information in a paragraph format; the three essay prompts also had identical instructions (i.e., writing an essay in response to each essay prompt).

The first set of writing prompts included a line graph and a cause and effect essay prompt. The line graph visually presented changes in the percentage of international students in four different hypothetical universities between 1999 and 2005 (see Figure 1 below). The essay prompt, in turn, instructed EAP students to think of some negative effects that technology addiction might have on people's physical and mental health and to *discuss 2 or 3 negative effects of technology addiction*.

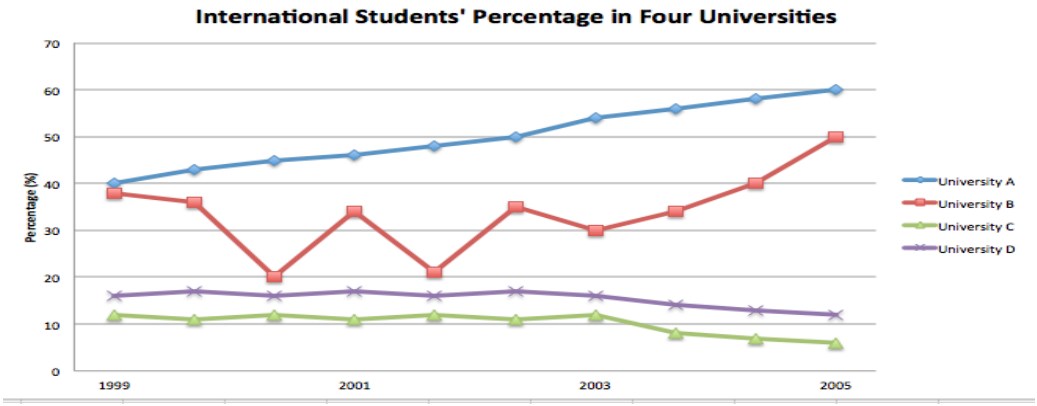


Figure 1. Task 1 (prompt 1)

The second set of writing prompts included a line graph and a compare/contrast essay prompt. The line graph was about the production of four different crops in a hypothetical country over a fifteen-year period (see Figure 2 below); the essay prompt required the EAP students to reflect on the differences between the education system in their home countries and the one in Canada and to *discuss 2 or 3 differences between the two*.

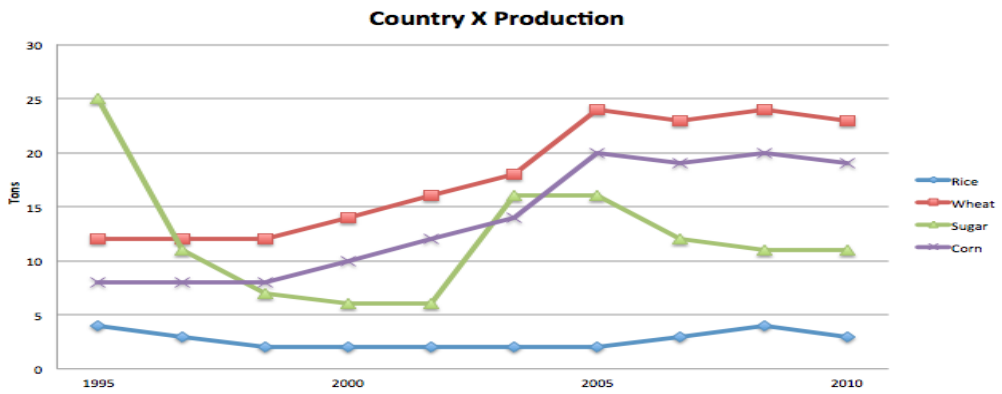


Figure 2. Task 1 (prompt 2)

The third set consisted of a line graph with fictitious data about the percentage of homeless people in four hypothetical countries from 1940 to 2000 (see Figure 3 below) and an essay prompt that instructed the students to express the extent to which they agree disagree with the following statement: *Some people are born intelligent and others are not.*

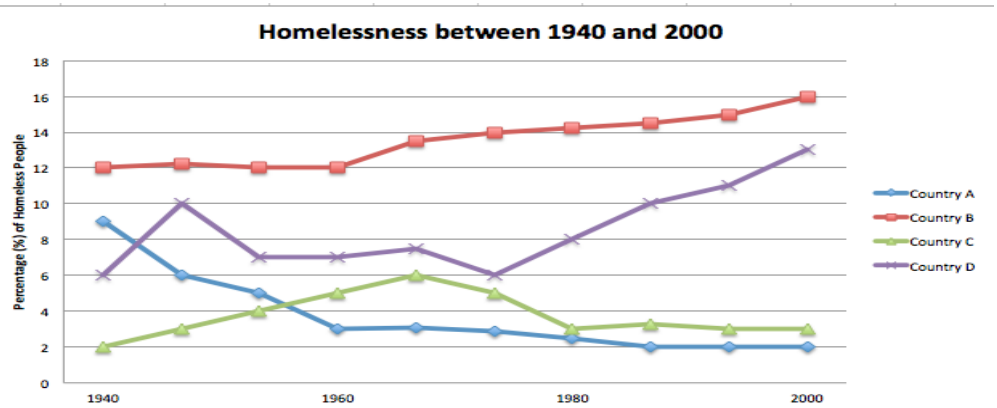


Figure 3. Task 1 (prompt 3)

The three sets of writing prompts were interchangeably administered at each stage of data collection following all possible combinations listed in Table 3 below.

Table 3

The Random Administration of the Three Sets of Writing Prompts

	Pretest	Posttest	Delayed Posttest
1st Combination	Prompt 1	Prompt 2	Prompt 3
2nd Combination	Prompt 1	Prompt 3	Prompt 2
3rd Combination	Prompt 2	Prompt 3	Prompt 1
4th Combination	Prompt 2	Prompt 1	Prompt 3
5th Combination	Prompt 3	Prompt 2	Prompt 1
6th Combination	Prompt 3	Prompt 1	Prompt 2

Designing three sets of writing prompts and interchangeably administering them at each stage of data collection is based on pedagogical and methodological grounds. Including both graphical and essay prompts as elicitation instruments was germane to the objectives of this study, that is, investigating the effects of formulaic sequences on EAP students' summaries of graphical information and responses to different essay prompts. It was also pertinent to the context of this research study (i.e., EAP). In addition to writing different types of essays, EAP students are usually introduced to economic discourse to practice reporting information in data sets, a task that is widespread in academic contexts and is increasingly viewed "as part of educating for science literacy" (Glazer, 2011, p. 201). Such a practice is introduced to help EAP students utilize supporting details from reliable sources (Oshima & Hogue, 2006; Williams, 2013), is provided in some writing textbooks (e.g. *LEAP Advanced: Reading and Writing* and *Writing Academic English*), and is common in English proficiency tests. These two tasks were also in line with the learning objectives of the selected EAP program, which focuses on summaries and different types of essays.

From a methodological viewpoint, the interchangeable administration of different

sets of writing prompts aimed to attenuate two major threats to internal validity—that is, the *practice effect* which refers to an improvement in the participants' performance because of the repeated administration of testing materials and the *instrumentation* effect which occurs when changes in participants' performance are the result of using inconsistent testing materials (Beglinger et al., 2005; Campbell & Stanley, 1963; Christ, 2007; Dörnyei, 2007; Kazdin, 1981). These two effects may invalidate the findings of a research study because they lead researchers to invalid conclusions in that they may inaccurately attribute changes in the participants' performance to an administered treatment while, in reality, these changes are the result of the experience the participants have gained because of repeatedly taking the same test or the use of different elicitation instruments with inconsistent levels of difficulty at different stages of data collection (Christ, 2007; Dörnyei, 2007; Kazdin, 1981).

One way to eliminate or at least minimize these two major threats to internal validity is the interchangeable use of comparable *alternate forms* (i.e., different testing materials) at each testing stage (Beglinger et al., 2005; Benedict & Zgaljardic, 1998; Christ, 2007). The use of Microsoft Excel graphing tools to create line graphs and the inclusion of general topics that are typical of language programs (i.e., technology, education, and intelligence) helped design three different sets of writing prompts with a relatively similar level of difficulty and functioned as *alternate forms*. Their random assignment, in turn, was intended to rule out the *practice effect* and the *instrumentation* effect and, hence, to bolster the internal validity of this research study.

Instruments for assessment. In addition to the three sets of writing prompts, two analytical rubrics (see Appendix B) were used as elicitation instruments in order to

collect the second part of data in this research study (i.e., textual evaluation). The two analytical rubrics were utilized because of their benefits in writing assessment, that is, offering descriptive guidelines with separate grading scales for different writing aspects to promote unbiased and consistent evaluation of students' work (Jonsson & Svingby, 2007; Rezaei & Lovorn, 2010; Spence, 2010). The descriptive criteria in both rubrics were broadly based on the common descriptors in analytical rubrics, that is, lexis, grammar, organization, etc. (Rezaei & Lovorn, 2010; Spence, 2010), which are also part of the guidelines for writing evaluation in the study context and are among the major markers of writing proficiency that include, but are not limited to, syntactic and lexical accuracy as well as coherence and cohesion (e.g. Ellis, 2015; González-Bueno & Pérez, 2000; Lembke, Deno, & Hall, 2003; Perkins, 1980; Polio, 1997; Schoonen et al., 2011).

The first rubric (see Table B1) was piloted in AlHassan's (2016) study and used in the present study to guide the EAP judge's assessment of Task 1 (i.e., graphical summaries). It had three separate scoring scales with three different descriptive words (i.e., unsatisfactory, average/good, and excellent) and three numerical grade ranges (i.e., 10-40 points, 50-70 points, and 80-100 points). It was divided into five categories that respectively focused on the accuracy and conciseness of content, the appropriateness of format and presence of essential paragraph elements, the precision and variety of lexis, the correctness of grammar and structure, and the indices of coherence and cohesion. The second analytical rubric (see Table B2) was created to inform the EAP judge's evaluation of Task 2 (i.e., essays). Similar to the first analytical rubric, it had five major categories and three scoring scales with descriptive words and grade ranges, but the descriptive statements in the second rubric were partially reworded to focus on the structure and

components of essays. The five categories of the second analytical rubric mainly centered on the relevance and sufficiency of controlling ideas and supporting details, the appropriateness of the essay format and the presence of all essay elements, the accuracy and variation of lexis, the knowledge of grammar and punctuation, and the degree of coherence and cohesion.

Instruments for pedagogical intervention. In order to introduce the 72 elicited formulaic sequences to the experimental groups and writing materials to the control groups, three sets of pedagogical instruments were used. Each of the three sets (see Appendix C, worksheets 1 through 30) consisted of 10 worksheets that were designed based on the recommendations provided in SLA research for teaching writing and/or integrating lexical items into classroom materials and included explanatory information (four worksheets), practice activities (four worksheets), or a combination of both explanation and practice (two worksheets). The three sets were also comparable in content in that the first two worksheets in each set tackled similar contents (i.e., topic and concluding sentences), and the remaining worksheets presented explanatory information that was followed by decontextualized group or independent practice with the former being the form of practice for the majority of the worksheets because of the effectiveness of communicative practice in L2 acquisition in general and lexical acquisition in particular (Ellis, 2003; Fan, 2009; Lewis, 2002). The three sets, however, differed in their focus and their presentation methods.

Instruments for the Experimental Focused Instruction Groups. The first set of pedagogical instruments consisted of 10 worksheets (see Appendix C, worksheets 1-10) that were piloted in AlHassan's (2016) study and then revised and utilized in the present

study to explicitly teach the use and function of the 72 target formulaic sequences in academic writing and to have the EAP students practice the use of the presented formulaic sequences in different types of writing activities.

Writing a topic sentence. The first worksheet designed for the focused instruction groups had a dual function. It first introduced a brief description of the function and the elements of a topic sentence along with two sample topic sentences for a hypothetical line graph. It also included three different types of graphs that were generated using the graphing tools in Microsoft Excel and two sentence frames to provide the students with both controlled and uncontrolled group practice of the formulaic sequences that can be used to construct topic sentences. In other words, the students were asked to work in groups of four to complete the sentence frames for the first two practice graphs and to write a topic sentence from scratch for the third one.

Using sentence frames with slots to construct stretches of meaningful discourse is believed to be particularly useful practice because such frames provide L2 learners with the repeated exposure and practice necessary for the acquisition of formulaic sequences that function as backbones for the creative construction of more accurate language segments (Lewis, 1997; Nattinger & DeCarrico, 1992; Willis, 2003; Woolard, 2000). Writing a topic sentence from scratch, in turn, was adapted from Nation's (1991, 2001) delayed copying technique. However, instead of having the students copy a topic sentence word for word as suggested by Nation (1991, 2001), they were encouraged to copy chunks from the examples provided in the same worksheet and use them creatively to construct a topic sentence for the third chart.

Writing a concluding sentence. The second worksheet focused on concluding sentences. Similar to the first worksheet, it introduced the definition and key characteristics of concluding sentences in academic writing and graphical tasks along with two sample sentences that included several formulaic sequences typical of conclusions and/or concluding sentences. It also included the same practice graphs that were introduced in the first worksheet. Instead of providing the participants with two sentence frames, they were dictated two concluding sentences and were instructed to check their sentences in groups of four. They also had to construct a third concluding sentence in groups following the examples provided and the two dictated sentences.

While having the students write their own concluding sentences was adapted from Nation's (1991, 2001) delayed copying technique, dictating two sentences to the participants was adapted from both dictogloss and classic dictation activities. In other words, the dictation procedure was identical to the one used in dictogloss in that each sentence was dictated twice at a normal pace without pauses and was reconstructed in groups (Wood, 2002). Traces of classic dictation were also present in this activity; that is, none of the language chunks in the dictated sentences was written on the board, and the dictation itself was composed of sentences rather than a text. Both dictation and dictogloss activities are particularly effective in promoting the acquisition of formulaic sequences as they prime L2 learners' noticing of formulaic sequences and encourage them to retain the dictated language in their working memories as chunks before producing it in writing (Nation, 1991, 2001; Wood, 2002).

Collocations. The third worksheet included 26 collocations that were categorized based on their semantic functions and syntactic structures. Each category included two

sets of collocations that had different grammatical structures (i.e., noun phrases and verb phrases) with sample sentences and focused on one type of graphical trends (i.e., increasing, decreasing, fluctuating, or stable). Creating a list of collocations was in line with Conzett's (2000) recommendation that language teachers develop lists of formulaic sequences that are peculiar of a topic and/or a task as a prewriting activity in order to familiarize their students with the lexical items that they can use in their language production. Moreover, categorizing collocations based on their shared stem words and grammatical structures and contextualizing them may promote L2 learners' comprehension of the underlying meaning of each set of lexical items, their understanding of the use of lexical items with different grammatical categories, their perception of language as being composed of chunks, and their accurate utilization of these chunks in their language production (Haworth, 1996; Nation & Newton, 1997; Nesselhauf, 2003; Willis, 2003).

Collocation practice. In order to help the students practice the use of the collocations presented in the third worksheet, the fourth worksheet consisted of 10 collocations that were chosen from the third worksheet and put in a box and five line graphs that were generated using the graphing tools in Microsoft Excel. This practice worksheet required the students to work in groups and complete two tasks. They first had to choose two collocations that best describe each line graph from the box. They were then instructed to write two semantically and syntactically accurate sentences to describe each line graph. Using sets of lexical items actively in communicative activities and associating them with images may promote deep processing and, thus, the successful

acquisition of the target lexis (Boers & Lindstromberg, 2009; Celce-Murcia & Olshtain, 2000; Fan, 2009; Hunt & Beglar, 2002, Lewis, 1997; Meunier, 2012).

Sentence builders. The fifth worksheet included 13 formulaic sequences that function as sentence builders. The function and usage of each formulaic sequence was briefly explained and illustrated in sample sentences that are related to different topics. The rationale for presenting a list of formulaic sequences along with an explicit description of their function and usage in academic writing is pertinent to the nature of this language phenomenon, which is partially transparent or totally opaque for many L2 learners (Boers & Lindstromberg, 2009; Willis, 2003). Moreover, illustrating the use of formulaic sequences in examples on different topics was an attempt to provide the EAP students with semi-concordance lines that highlighted both the semantic and syntactic functions of the target formulaic sequences (Jones & Haywood, 2004; Lewis, 2008; O'Keefe et al., 2007).

Sentence builders practice. The sixth worksheet was developed in order to have the students practice the use of the formulaic sequences presented in the fifth worksheet. To this end, they were instructed to use 11 formulaic sequences along with sentence clues in order to compose meaningful sentences in groups of four. This practice worksheet, first and foremost, provided students with sentence-builder practice which, as suggested by Nattinger and DeCarrico (1992) and Willis (2003), is likely to promote academic writing skills. Moreover, presenting clues related to different topics to practice highly productive sentence builders could raise L2 learners' awareness of the creative and innovative use of these formulaic sequences (Nattinger & DeCarrico, 1992). Finally, having L2 learners use formulaic sequences that were previously introduced to construct

meaningful sentences in a communicative activity provides the repeated exposure and deep processing necessary for their acquisition (Fan, 2009; Lewis, 1997; Woolard, 2000).

Cause/effect & compare/contrast. Nine formulaic sequences that can be used to either compare/contrast or reveal causative relations between discourse elements in academic writing, along with a brief explanation and sample sentences, were included in the seventh worksheet. The rationale behind designing this worksheet is identical to that of the fifth worksheet. In other words, because of the partial transparency and sometimes total opacity of formulaic sequences (Boers & Lindstromberg, 2009), the provided explanation was intended to familiarize the students with the semantic functions and syntactic structures of the target formulaic sequences in academic writing. Moreover, classifying formulaic sequences into two distinct discursive functions and contextualizing them were based on Conzett's (2000) and Jones and Haywood's (2004) suggestions that presenting semantically categorized formulaic sequences in context (e.g. concordance lines) is likely to familiarize L2 learners with the syntactic and semantic features of this language phenomenon and, hence, promote its accurate utilization in academic writing.

Mingle and sentence writing activities. The nine formulaic sequences presented in the seventh worksheet were integrated into two different activities, one communicative and one individual, in the eighth worksheet. In the first activity, seven sentences were constructed using eight different formulaic sequences. Each student was given one sentence and was instructed to memorize his/her sentence, orally share it with his/her classmates, listen to his/her classmates' sentences, and, then, write them down. All the students had to repeat the same process until they compiled a list of seven different sentences, including theirs. In the second activity, the students were instructed to work

independently and write seven meaningful and grammatically accurate sentences using seven different formulaic sequences. Both activities were adapted from different sources because of their pedagogical utility.

The mingle activity was adapted from both the memorization technique recommended by Conzett (2000), Nation (2001, 2002), and Nattinger and DeCarrico (1992) and the mingle jigsaw suggested by Wood (2002, 2010b). Memorization—which is a technique in itself and an essential part of the mingle activity—together with the oral sharing of information, can promote deep processing of formulaic sequences, may increase the chances of committing formulaic sequences to L2 learners' long-term memory, and might ultimately lead to more accurate language production (Nattinger & DeCarrico, 1992; Nesselhauf, 2003; Wood, 2010b). Having L2 learners put formulaic sequences in sentences that are both syntactically and semantically accurate was, in turn, adapted from Lewis (1997) and Nation (2001, 2002). To them, using lists of formulaic sequences actively in the language classroom and instructing L2 learners to put lexical items in meaningful sentences are two techniques that may elevate the acquisition of formulaic sequences and augment L2 learners' writing proficiency.

Discourse organizers. The ninth worksheet was very similar to the fifth and seventh worksheets in that it included a number of semantically related formulaic sequences, a total of 17, along with brief descriptions and examples that highlighted their discursive functions and syntactic structure in academic writing. Similar to the objectives behind designing the fifth and seventh worksheets, creating a semantic list of formulaic sequences, along with a brief explanation and some sample sentences, was intended to raise the EAP students' awareness of their use and function, which may remain otherwise

unnoticed due to their complex nature (Boers & Lindstromberg, 2009). Moreover, presenting a list of formulaic sequences as a prewriting activity is likely to elevate L2 learners' writing proficiency as they can master discourse elements that help them write coherently (Conzett, 2000; Cortes, 2004). Furthermore, the ninth worksheet reintroduced some of the formulaic sequences that were present in the first and second worksheets in order to elaborate on their use and function, which can be seen as another instance of repeated exposure to the target formulaic sequences that may promote the chances of their acquisition (Jones & Haywood, 2004; Lewis, 1997; Woolard, 2000).

Discourse organizers practice. The last worksheet was designed to offer some independent practice for the discourse organizers that were presented in the ninth worksheet. To this end, the tenth worksheet included ten different sentences, each of which was followed by a formulaic sequence between brackets. Each student had to rewrite the ten sentences using the formulaic sequences between brackets. This activity was adapted from one of the dividing-up texts techniques suggested in Nation (1991, 2001) and Conzett's (2000) discourse organizers activity. Nation (1991, 2001) explains that a delayed copying activity—which involves rewriting a text through phrase for phrase rather than word for word copying—may improve L2 learners' academic writing skills because the copying process may enhance the deep processing of chunks of language, which is likely to result in their successful acquisition.

Conzett (2000), in turn, explains that engaging L2 learners in activities that promote the memorization of formulaic sequences, especially discourse organizers, may help L2 learners produce coherent and cohesive written texts. In this respect, the tenth worksheet was designed to promote, through the copying technique suggested by Nation

(1991, 2001), the internalization of a number of formulaic sequences that mainly function as discourse organizers to help them, at later stages, write both coherently and cohesively, as suggested by Conzett (2000). It also provided the students with multiple exposures to the target formulaic sequences, which can also augment the chances for their acquisition and internalization into the students' linguistic repertoire (Jones & Haywood, 2004; Lewis, 1997; Woolard, 2000).

While the pedagogical validity of the first set of instruments was enhanced, at least partially, by designing it based on the recommendations of experts in the field of SLA, the pedagogical utility of this set could be ascribed to its content. The content of the first set of instruments was in line with the learning objectives of the program in which the study took place. As mentioned earlier, part of the writing assessment guidelines in the selected EAP program focuses on the upper-intermediate and advanced students' ability to accurately utilize, at least to some extent, one class of formulaic sequences in their writing (i.e., collocations). The content of the first set of instruments, therefore, helped not only implement one of the three target pedagogical interventions but also address the needs of the students who were registered in the groups that were assigned focused instruction of formulaic sequences.

Instruments for the experimental saliency groups. The second set of pedagogical instruments was composed of 10 worksheets that were identical in content (i.e., metalinguistic explanations as well as examples) and order to the instruments designed for the experimental focused instruction groups, and so were the pedagogical utility and validity of the content of the second set of pedagogical instruments. However, the different nature of the second pedagogical intervention (i.e., salient input) required the

use of some typographic techniques to make the 72 target formulaic sequences salient in the second set of instruments (see Appendix C, worksheets 11-20).

The use of the same pedagogical instruments that were alike in content and order but different in format to deliver two different types of pedagogical interventions to different experimental groups (i.e., focused instruction and saliency) was intended to control for an extraneous variable that may have invalidated the results of this research study, that is, the effects of different types of pedagogical materials on learning gains. By only adjusting the format of the worksheets used with the focused instruction groups, any differences in the performance of the different experimental groups could be attributed to the treatment effects (i.e., focused instruction vs. saliency) rather than another possibly intervening and unrelated variable (i.e., the effects of different types of activities).

The target formulaic sequences were made salient in the second set of instruments by using some of the *input enhancement* techniques suggested in the literature. *Input enhancement*, also termed *input-salience-creation*, *textual enhancement*, and *textual enhancement of input*, is a technique utilized to deliberately make opaque and/or complex linguistic aspects perceptually salient in input (Han, Park, & Combs, 2008; Izumi, 2002; Simard, 2009; Smith, 1991, 1993). Such a technique, which is externally induced by altering '*the physical appearance*' of certain linguistic aspects, aims to implicitly direct L2 learners' attention to the target linguistic aspects to promote their deep processing and acquisition (Izumi, 2002; Rutherford & Smith, 1985; Simard, 2009; Smith, 1991, 1993).

Visual *input enhancement* of less salient linguistic aspects in written texts can be induced in many ways: coloring, boldfacing, italicizing, capitalizing, shadowing and/or underlining the target linguistic aspects; adding a typographical symbol (i.e., an asterisk)

to a word, a phrase, or a sentence; using different fonts or altering font sizes; providing examples with or without metalinguistic explanations; and explicitly referring to the importance of certain linguistic aspects (Izumi, 2002; Meunier, 2012; Simard, 2009; Smith, 1991, 1993). Each of these visual enhancement techniques can be either manipulated individually or in a combined form, which is believed to be more effective in flagging the target linguistic aspects (Han et al., 2008; Simard, 2009). Visual enhancement cues have also been used and/or recommended as a technique to make formulaic sequences salient in input (e.g. Bishop, 2004a, 2004b; Boers & Lindstromberg, 2012; Jones & Haywood, 2004; Meunier, 2012).

Of the proposed visual enhancement techniques, boldfacing, highlighting, italicizing, capitalizing, underlining, and explicit reference were used either one at a time or in a combined form to make the 72 target formulaic sequences salient in input. Boldfacing was utilized to make the formulaic sequences that were present in the first and the second worksheets salient in input. Color-coding was used in the third worksheet to direct the students' attention to formulaic sequences and promote their recognition of the different grammatical categories to which they belong. Underlining, capitalizing, and italicizing were also used independently to respectively flag the formulaic sequences that were included in the sixth, eighth and tenth worksheets. The formulaic sequences in the four remaining worksheets were made salient by using combined methods; that is, while the formulaic sequences in the fifth, seventh, and ninth worksheets were made salient by italicizing and boldfacing them, the ones in the fourth worksheet were presented in a separate box and their importance was orally referred to during the training period.

In addition to the visual enhancement techniques that were applied to the second set of instruments, six sets of discussion questions (see Table C1) accompanied the presentation of the worksheets that included explanatory information and were explicitly taught to the other groups, namely, worksheets one, two, three, five, seven, and nine. These discussion questions had both equalization and amplification purposes. They were, first and foremost, designed to equalize the time spent on pedagogical intervention across all groups. The students in the groups that were assigned salient formulaic sequences were required to check the content of each of the above-mentioned worksheets and answer the related discussion questions in groups of four. In so doing, the time spent on explicit instruction in the experimental focused instruction and control groups was compensated for by the implementation of discussion questions.

The discussion questions also served as an amplifying technique that accompanied the visual *input enhancement* cues. According to Han et al. (2008) and Smith (1991), L2 learners may not notice textually induced salient items in input, or they may notice these items superficially without any subsequent effects on their learning and/or acquisition of these aspects. To increase the chances for noticing and deep processing, the discussion questions focused on the target formulaic sequences and their functions in academic writing, explicitly directed the EAP students' attention to the form and function of the 72 target formulaic sequences, and helped reduce the chances of disregarding them in input.

Instruments for the control groups. The third set of pedagogical instruments was designed for the control groups and was used to compensate for the time spent on presenting formulaic sequences to the experimental focused instruction and saliency

groups. This set was comparable in its quantity, form, and pedagogical objectives to the ones developed for the experimental groups, that is, being composed of 10 worksheets that included explanatory, practice, and combined explanatory and practice worksheets and providing the EAP students with both communicative and independent practice. However, the third set differed in its content from the first two sets in that it centered on four linguistic components: topic and concluding sentences, sentence types, articles, and affixes (see Appendix C, worksheets 21-30).

The selection of these linguistic components was partly influenced by some of the writing assessment guidelines of the selected EAP program, which explicitly focused on the implementation of paragraph components, the variety of sentence types, and the accuracy of syntax at the word and sentence levels. Focusing on these components was, therefore, intended to increase the pedagogical utility of this set of instruments by introducing or reinforcing some of the linguistic aspects that the EAP students were expected to master. It was also partly informed by the literature on teaching L2 academic writing skills. Different composition approaches focus on the importance of addressing key textual elements (i.e., topic and concluding sentences, supporting sentences, etc.) and linguistic aspects (e.g. syntax, lexis, etc.) in the language classroom whether directly (i.e., the traditional approach and the current-traditional rhetoric) or indirectly (i.e., the process-based, the genre-based, and the content-based approaches) in order to promote L2 learners' writing proficiency (Hyland, 2007; Mitchell, 2004; Raimes, 1991; Seow, 2002; Shih, 1986; Silva, 1990).

Focusing on articles, sentence types, and word forms was also informed by the literature on error analysis of L2 written texts, which lists tenses, articles, subject-verb

agreement, sentence structure, punctuation, and lexical choice as universally common errors in the writing production of L2 learners—including advanced learners—despite being engaged in different types of communicative activities and being exposed to rich input (Butler, 2002; Connors & Lunsford, 1988; Ellis, 2005c; Ferris, 2002; Ferris & Roberts, 2001; Harper & de Jong, 2004; Laufer, 2006; Richards, 2002; Swan, 2002). The recurrence of these syntactic errors has been mainly ascribed to the complexity of these linguistic aspects, their low communicative functions, their limited perceptual saliency, or L1 transfer (Butler, 2002; Richards, 2002). Unless addressed in an explicit and systematic manner, L2 learners may continue to struggle with these linguistic aspects despite being exposed to rich input and using the second language for many years (Butler, 2002; Ellis, 2005a; Harper & de Jeong, 2004).

Writing topic and concluding sentences. The first and the second worksheets included both explanatory information and writing practice. As for the explanatory information, the first worksheet introduced the function and constituents of a topic sentence in academic writing along with two sample topic sentences. It also referred to the way in which a topic sentence should guide the selection of all subsequent supporting sentences and supporting details. The second worksheet, in turn, explicated the function and component parts of a concluding sentence in both academic paragraphs and essays and presented two different examples. The practice component of the first and second worksheets included three writing prompts for which the EAP students were required to write topic and concluding sentences in groups of four following the examples provided in each worksheet.

Addressing topic and concluding sentences was adapted from both the traditional approach and the current-traditional rhetoric, which stress the importance of mastering the constituents of paragraphs and five-paragraph essays (Mitchell, 2004; Raimes, 1991; Silva, 1990). Teachers following these approaches typically view writing as a process of '*guided imitation*'; they break written texts down into their constituent parts, address these parts explicitly in the language classroom, and then provide L2 learners with both controlled and uncontrolled practice in order to help them master these constituents and ultimately promote their writing proficiency (Badger & White, 2000; Curry & Lillis, 2003; Silva, 1990). The explicit instruction and controlled practice of topic and concluding sentences were, therefore, intended to help EAP students in the control groups master these constituent parts and, hence, improve their academic writing skills.

Sentence types. Four of the worksheets for the control groups focused on clause and sentence types and were informed by information and sample practice activities available online (e.g. Purdue Owl) and ESL/EAP textbooks (e.g. *Longman Academic Series 4*, *Pathways 3: Listening,* and *Speaking and Critical Thinking*, *Fundamentals of English Grammar*). The third worksheet first tackled the differences between dependent and independent clauses with examples and then introduced different types of simple sentences, that is, simple sentences with single vs. compound subjects and/or verbs. To practice the content of the third worksheet, the EAP students were required to work in groups and complete three writing activities, which constituted the content of the fourth worksheet. The students were first required to read six sentences and underline the independent clause in each one. They, in the second activity, had to decide whether five

strings of words were complete sentences or fragments. They were then instructed to compose five different types of simple sentences with their group members.

The fifth worksheet explained different types of compound and complex sentences. Similar to the third worksheet, the fifth worksheet listed some sample sentences and was followed by practice activities in the sixth worksheet. The first activity in the sixth worksheet required the students to work in groups of four and label six different sentences as simple, compound or complex. In the second activity, the students were required to negotiate chunks of language with their group members and try to write different types of semantically and syntactically accurate sentences using these chunks. Another group practice of sentence types was created in the sixth worksheet. Instead of using chunks of language, the students were required to use a list of words to construct either a compound or a complex sentence.

As for the pedagogical utility of the third and fifth worksheets, providing explanatory information on the two major types of clauses as well as the different types of English sentences had a dual purpose. Since fragments, comma splices, and run-on sentences are among the major syntactic errors that are recurrent in L2 learners' written texts (Connors & Lunsford, 1988; Ferris, 2002; Ferris & Roberts, 2001), providing explanatory information was an attempt to eliminate or at least reduce the recurrence of such errors in the EAP students' writing. The third and fifth worksheets were also designed to raise the students' awareness of the form and structure of different types of sentences and to reinforce the importance of implementing various types of sentences in their writing production, which is considered a marker writing proficiency (Celce-Murcia & Olshtain, 2000; Ferris, 2002; Llach, 2011; Reppen, 2002).

The practice worksheets, in turn, were designed to provide the students with two different types of practice that were likely to help them master the production of different sentence types. As mentioned above, sentence structure is one of the challenging linguistic aspects to many L2 learners. Different types of clauses are introduced at different stages of language learning (i.e., lower intermediate or intermediate), yet they are mastered at later stages (Ellis, 2004). Having L2 learners assign linguistic labels to different chunks of language are common activities in writing textbooks (e.g. *Longman Academic Series 4*) and online practice websites (e.g. ProProfs Quiz Maker). Such practice is, as Ellis (2004) suggests, '*a test of metalanguage*' that plays a role in promoting L2 learners' syntactic accuracy. Requiring the students to use clues to compose meaningful sentences aimed at controlling their output by pressuring them into drawing on advanced linguistic forms, which is likely to expand their linguistic repertoire and ultimately improve their linguistic competence (Ellis, 2002, 2005c; Richards, 2002).

Articles. The seventh and eighth worksheets concentrated on the use of definite and indefinite articles with different types of nouns and were informed by information available on online websites (e.g. Purdue Owl) and ESL/EAP textbooks (e.g. *Fundamentals of English Grammar*, *English Grammar in Use: Intermediate*, and *Longman Academic Series 5*). The seventh worksheet consisted of a description of English nouns (i.e., singular, plural, and uncountable). It also focused on the use of the definite article *the* with plural countable and uncountable nouns and the use of *a*, *an*, and *the* with singular nouns. The eighth worksheet was designed to have the students practice in groups the use of definite and indefinite articles in three different activities. While the first activity required the students to fill in the gaps with *a*, *an*, or *the*, they had, in the

third activity, to decide which nouns should be used with *a*, *an*, *the*, or zero article. Editing five sentences for article errors was the second activity in the practice worksheet.

Similar to the rationale for designing the explanatory worksheets on sentence types, providing explanatory information on the use of articles with different types of nouns was intended to raise the students' awareness of the proper use of English articles. English articles are among the major universally challenging linguistic aspects for many L2 learners, including advanced learners; they are introduced at the very early stages of language learning, yet the majority of L2 learners do not master them until they achieve an advanced proficiency level (Butler, 2002; Ellis, 2005c). Despite their high frequency in input, the remarkable complexity of English articles—which may be attributed to the lack of semantic properties, the effects of L1 transfer, or the overgeneralization of L2 rules—entails providing systematic and structured explanation of their usage to promote L2 learners' syntactic accuracy (Butler, 2002; Ferris, 2002; Young, 1996).

The different types of practice activities in the eighth worksheet had different pedagogical utilities. Having L2 learners complete sentences by using the definite/indefinite articles or Ø for zero articles may help them properly use this complex linguistic aspect in their language production (Butler, 2002; Celce-Murcia & Olshtain, 2000). Splitting this practice into two successive component parts that first focused on the use of articles and then the use of articles and zero articles, on the other hand, was informed by Celce-Murcia and Olshtain's (2000) suggestion that articles practice should focus on the use of definite and indefinite articles in sentences prior to being extended to cases with zero article. Editing for form, in turn, is a major step of the writing process that many L2 learners need training at in order to improve their linguistic accuracy

(Ferris, 2002). It is also one form of *Grammatically Judgment Tests* which involves semantic and syntactic processing and requires L2 learners to access their linguistic knowledge and, hence, helps measure their implicit and/or explicit knowledge of a target linguistic form (Ellis, 2004, 2005c).

Affixation. The last two worksheets focused on word-derivation rules and were informed by information available on online websites (e.g. ThoughtCo. & Lexical Tools) as well as ESL/EAP textbooks (e.g. *English Grammar in Use: Intermediate* & *Longman Academic Series 5*). The ninth worksheet listed a number of prefixes that can be used to derive opposite or new words along with examples. It also included a number of suffixes that can be used to derive nouns from verbs, adverbs and verbs from adjectives, and adjectives from nouns or verbs. The tenth worksheet included two practice activities that focused on the use of prefixes and suffixes and were completed independently by each student. Teaching word-formation rules is crucial for improving academic writing skills as it equips L2 learners with the rules necessary for deriving new words, promotes their linguistic accuracy, and helps them achieve native-like proficiency (Celce-Murcia & Olshtain, 2000). Providing L2 learners with different types of word-focused practice also promotes the acquisition process of L2 vocabulary (Laufer, 2005, 2006).

Prior to concluding this section, two points that relate to the design of the three sets of pedagogical instruments are worth highlighting. The focus of the first two worksheets in each set of instruments was made identical (i.e., topic and concluding sentences) in order to provide the students in different groups with equal opportunities to practice the same language aspects when possible. Matching the content of the three sets was also an attempt to promote the validity of the pedagogical instruments. In other

words, teaching two paragraph elements to the experimental groups only might improve their writing production due to practicing these component parts rather than learning different types of formulaic sequences to construct topic and concluding sentences. Such an intervening variable was controlled by teaching the constituent parts and functions of both topic and concluding sentences to the experimental and control groups.

The method of presenting information and the types of practice were also made similar as much as possible. Despite their different focus, the three sets of pedagogical instruments included explanatory information that was followed by decontextualized practice activities which were completed in groups (five worksheets) or independently (one worksheet). Developing different sets that were comparable in form and delivery was intended to control for the effects of presenting different amounts of metalinguistic information, integrating different types of activities (i.e., contextualized vs. decontextualized), and implementing different types of practice (communicative vs. individual) on learning gains. By integrating the same amount and type of explanatory information, activities, and practice, the effects of explanatory information, type of activities, and type of practice on learning gains were controlled.

Adapted Pedagogical Approaches

Since this research study aimed at investigating the effectiveness of two approaches to presenting formulaic sequences in improving EAP students' academic writing skills when compared to EAP students who were taught writing and structural materials, major concepts proposed in SLA research and pedagogy were amalgamated to inform delivering the three different sets of pedagogical instruments. The delivery of each set was mainly based on concepts from several L2 instructional approaches, namely

form-focused instruction, *consciousness-raising*, *planned lexical instruction*, and *the lexical approach*.

Form-focused instruction. *Form-focused instruction* refers to an instructional approach that places, at its core, addressing the formal properties of complex linguistic aspects (Collins, 2012; Dörnyei, 2013; Ellis, 2005b). It can be implemented in the language classroom either intensively by designing exercises and activities that focus on preselected linguistic forms or extensively by providing rich input and addressing problematic linguistic forms as they arise during communicative tasks (Collins, 2012; Ellis, 2005b; Ellis, Basturkmen, & Loewen, 2002). The intensive versus extensive dichotomy of *form-focused instruction* is associated with two distinct terms in the literature, that is, *focus on forms* and *focus on form***,** respectively (Collins, 2012; Ellis, 2005b, 2015; Ellis et al., 2002).

Integrating *form-focused instruction* in the language classroom involves four different pedagogical strategies (Dörnyei, 2013; Ellis, 1998). The first pedagogical strategy of *form-focused instruction* is the *input-based option* which aims to raise L2 learners' awareness of a target linguistic aspect through *input flooding* (i.e., focusing on input that is dense with the target linguistic form), *enhanced input* (i.e., artificially manipulating input by using typographical textual enhancement strategies), or *structured input*, that is, having L2 learners process, analyze, and demonstrate their understanding of a target linguistic aspect in a text (Dörnyei, 2013; Ellis, 1998). When the input is geared at a specific linguistic aspect, it provides L2 learners with positive linguistic evidence, that is, information about what is grammatically accurate and acceptable in the target language (Ellis, 1998). *Direct and indirect instruction* constitutes the second pedagogical

strategy of *form-focused instruction* (Dörnyei, 2013; Ellis, 1998). While *direct instruction* provides L2 learners with oral or written metalinguistic explanation of the target linguistic form and may or may not be followed by practice activities, *indirect instruction* enables L2 learners to construct linguistic rules through *consciousness-raising* or analytical tasks (Dörnyei, 2013; Ellis, 1998).

Production practice, the third pedagogical strategy of *form-focused instruction*, focuses on learners' output and is also referred to as *production options* (Dörnyei, 2013; Ellis, 1998). It pushes L2 learners to practice the use of the target linguistic aspects either in highly controlled and structured exercises that involve '*text manipulation*' (e.g. a fill-in-the-blank exercise) or in '*text creation*' activities, which are less structured and require L2 learners to create their own stretches of discourse while using the target linguist form (Dörnyei, 2013; Ellis, 1998). *Negative linguistic evidence* is the last pedagogical strategy of *form-focused instruction* that is typically implemented at the *production-practice stage* (Dörnyei, 2013; Ellis, 1998). It provides L2 learners with negative linguistic evidence by signaling out grammatically inaccurate and unacceptable constructions to reduce the frequency of errors and promote linguistic accuracy (Dörnyei, 2013; Ellis, 1998). This form of negative linguistic evidence (i.e., corrective feedback) can take a variety of forms including recasts, repetition, elicitation, explicit correction, and metalinguistic feedback (Dörnyei, 2013; Ellis, 1998).

Consciousness-Raising. As a term, *consciousness-raising* refers to a number of pedagogical techniques used to deliberately draw L2 learners' attention to the formal characteristics of a target linguistic aspect in order to facilitate its acquisition (Rutherford & Smith, 1985; Smith, 1981, 1991, 1993). These techniques range in their degree of

'*explicitness*' (i.e., gestures, typographical cues, and metalinguistic explanation) and '*elaboration*,' that is, how long the process of raising awareness lasts and how recurrent it is (Smith, 1981, 1991). Among consciousness-raising techniques are utilizing textual enhancement strategies (e.g. highlighting, underlining, and italicizing), artificially creating input replete with the target linguistic aspect, and providing positive and negative linguistic evidence (Rutherford & Smith, 1985; Smith, 1991, 1993). The latter technique can be provided through instruction (i.e., elaborating on correct and incorrect linguistic forms) or written corrective feedback by making correct stretches of discourse more salient or flagging inaccurate constructions with or without metalinguistic explanation (Smith, 1991, 1993).

As a method of instruction, *consciousness-raising* aims at raising L2 learners' explicit and conscious knowledge of complex linguistic aspects (Ellis, 2002, 2015). In its essence, *consciousness-raising* is similar to *practice*, a traditional approach in which a complex linguistic aspect is isolated, explicitly addressed at a presentation stage, repeatedly practiced at a controlled practice stage, and then produced in an uncontrolled manner at a production stage (Ellis, 2002). Similar to *practice*, *consciousness-raising* instruction involves isolating and addressing a complex linguistic aspect in the language classroom (Ellis, 2002, 2015). However, it differs from *practice* in that the target linguistic aspect is addressed inductively by providing L2 learners with rich data, sometimes with metalinguistic explanation, from which they construct a linguistic rule (Ellis, 2002, 2015). Another difference between *practice* and *consciousness-raising* is that the latter does not necessarily involve repeated practice, and explanation is only provided in the form of clarification after the presentation stage (Ellis, 2002, 2015).

Planned lexical instruction (PLI). *Planned lexical instruction* is an approach proposed by Laufer (2005) to overcome the shortcomings of a common assumption that L2 vocabulary is naturally and incidentally acquired through exposure alone. Laufer (2005) argues that L2 learners may not be able to build a rich repertoire of vocabulary based on extensive exposure during communicative and reading activities because they may not notice new words, or they may notice them without any consequences on learning. According to the PLI, the main source of L2 vocabulary acquisition is not incidental learning but rather such an instructional approach that focuses on lexical items as PLI (Laufer, 2005, 2006). *Planned lexical instruction* is compatible with *form-focused instruction* in that both approaches attend to target linguistic aspects inductively or deductively (Laufer, 2005). However, they have different linguistic aspects at their core, with the former focusing on single and/or multiword lexical items and the latter centering on grammar (Laufer, 2005, 2006).

Teachers who opt for *planned lexical instruction* provide their students with the exposure, information, and practice necessary for vocabulary acquisition through decontextualized vocabulary-based activities as in the case of *focus on forms* or communicative tasks that are typical of *focus on form* (Laufer, 2005). This instructional approach can be preplanned (i.e., using a preselected list of lexical items) or incidental, that is, focusing on new lexis during a communicative activity (Laufer, 2005). Whether implemented inductively or deductively, PLI helps overcome the shortcoming of L2 incidental vocabulary acquisition by assigning an active role to instructors, that is, exposing L2 learners to rich input repeatedly and confirming their comprehension and acquisition of the target lexis (Laufer, 2005). In brief, “the value of PLI for instructed

learning context lies in ensuring noticing, providing correct lexical information, repeated exposures to words, and ample opportunities for creating and expanding knowledge through a variety of word focused activities" (Laufer, 2005, p. 324).

The Lexical Approach. As explicated in the second chapter, *the Lexical Approach* places lexis at the heart of L2 instruction and advocates the importance of allocating a considerable amount of classroom time to multiword lexical items in order to ensure building a large repertoire of lexis that is essential for proficient language production (Lewis, 1997, 2002, 2008). *The Lexical Approach* is compatible with CLT in that it mainly focuses on meaning-based communicative activities that target L2 learners' mental lexicon (Lewis, 2008). Unlike the latter, however, *the Lexical Approach* does not place L2 learners' language production at the center of the designed activities. Instead, learning is promoted through different forms of pedagogical practice that may incrementally convert rich input into intake (Lewis, 2008).

Participation, the first practice technique proposed in *the Lexical Approach*, encompasses various forms that do not exclusively focus on L2 learners' production (Lewis, 2008). Learners' participation can be achieved through activities that focus on speaking, listening, reading, noticing, and/or reflection; practice can also be achieved through repeatedly exposing L2 learners to the target lexis, promoting their noticing of the target lexical items (e.g. through repetition), or raising their conscious awareness of their forms and functions, for example, by having L2 learners identify the commonalities and differences between lexical items (Lewis, 2008). These forms of practice, as Lewis (2008) explicates, may increase the likelihood of converting input into intake. Integrating negative linguistic evidence in teachers' feedback by highlighting an error in a lexical

item prior to correcting it is the last form of practice in *the Lexical Approach* and is likely to promote lexical acquisition and accuracy (Lewis, 2008).

Despite the major differences in their core concepts, the above-reviewed approaches share in common two major aspects: the importance of focusing on complex linguistic aspects—whether syntactic or lexical—and the efficacy of dividing instruction into sequential presentation and practice stages. They were, therefore, adapted to deliver the three different types of pedagogical intervention, that is, focused instruction of formulaic sequences, salient formulaic sequences, and writing materials.

Focused instruction of formulaic sequences. Delivering the target pedagogical intervention in the experimental focused instruction groups was informed by the concepts proposed in the four instructional approaches reviewed above. Prior to referring to the information presented in each explanatory worksheet, the form and function of the listed formulaic sequences were explicitly addressed on the board with examples and metalinguistic explanation that highlighted their accurate use. The presentation stage was followed by timed practice, in groups or independently, using the related practice worksheet. After completing each timed practice, all the answers were checked as a class. Incorrect answers were explicitly pointed out and corrected by other students or the researcher, and further linguistic clarification was provided when needed. The same process was repeated till all the worksheets were covered.

Explicitly teaching the target formulaic sequences to the experimental focused instruction groups was adapted from *planned lexical instruction* (i.e., deductively teaching a preselected list of lexical items), *form-focused instruction* (i.e., input flooding and explicit instruction with metalinguistic explanation and positive linguistic evidence),

and *consciousness-raising* techniques (positive linguistic evidence through metalinguistic information). Having the students practice the use of the target formulaic sequences in structured and semi-structured worksheets was informed by the types of practice proposed in *planned lexical instruction* (i.e., decontextualized vocabulary practice) and *form-focused instruction* (i.e., *text creation* and *manipulation*). Explicitly correcting errors and providing further metalinguistic clarification as needed, in turn, was based on the negative linguistic evidence proposed in *form-focused instruction*, *the lexical approach*, and *consciousness-raising* techniques and the follow-up metalinguistic clarification recommended in *form-focused instruction* and *consciousness-raising*.

Salient formulaic sequences. The delivery of the second target pedagogical intervention was based on integrating major concepts from the four instructional approaches reviewed above. The students in the experimental saliency groups were asked to negotiate the content of each explanatory worksheet in groups and answer some comprehension questions. They, then, practiced the use of the target formulaic sequences in timed-group or timed-independent activities. Their answers were checked as a class. The inaccurate answers were explicitly pointed out, and more answers were elicited from students till the correct answer was articulated by one of the students or eventually by the researcher. The same procedure was repeated for all the worksheets.

Having the students negotiate the content of materials was mainly informed by the input stage proposed in *form-focused instruction* (i.e., providing L2 learners with flooded-, enhanced-, and structured-input) and the inductive instructional approach proposed in *planned lexical instruction*, *the Lexical Approach*, and *consciousness-raising*, that is, focusing on preselected vocabulary items. Practicing the use of formulaic

sequences in decontextualized activities was mainly based on some of the practice activities proposed in *planned lexical instruction* and *form-focused instruction*. The sample sentences that the students had to check in the explanatory worksheets as well as the explicit identification of inaccurate answers during the practice stage provided positive and negative linguistic evidence respectively, which are proposed in *form-focused instruction*, *consciousness-raising* techniques, and *the Lexical Approach*. Collecting more answers from the students till the correct answer was articulated by one of the students, in turn, was one of the negative evidence techniques proposed in *form-focused instruction*.

Writing materials. The pedagogical techniques followed to deliver the third target pedagogical intervention in the control groups were identical to those implemented in the experimental focused instruction groups; however, since the third set of materials focused on linguistic aspects other than formulaic sequences, it was mainly based on *form-focused instruction* and *consciousness-raising* techniques. Each of the four major concepts selected for the control group was first explicitly addressed on the board using metalinguistic explanation and sample sentences and then referred to in the explanatory worksheets. This presentation stage was followed by timed-group and timed-independent practice stages. After completing each timed-activity, the students' answers were checked as a class, inaccurate responses were explicitly signaled out and corrected by other students or the researcher, and further metalinguistic explanation was provided when needed.

The explicit metalinguistic explanation of each linguistic aspect was adapted from the deductive instructional approach proposed in *form-focused instruction* and the

positive linguistic evidence through metalinguistic explanation suggested as one of the *consciousness-raising* techniques. Listing sample sentences during the presentation stage on the board and in the worksheets was based on the input flooding technique suggested in these approaches to provide L2 learners with positive linguistic evidence. Explicitly pointing out errors prior to correcting them at the practice stage, in turn, provided the students with negative linguistic evidence following the explicit correction technique that is proposed in the negative evidence stage of both *form-focused instruction* and *consciousness-raising* techniques.

Procedure and Data Collection

The study was divided into six stages during which different types of pedagogical intervention, multiple rounds of data collection, and evaluation of textual data took place.

Stage 1: Data collection (pretest). Before administering any pedagogical intervention, the upper-intermediate and advanced EAP students in all groups completed two timed writing tasks that functioned as the pretest for this research study. Each student was randomly assigned one of the three designed sets of writing prompts. The random assignment of the three sets was ensured by collating them in numerical order prior to visiting each class group (i.e., prompt 1→ prompt 2→ prompt 3) and sequentially distributing them to students based on their seating arrangement. The completion of the pretest was precisely timed in all groups; the students were given 20 minutes to complete the summary task and 40 minutes to respond to the essay prompt. At this stage, 131 written texts were collected: 65 graphical summaries and 66 essay responses. Collecting pretests from all groups was an essential stage as it helped (a) identify individual differences between groups prior to delivering any pedagogical intervention and (b)

reveal any statistically significant differences in the participants' performance within and across groups after receiving one of the three target treatments (Dörnyei, 2007; Spada, 2005).

Stage 2: Pedagogical intervention. After collecting the pretest, the three randomly assigned pedagogical interventions were delivered by the researcher to ensure that they were implemented as planned. The researcher delivered the target pedagogical intervention to each group because part of this research study focused on the pedagogical utility of different instructional approaches to formulaic sequences. According to Dörnyei (2007) and Spada (2005), one of the major challenges of classroom research is making certain that a treatment is delivered as intended. This can be notoriously difficult if researchers involve the class instructors in delivering a target treatment since (a) different instructors have distinct teaching styles and beliefs that may, intentionally or unintentionally, intervene in the delivery of an intended pedagogical treatment and (b) researchers may not be granted access to a classroom to track the things that take place during a pedagogical intervention (Dörnyei, 2007; Spada, 2005).

The three target pedagogical interventions were delivered between the third and tenth week of Summer 2016 and Fall 2016 terms, depending on the weeks devoted by the groups' instructors for the researcher's visits. Each target pedagogical intervention in each group lasted for five hours, was based on integrated instructional approaches (see the previous section), and was delivered over two, three, or four weeks, depending on the time allocated for the researcher by the groups' instructors (i.e., between one and three hours over two or more sessions). The five intervention hours, which took place over two or more weeks, were divided into a presentation stage—during which the explanatory

materials were addressed either inductively or deductively—and a practice stage at which the remaining practice materials were completed by all the EAP students in the form of group or independent practice.

The researcher, at this stage, also briefly tackled English tenses and subject-verb agreement in the seven groups using visual illustrations. Although it was not part of the research design, the errors that both participants and non-participants made in their pretest and during the practice activities, the clarification that some students requested regarding the use of tenses, and the fact that English tenses and subject-verb agreement are among the major challenging English aspects for L2 learners (Connors & Lunsford, 1988) led to the researcher's decision to address these linguistic aspects in all the groups.

Stage 3: Data collection (posttest). The second round of data collection took place immediately at the end of each pedagogical intervention in order to elicit the posttest. At this stage, each EAP student was randomly assigned a set of writing prompts different from the one he/she had completed as a pretest. The random assignment of the posttest, along with the delayed posttest, was predetermined prior to visiting each group by sequentially using the six possible combinations presented in Table 3 above and writing students' names on the writing prompts prior to visiting each class group in order to eliminate the possibility of accidently giving students a set they completed at an earlier stage of data collection. The same timing procedure was followed at this stage (20 minutes for summaries and 40 minutes for essays). At this stage of data collection, a total of 132 written texts (i.e., 66 graphical summaries and 66 essay responses) were elicited from the 67 EAP participants. These 132 written texts functioned as the posttest, which

was collected to identify any immediate learning gains and differences, both within and between different groups, which could be attributed to the training period.

Stage 4: Written corrective feedback. All the EAP students received written corrective feedback on their pretests and posttests approximately one week after the collection of the posttest following the *consciousness-raising* techniques proposed by Smith (1991, 1993). The provided written corrective feedback focused on both errors—which refer to "the deviation [that] arises as a result of lack of knowledge" (Ellis, 1994, p. 51)—and mistakes—which occur "when learners fail to perform their competence" (Ellis, 1994, p. 51). The written corrective feedback was in the form of repeated typographical cues that were sometimes accompanied with written metalinguistic explanation. In other words, all the errors and mistakes in each text were underlined and labeled (e.g. s/v, tense, missing article, etc.). When there was evidence in the text that a certain inaccuracy was an error rather than a mistake, the error was corrected once or twice for illustration purposes, and a very brief metalinguistic explanation was included at the end of the written text. The written corrective feedback also focused on the presence and appropriateness of introductory and concluding sentences/paragraphs. No feedback was, however, provided on coherence, cohesion, or text length unless the text was incomplete (i.e., being composed of a couple of sentences in the case of summaries or one or two paragraphs in the case of essays).

From a pedagogical perspective, providing the students with written corrective feedback was mainly intended to promote the students' writing proficiency and accuracy. Several researchers (e.g. Ellis, 2015; Ferris & Roberts, 2001; Gass & Mackey, 2007; Hyland, 2003a) and academic writing instructional approaches (e.g. the product-based

approach) stress the efficacy of corrective feedback in augmenting L2 learners' academic writing skills. Hyland (2003a), for example, suggests that

> (f)eedback is central to learning to write in a second language. Not only can it provide writers with a sense of audience and sensitize them to the needs of readers, but it offers an additional layer of scaffolding to extend writing skills, promote accuracy and clear ideas, and develop an understanding of written genres. (p. 207)

Other researchers (e.g. Coxhead & Byrd, 2007; Li & Schmitt, 2009; Wood, 2015) argue that corrective feedback may play a role in promoting the accurate use of formulaic sequences in language production. Providing written corrective feedback was also an attempt to meet the students' expectations in the selected program, for most teachers, if not all, to the researcher's knowledge, provide written corrective feedback on the students' practice and assessed written texts.

It is worth noting here that the written corrective feedback was handed back to the students prior to the collection of the delayed posttest because some teachers requested their students receive feedback prior to their midterms, which coincided before the third stage of data collection in most of the groups. The researcher provided all the groups with written corrective feedback before the third round of data collection because of the possible positive effects corrective feedback may have on the participants' performance. Providing all the groups with corrective feedback at the same point in time can be seen as an attempt to control for a major intervening factor (i.e., feedback) that may have significant effects on the performance of the groups that received feedback before the

third stage of data collection, which may invalidate the results (Dörnyei, 2007; Mark & Reichardt, 2009; Vogt, 2007).

Stage 5: Data collection (delayed posttest). The delayed posttest, usually administered two or more weeks after the collection of the posttest (Nation & Web, 2011), took place approximately four weeks after the collection of the posttest. Each student received the last randomly assigned set of writing prompts. The same timing procedure that was followed at the pretest and posttest stages was applied to the collection of the delayed posttest. At the final stage of data collection, a total of 119 written texts (i.e., 60 summary tasks and 59 essays) were collected from the 67 EAP participants in order to explore the participants' retained knowledge within each group and identify any possible differences across the groups. The same techniques were used to provide the students with written corrective feedback on their delayed posttest. One point worth mentioning here is that the incomplete numbers of summaries and essay responses at each data collection stage (i.e., fewer than 134 texts) were due to the absence of some participants at a data collection stage or the fact that some participants completed a single writing task (i.e., either the summary or the essay) at one stage of data collection. The three data collection stages resulted in collecting 382 written texts (i.e., 191 graphical summaries and 191 essay responses), which constituted the textual data in this study.

Stage 6: evaluation. To identify any improvement in the participants' evaluation within and/or between different groups, the textual data were blindly evaluated by the EAP judge based on the two analytical rubrics. The collected summaries and essays were typed and sequentially organized based on the prompt number (i.e., prompt 1→ prompt 2→ prompt 3) without any reference to the order in which they were collected or the

proficiency level of the student who produced them to ensure blind and objective evaluation. The EAP judge was required to assign numeric grades out of 100 following the analytical rubrics; no feedback was required. Given the tedious and time consuming nature of writing evaluation, the EAP judge was given, as honorarium, one Canadian dollar for each paragraph and two Canadian dollars for each essay she evaluated.

Quantifying Textual Data

Three different techniques were used in order to quantify the textual data and derive numerical variables with different categorical counts.

The textual data were first quantified based on *quantitative content analysis*, a deductive, systematic, and objective coding technique that is manually or electronically applied to various types of textual data to tally instances of target words, phrases or other linguistic aspects (Dörnyei, 2007; George, 2009; Jeong 2013; Neuendorf, 2017). Unlike inductive coding techniques in which codes emerge from data, *quantitative content analysis* is driven by a coding scheme that includes a list of categories identified a priori to elicit *frequency* and/or *non-frequency* counts in textual data (Dörnyei, 2007; George, 2009; Jeong 2013; Neuendorf, 2017). While *frequency* counts concern with "the number of times one or more content characteristics occur," the *non-frequency* counts focus on "the mere occurrence or nonoccurrence of attributes" (George, 2009, p. 145).

Quantitative content analysis was manually applied to the textual data using a predefined categories list that included all the tackled formulaic sequences during the presentation stage (see Table D1) in order to identify the frequency (i.e., *frequency counts*) and occurrence (i.e., *non-frequency counts*) of the target formulaic sequences in both tasks at the three stages of data collection. The predefined list of categories was first

used to count all the formulaic sequences that the 67 participants used at the three production stages including the repeated ones (i.e., frequency) and, then, to identify the distinct types of formulaic sequences (i.e., occurrence) in the textual data. All other multiword sequences that are deemed formulaic in the literature and were used by the participants in both tasks at the three production stages were disregarded because they were neither tackled during the presentation stage nor included in the predefined categories list.

One point worth noting is that although some research studies only focused on accurate formulaic sequences (e.g. Peters & Pauwels, 2015; Serrano, Stengers, & Housen, 2015), all the target formulaic sequences that the 67 participants produced, including the inaccurately constructed or used ones, were counted at this stage and, then, classified into correct and incorrect formulaic sequences. Counting both correct and incorrect formulaic sequences was informed by Nation and Webb's (2011) "sensitive tests of vocabulary knowledge" which "gives credit to partial knowledge of a word" and takes into account the nature or vocabulary acquisition as being "a cumulative process" (p. 303). In other words, vocabulary knowledge, including multiword lexis, is not an *all-or-nothing* form of knowledge (Laufer & Goldstein, 2004; Nation & Webb, 2011). It is rather a multifaceted, incremental process that includes the gradual control over the semantic, syntactic, and morphological aspects of lexical items (Laufer & Goldstein, 2004; Nation & Webb, 2011).

The data quantification phase also included manual and computer-aided counting. During the manual counting phase, all types of linguistic inaccuracies (e.g. tenses, spelling, punctuation, etc.) were manually counted and collectively referred to as errors

because exploring the type of inaccuracy (i.e., error vs. mistake) was beyond the scope of this research study. The number of words generated by the 67 EAP students in both tasks at the three production stages was also counted using the Word Count function in Microsoft Word (see Appendix D, Analyzed Texts). Counting errors and words was an attempt to address one of the major objectives of this research study, that is, measuring the EAP students' linguistic accuracy and writing proficiency. Among the most widely accepted indicators of academic writing proficiency are lexical precision, grammatical correctness, error-free phrases or clauses, and word counts, especially in timed writing tasks (Ellis, 2015; González-Bueno & Pérez, 2000; Hinkel, 2004; Ishikawa, 1995; Llach, 2011; Perkins, 1980; Polio, 1997; Schoonen et al., 2011). While acknowledging the importance of coherence, cohesion, and rhetorical skills as major indicators of writing proficiency, several studies (e.g. González-Bueno & Pérez, 2000; Lembke et al., 2003; Perkins, 1980) utilized counts of errors, words, and/or error-free clauses as determiners or, at least predictors, of writing proficiency.

The three quantifying techniques resulted in extracting 24 major variables that included frequency, occurrence, words, and errors and provided the basis for creating 42 new variables that represented either subcategories (i.e., correct formulaic sequences, wrong formulaic sequences, 3 distinct types of erroneous formulaic sequences) or percentages (correct formulaic sequences and errors) of these variables. Other variables were also developed to include the EAP judge's evaluation, split data by groups and proficiency levels, assess normality, and compute correlations (see Appendix D, List of Variables). The data quantification stage resulted in creating 100 continuous variables with an interval scale of measurement—in which there are equal and meaningful intervals

between adjacent values—and two discrete variables with a nominal level of measurement, in which values are arbitrarily assigned to categorize data (Dörnyei, 2007; Vogt, 2007). These 102 variables constituted two data sets for summaries and essays, henceforth respectively referred to as task 1 and task 2 data sets (see Tables D2 to D7).

Statistical Analyses

After manually imputing the missing scores in task 1 and task 2 data sets, several screening, descriptive, and inferential statistical tests were computed on both data sets using the Statistical Package for Social Sciences (SPSS), version 24.

Missing data: Analysis and imputation. Since the three rounds of data collection resulted in eliciting 382 written texts rather than 402 had there been no missing data, the percentage and nature of missing data in each data set were investigated prior to running any statistical analyses. While missing data is a very common problem in research studies with repeated measurements, investigating the nature and percentage of missing values at a preliminary analysis stage is not only recommended but also crucial because (a) many statistical tests with missing values assume that the missing data mechanism is *missing completely at random* (MCAR) and (b) most imputation techniques, which help eliminate biased results and increase statistical power, require that the data are MCAR (Little, 1988; Meyers et al., 2006; Schlomer, Bauman & Card, 2010; Tabachnick & Fidell, 2001, 2014).

Given the recommendations proposed in the literature, the 72 major continuous variables were statistically analyzed in order to explore the percentage and nature of missing values in each data set using Analyze Patterns and Missing Value Analysis functions on SPSS. The preliminary screening, summarized in Figure 4 and Figure 5,

revealed that each variable in each data set had at least one missing score. Of the 67 cases, 10 cases (14.93%) had missing values in task 1 data set, and 11 cases (16.42%) had missing values in task 2 data set. Moreover, the overall numbers of missing values in task 1 data and task 2 data sets were 120 (4.975%) and 121 (5.017%) respectively.

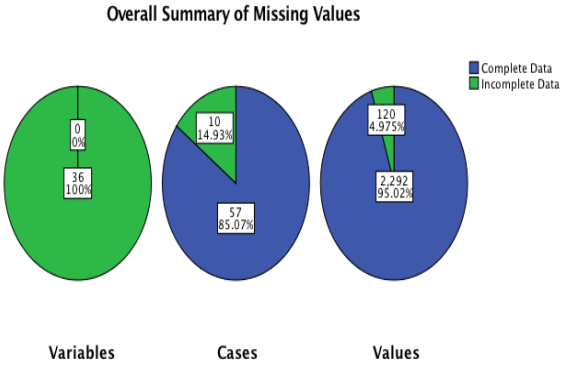


Figure 4. Missing cases in task 1 data set

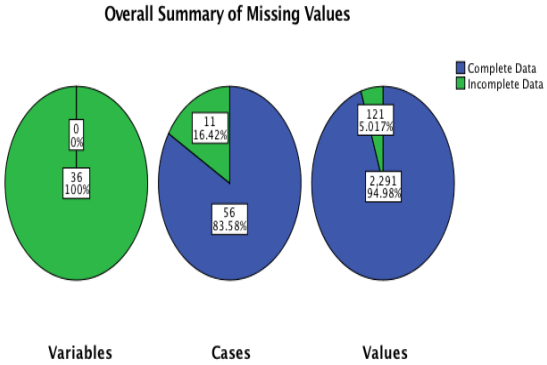


Figure 5. Missing cases in task 2 data set

As for the percentage of missing values within variables, the highest percentage in both data sets was in the delayed posttest variables, as shown in Table 4 below. The lowest percentage of missing values, on the other hand, was in the posttest variables of task 1 data set and the pretest and posttest variables of task 2 data set.

Table 4

The Number and Percentage of Missing Cases

	Task 1 Data Set						Task 2 Data Set					
	Total Number			Total Percentage			Total Number			Total Percentage		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Frequency	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Occurrence	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Correct	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Wrong	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Use	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Structure	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Spelling	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Error	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Words	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Correct%	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Error%	2	1	7	3%	1.5%	10.4%	1	1	8	1.5%	1.5%	11.9%
Evaluation	2	1	7	3%	1.5%	10.4%	1	1	9	1.5%	1.5%	13.4%

In order to statistically evaluate any recurrent patterns of missing data, *Little's Missing Completely at Random* test, also referred to as *Little's MCAR* test, was computed on both data sets. *Little's MCAR* is an '*omnibus*' statistical test that simultaneously analyzes multiple variables to determine whether the missing values in a data set are MCAR, that is, $p \geq .05$ (Schlomer et al., 2010; Van Ness, Murphy, Araujo, Pisani & Allore, 2007). The results of the two *Little's MCAR* tests statistically affirmed that the missing data mechanism in task 1 data set (x^2=.001, *df*= 72, *p*=1.000) and task 2 data (x^2= 51.58, *df*= 107, *p*=1.000) was MCAR.

Since the overall percentage of missing values in each data set was approximately 5%, the percentage of missing values within variables did not exceed the suggested cut-off percentage in the literature for variable deletion (i.e., 15% or higher), and the missing data mechanism was MCAR (i.e., $p>.05$), the missing values were not likely to affect the subsequent statistical analyses and, hence, could be either ignored or imputed using different statistical procedures (Meyers et al., 2006; Tabachnick & Fidell, 2001, 2014). Among the procedures suggested in the literature are *listwise* and *pairwise deletion*, *mean substitution*, *expectation maximization*, *regression*, *multiple imputation*, and *manual imputation* (Meyers et al., 2006; Myers, Well & Frederick, 2010; Tabachnick & Fidell, 2001, 2007a, 2007b, 2014).

Listwise and *pairwise deletions*, the default options in many statistical packages including SPSS, are relatively good options when a small percentage of cases in a large data set is missing (i.e., 5% or less); however, these procedures result in sample size reduction, which is a serious issue for small sample sizes and may sometimes provide biased results if the cases with missing values differ systematically from those with

complete values (Fox-Wasylyshyn & El-Masri, 2005; Meyers et al., 2006; Myers et al., 2010; Tabachnick & Fidell, 2001, 2014). *Mean substitution* is an imputation method in which the mean of a variable is calculated and the mean score is used to substitute all missing cases within that variable; despite its popularity and widespread application, *mean substitution* is believed to reduce the variance of a variable as a result of repeatedly using a single score, which is closer to the mean score rather than the missing value itself (Meyers et al., 2006; Pallant, 2016; Tabachnick & Fidell, 2001).

Expectation maximization is an imputation technique applied to MCAR data, and it provides good and realistic estimates of the missing values; however, it does not take into account the repeated-measures nature of variables in within-subjects designs and is typically used for studies whose results are not based on inferential statistics (Tabachnick & Fidell, 2001, 2014). *Regression* is another imputation technique that can be applied using statistical software. Similar to the other imputation techniques, it does not take into account the repeated-measures nature of variables in within-subjects research designs, is likely to reduce the variance of a data set, and sometimes imputes scores that are far from being real estimates of the missing scores (Tabachnick & Fidell, 2001, 2014).

Multiple imputation is among the most sophisticated and accepted techniques that can be applied to different types of data (i.e., repeated-measures and grouped data); it does not require the data to be MCAR and can be used with inferential statistics (Tabachnick & Fidell, 2001, 2014). However, it generates several different data sets (typically five) that can be overwhelming at the analysis stage and "does not provide the full richness of output that is typical with other methods" (Tabachnick & Fidell, 2014, p. 106). *Manual imputation* is a technique proposed in Tabachnick and Fidell (2001, 2007a,

2007b, 2014) to estimate missing scores in research studies that are solely or partially based on repeated measurements. According to Tabachnick and Fidell (2001, 2007a, 2007b, 2014), *manual imputation* is a reliable and systematic technique to impute missing data in studies whose designs have a repeated-measures component. It is different from other imputation techniques that "do not take into account the commensurate nature of the measurement; that the DV [i.e., dependent variable] is measured on the same scale for all its occurrences… [and] the consistency afforded by the repeated measurement; that is, individual differences" (Tabachnick & Fidell, 2007a, p. 292).

The reliability of this technique is inherent in the required multiple values that relate to a participant's performance over time within a group, and, therefore, it takes into account the repeated-measures nature of data more than any other computer-based imputation techniques (Tabachnick & Fidell, 2001, 2007a, 2007b, 2014). In other words, to manually impute a missing score, several values are needed: (1) the sum of the participant's known values, (2) the sum of the variable in which the missing value occurs, (3) the sum of all variables for that level of measurement, (4) the number of participants, and (5) the number of measurements (Tabachnick & Fidell, 2001, 2007a, 2007b, 2014).

The concerns regarding the reduction of sample size, the likelihood of obtaining biased results from complete data sets, and the limitations of the above-reviewed computer-based imputation techniques (Myers et al., 2010; Pallant, 2016; Tabachnick & Fidell, 2001, 2014) informed the decision of opting for the *manual imputation* technique proposed in Tabachnick and Fidell (2001, 2007a, 2007b, 2014). To this end, the following three equations were created based on Tabachnick and Fidell's (2001, 2007a,

2007b, 2014) equation and applied to each data set in order to impute missing raw scores within each of the six groups at each stage of data collection:

Pre: $\frac{N(participant's\ post+delayed)+Time(Pre{:}Total)-Sum\ of\ the\ total\ pre,post\ and\ delayed}{(Time-1)(Total\#of\ participants-1)}$

Post: $\frac{N(participant's\ pre+delayed)+Time(post{:}Total)-Sum\ of\ the\ total\ pre,post\ and\ delayed}{(Time-1)(Total\#of\ participants-1)}$

Delayed: $\frac{N(participant's\ pre+post)+Time(delayed{:}Total)-Sum\ of\ the\ total\ pre,post\ and\ delayed}{(Time-1)(Total\#of\ participants-1)}$

These formulas are based on the equation proposed in Tabachnick and Fidell (2001, 2007a, 2007b, 2014); mathematical symbols are replaced with verbal explanation to clarify the values that were used to impute missing scores at different stages of data collection.

Based on the above listed equations, all missing scores were imputed, and all the imputed scores were added to Tables D2 through D7 and highlighted in yellow.

After manually imputing the missing scores, the overall mean scores for the complete and imputed data sets were computed and the mean differences were manually calculated by subtracting the mean score for the complete variable from the one for the imputed variable in order to identify any major differences between both data sets. As shown in Table 5 below, 16 out of 36 (i.e., 44.44%) of the imputed variables in task 1 data set had slightly higher mean scores than the variables in the complete data sets, while one imputed variable had a mean score identical to the complete one. Moreover, the mean differences between the imputed and the complete data sets ranged from 2.07 to .01, with the majority of mean differences (91.67%) below 1. Most of the imputed variables in Task 2 data set, on the other hand, had lower mean scores than those in the complete data set, as shown in Table 5 below, with an exception of four mean scores (i.e., 11.11%); two variables in the imputed data set had identical mean scores to the variables in the complete data set; and the mean differences ranged from 1.91 to .01, with the majority of mean differences below 1 (91.67%).

Table 5

Mean Scores and Mean Differences for Complete and Imputed Data Sets

		Task 1			Task 2		
		T1	T1	T3	T1	T2	T3
Frequency	Complete	13.31	15.97	16.05	9.36	11.65	10.34
	Imputed	13.42	16.11	15.71	9.29	11.55	10.16
	Difference	-.11	-.14	.34	.07	.1	.18
Occurrence	Complete	8.37	10.12	10.05	5.65	7.38	6.39
	Imputed	8.40	10.21	9.89	5.61	7.34	6.32
	Difference	-.03	-.09	.16	.04	.04	.07
Correct	Complete	8.06	10.88	11.07	5.18	7.14	6.73
	Imputed	8.25	10.94	10.78	5.17	7.10	6.64
	Difference	-.19	-.06	.29	.01	.04	.09
Wrong	Complete	5.25	5.09	4.98	4.18	4.52	3.61
	Imputed	5.17	5.16	4.93	4.12	4.45	3.53
	Difference	.08	-.07	.05	.06	.07	.08
Use	Complete	.43	.41	.28	.48	.67	.41
	Imputed	.41	.41	.26	.47	.67	.43
	Difference	.02	0	.02	.01	0	-.02
Structure	Complete	4.26	3.79	3.88	3.11	3.24	2.69
	Imputed	4.22	3.85	3.82	3.06	3.19	2.61
	Difference	.04	-.06	.06	.05	.05	.08
Spelling	Complete	.55	.89	.82	.59	.61	.51
	Imputed	.53	.90	.85	.58	.60	.48
	Difference	.02	-.01	-.03	.01	.01	.03
Errors	Complete	22.18	17.29	16.40	41.48	37.12	34.03
	Imputed	22.02	17.17	16.44	41.04	36.85	33.84
	Difference	.16	.12	-.04	.44	.27	.19
Words	Complete	120.72	115.08	117.92	220.53	223.88	221.93
	Imputed	121.08	114.73	115.85	218.62	222.18	221.15
	Difference	-.36	.35	2.07	1.91	1.70	.78
Correct%	Complete	60.68	66.86	68.15	52.89	61.21	65.12
	Imputed	61.43	66.76	66.99	53.59	61.65	65.44
	Difference	-.75	.1	1.16	-.7	-.44	-.32
Errors%	Complete	18.32	15.20	14.24	18.88	16.88	15.79
	Imputed	18.14	15.12	14.71	18.79	16.80	15.79
	Difference	.18	.08	-.47	.09	.08	0
Evaluation	Complete	68.57	68.17	66.63	63.27	64.80	64.12
	Imputed	68.76	68.37	65.79	62.91	64.32	62.83
	Difference	-.19	-.2	.84	.36	.48	1.29

Note. Red: Mean scores for imputed variables > those in complete data sets; green: Identical mean scores

The mean differences between the complete and imputed data sets indicate that the missing values, especially in Task 2 data set, are associated with cases that had lower scores than those who had complete scores. This was inferred from the lower mean scores for the majority of the imputed variables (i.e., 83.33%) for Task 2. The same seems to hold true for Task 1 data set; that is, 52.78% of the complete variables have higher mean scores than the imputed ones. Since the differences between most of the imputed and completed variables were very minimal (i.e., less than 1), the imputed data sets were used at the data analysis stage in order to avoid sample size reduction and/or biased results and boost statistical power (Myers et al., 2010; Pallant, 2016; Tabachnick & Fidell, 2001, 2014).

Descriptive statistics. Computing *descriptive statistics* is a preliminary analytical phase of almost all quantitative research studies (Dörnyei, 2007; Tabachnick & Fidell, 2001, 2007a; Vogt, 2007). Two major categories fall under *descriptive statistics*: *measures of central tendency* (i.e., mean, median, and mode) that help summarize the data in an incisive manner and *measures of variability* (i.e., standard deviation, minimum, maximum, and range) that measure the variance and distribution of scores in a dataset (Dörnyei, 2007; Tabachnick & Fidell, 2007a; Vogt, 2007). Although this preliminary analysis stage does not lead to drawing conclusions, it is seldom optional as concisely describing data and measuring its dispersion are essential components of any statistical analysis (Dörnyei, 2007; Tabachnick & Fidell, 2001, 2007a; Vogt, 2007).

Measures of central tendency and *measures of variability* were computed on the raw scores in both data sets. They were run to summarize the raw scores for the frequency, variety, and accuracy of the formulaic sequences that were utilized in the 67

EAP students' summaries and essay responses; the number and percentage of errors they made when summarizing graphical information and responding to essay prompts; the number of words they generated for both task 1 and task 2; and the EAP judge's evaluation of summaries and essays at the three production stages.

Mixed analysis of variance (ANOVA). *A mixed ANOVA*—also known as *mixed between-within-subjects ANOVA*, *split-plot ANOVA*, and *SPANOVA*—is a univariate statistical test that is computed on data collected from multiple groups at different time points (Meyers et al., 2006; Pallant, 2016; Stevens, 2002; Tabachnick & Fidell, 2001, 2007a, 2014). It melds features of *randomized-group ANOVAs*, which test mean differences between different groups, and *repeated-measures ANOVAs*, which examine mean differences over time within a single group (Meyers et al., 2006; Pallant, 2016; Stevens, 2002; Tabachnick & Fidell, 2001, 2007a, 2014). In its basic form, a *mixed ANOVA* is computed on data with one between-subjects variable and one within-subjects variable, but it can be extended to a *mixed factorial ANOVA* that includes multiple independent variables (Meyers et al., 2006; Tabachnick & Fidell, 2001, 2007a, 2014).

Mixed (factorial) ANOVAs are followed by a post *ANOVA* test, known as *simple-effects analysis*, when, and only when, a statistically significant interaction between two or more variables is identified (Meyers et al. 2006; Pallant, 2016; Tabachnick & Fidell, 2007a). This post *ANOVA* test is computed by holding an independent variable or interaction constant (usually the nonsignificant one) in order to examine the differences within each group over time and/or differences between different groups at each time point (Meyers et al., 2006; Myers et al., 2010; Tabachnick & Fidell, 2007a). The power of the *Mixed (factorial) ANOVA* test lies (a) in its relative robustness to unequal sample

sizes and slight departures from normality provided that there are no univariate outliers or violations to the homogeneity of variance and sphericity assumptions and (b) in simultaneously identifying statistically significant differences and/or interactions between and within different groups (Fidell & Tabachnick, 2003; Meyers et al., 2006; Pallant, 2016; Tabachnick & Fidell, 2001, 2007a).

Given the statistical power and efficiency of *mixed (factorial) ANOVA*, multiple *mixed factorial ANOVAs* with two between-subjects variables (i.e., group and proficiency level) and one within-subjects variable (i.e., repeated measurements) were computed on task 1 and task 2 data sets in order to answer the first three research questions that guided this study. When running each *mixed factorial ANOVA*, several statistical procedures were used. First, a *Bonferroni* correction was applied to each round of statistical analysis in order to control for the inflated *Type I error* (i.e., falsely rejecting the null hypothesis) in the within-subjects factor, which typically results from running multiple univariate tests (Meyers et al. 2006; Stevens, 2002; Tabachnick & Fidell, 2007a).

Second, the *Tukey-Kramer* post hoc test was computed to identify any statistically significant differences between different groups. *Tukey-Kramer* test was selected because it is one of the most popular post hoc tests for three or more groups, and it neither requires equal sample size nor results in inflated *Type I error* (Hilton & Armstrong, 2006; Meyers et al., 2006; Myers et al., 2010; Tabachnick & Fidell, 2001, 2007a). Third, the *multivariate test* results were reported when the sphericity assumption was violated (i.e., $p<.05$), for multivariate results do not require sphericity (Fidell & Tabachnick, 2003; Pallant, 2016; Tabachnick & Fidell, 2007a). Fourth, when interpreting the *mixed factorial ANOVA* results, an alpha level of $p\leq.05$ was used, and more conservative alpha levels of

$p \leq .025$ and $p \leq .01$ were respectively followed when the homogeneity of variance assumption was moderately (i.e., at one or two levels) and severely (i.e., at all three levels) violated (Fidell & Tabachnick, 2003; Tabachnick & Fidell, 2001, 2007a).

Finally, each *mixed factorial ANOVA* whose results yielded a statistically significant two- or three-way interaction was followed by a *simple-effects analysis*, which was computed by splitting the data and running separate *mixed ANOVAs*. When a significant three-way statistically significant interaction (i.e., time*group*proficiency) was detected and neither of the two-way interactions was significant, the data set was split based on the proficiency grouping variable, and the SPSS syntax was modified by adding the following text (Hinton, McMurray, & Brownlow, 2014; IBM Support, n.d.):

COMPARE (Groups) ADJ(BONFERRONI)

EMMEANS=TABLES(Groups*Time) COMPARE (Time) ADJ(BONFERRONI)

The same procedure was followed when any statistically significant interaction was detected. Although Tabachnick and Fidell (2007a) recommend holding the nonsignificant interaction constant before running a *simple-effects analysis*, the data were not split based on the groups variable because a major objective of this research study was to compare the performance of the EAP students who were assigned different pedagogical interventions rather than to explore the effects of each instructional method on the writing production of EAP students at different proficiency levels.

It is worth noting here that although running separate *mixed factorial ANOVAs* is discouraged in the literature because of the possible inflation of *Type I error* that results from running multiple univariate tests (Meyers et al., 2006; Tabachnick & Fidell, 2001), multiple *mixed ANOVAs* rather than a single *multivariate analysis of variance*

(*MANOVA*) were performed because the data violated two major assumptions of *MANOVA*. The use of *MANOVA* is conditional on several statistical assumptions, of which sample size (i.e., $n>$ the number of dependent variables) and the presence of strong negative correlations between dependent variables are crucial (Meyers et al., 2006; Stevens, 2002; Tabachnick & Fidell, 2001). In both data sets, there were fewer participants than dependent variables and the correlation assumption was violated: some variables in both data sets had very strong positive correlations (e.g. $r>.80$), other variables did not correlate ($p>.05$), and very few variables had statistically significant negative correlations.

Running *MANOVA* on positively correlated variables and/or uncorrelated variables is believed to be both 'wasteful' and 'counterproductive' because strong positive correlations do not only indicate that the variables measure the same aspects or represent subscales of major constructs but may also lead to *multicollinearity*, which yields invalid results (Meyers et al., 2006; Tabachnick & Fidell, 2001). Therefore, adapting a more conservative univariate approach by applying a Bonferroni correction to each *mixed ANOVA* is recommended over MANOVA in such cases (Meyers et al., 2006; Tabachnick & Fidell, 2001). Running separate *mixed factorial ANOVAs* was, therefore, considered more appropriate and reliable than computing a *MANOVA*. Moreover, although a less stringent alpha level (e.g. .10 or .15) is recommended for studies with small sample sizes ($n\leq 20$) to control for *Type II error* (i.e., failing to reject the null hypothesis despite the presence of a statistically significant difference) and increase power (e.g. Pituch & Stevens, 2016; Stevens, 2002; Vogt, 2007), combining a Bonferroni adjustment with an alpha level of $p\leq.05$ and sometimes more conservative ones in this

study was intended to fully control for any possible inflated *Type I error* that may result from running multiple univariate tests.

Pearson correlation coefficients. The *Pearson correlation coefficient*, also known as *Pearson product-moment correlation coefficient* or *Pearson r*, is a bivariate statistical test performed on continuous raw or mean scores to index the degree, direction, and proportion of association between two variables (Bland, 2015; Bland & Altman, 1995a, 1995b; Dörnyei, 2007; Meyers et al., 2006; Vogt, 2007). In *repeated-measures* designs, a *Pearson correlation coefficient* is typically computed on the mean scores for each case in order to investigate whether an overall increase in one variable is accounted for by an overall increase in another variable (Bland, 2015; Bland & Altman, 1995a, 1995b). Since the data of this study were collected at different time points, multiple *Pearson correlation coefficients* were computed on the average scores for the frequency and occurrence of the target formulaic sequences, the number of errors, the number of words, and the evaluation of the EAP judge to identify any statistically significant positive and/or negative associations between these variables and, hence, answer the fourth and fifth research questions.

Screening tests. Since the validity of the *mixed ANOVA* and the *Pearson correlation coefficient* results is contingent on a number of major statistical assumptions that the data need to meet, three screening tests were run to assess these assumptions prior to computing inferential statistical tests.

Outliers. Screening for outliers (i.e., values that have extreme high or low scores) prior to computing inferential statistics is essential since most statistical tests, including *mixed factorial ANOVAs* and *Pearson correlation coefficient*, are sensitive to outliers as

they inflate *Type I error* and/or *Type II error* (Dörnyei, 2007; Fidell & Tabachnick, 2003; Meyers et al., 2006; Pituch & Stevens, 2016; Stevens, 2002; Tabachnick & Fidell, 2007a). Given their adverse effects on the validity of the obtained results, any identified outliers should be deleted or altered, or variables including outliers should be transformed prior to running major analyses (Meyers et al., 2006; Tabachnick & Fidell, 2007a, 2014).

Outliers can be identified graphically by generating *boxplots*, in which any potential outliers are located outside the boundaries of the whiskers (Meyers et al., 2006; Tabachnick & Fidell, 2001, 2014). They can also be detected statistically by computing *z scores*, which "indicates how many standard deviation units a particular score lies from the mean of the distribution" (Meyers et al., 2006, p. 129). When a *z score* for a particular value is greater than +/-3.29 in large enough data sets (i.e., $n \geq 100$) or +/-2.58 in smaller data sets, that particular score is an outlier (Fidell & Tabachnick, 2003; Meyers et al., 2006; Stevens, 2002; Tabachnick & Fidell, 2001, 2007a, 2014). If the detected outliers are true scores from the target population rather than the result of data entry errors, they should be adjusted rather than omitted (Fidell & Tabachnick, 2003; Tabachnick & Fidell, 2001, 2007a). This can be done by either transforming the target variable(s) or adjusting the score of an outlier by replacing it with the following highest or lowest score (Fidell & Tabachnick, 2003; Tabachnick & Fidell, 2007a).

In order to screen both data sets for outliers, *boxplots* and *z scores* were run on the variables that represented the three production stages of each of the six groups prior to computing *mixed factorial ANOVAs* and on the mean scores for all groups combined before running *Pearson correlation coefficients* (see Appendix E). When the *z scores* for the potential outliers identified in the boxplot exceeded +/-2.58, the detected outliers were

replaced with the next most extreme score whose *z score* did not exceed +/-2.58 within the same variable (see Appendix E). The decision to alter outlying scores was based on two major arguments proposed in the literature: (1) when outlying scores are not due to data entry errors, they should be altered rather than deleted in order not to distort the data, and (2) data transformation is a possible but not recommended option as it generates scores that are different from the actual scores (Fidell & Tabachnick, 2003; Meyers et al., 2006; Tabachnick & Fidell, 2001, 2007a, 2014).

Normality tests. After screening for and altering outliers, the distribution of scores was assessed statistically, which was another crucial step because most inferential statistics, including *mixed factorial ANOVAs* and *Pearson correlation coefficients*, assume that the data are approximately normally distributed (Dörnyei, 2007; Fidell & Tabachnick, 2003; Meyers et al., 2006; Stevens, 2002; Tabachnick & Fidell, 2001, 2007a). While a visual assessment of distribution (e.g. histograms) is very common in studies that have a large enough sample ($n \geq 100$), it does not "provide a definitive indication of a normality violation" (Meyers et al., 2006, p. 68). A more formal and accurate method of assessing normality, especially in the case of small and/or unequal sample sizes, is through the computation of the skewness and kurtosis *z scores* and the *Shapiro-Wilk* normality test (Fidell & Tabachnick, 2003; Ghasemi & Zahediasl, 2012; Meyers et al., 2006; Stevens, 2002; Tabachnick & Fidell, 2001).

Skewness and kurtosis values are part of *descriptive statistics* that measure that 'asymmetry' and 'peakedness' of distribution in a dataset (Fidell & Tabachnick, 2003; Kim, 2013; Tabachnick & Fidell, 2001, 2007a). When assessing the distribution of data, these values are divided by their standard error scores in order to elicit *z sco*res that

statistically depict any normality violations (Fidell & Tabachnick, 2003; Ghasemi & Zahediasl, 2012; Kim, 2013). The *Shapiro-Wilk* normality test is another statistical procedure that is used to test the normality assumption; it provides more powerful and reliable results than other tests (e.g. Kolmogorov-Smirnov), and its results are robust in the case of a small sample size, that is, $n < 20$ (Meyers et al., 2006; Razali & Wah, 2011; Shapiro & Wilk, 1965; Stevens, 2002; Tabachnick & Fidell, 2007a). Normally distributed data with a small sample size have skewness and kurtosis *z scores* that fall within a -1.96 and +1.96 range and a nonsignificant *Shapiro-Wilk* test results, that is, $p > .01$ (Ghasemi & Zahediasl, 2012; Kim, 2013; Meyers et al., 2006; Razali & Wah, 2011; Stevens, 2002; Tabachnick & Fidell, 2007a).

In order to assess the distribution of the dependent variables in each data set, skewness and kurtosis *z scores* and *Shapiro-Wilk* normality tests were calculated because of the small sample size in this research study and the statistical rigor of these two statistical tests (see Appendix F). These values were calculated for the variables that represented the sum of scores within each group for *mixed factorial ANOVAs* and the mean scores for *Pearson correlation coefficients*. Computing normality tests on the sum and the mean scores for the three production stages rather than the raw scores at each production stage was informed by (a) the suggestion that assessing the normality of sampling distribution in studies with a repeated-measures component prior to computing *ANOVAs* requires generating summative variables because "the scores that are tested are not the raw scores, but rather the sums of raw scores" (Tabachnick & Fidell, 2007a, p. 247) and (b) the fact that *Pearson correlation coefficients* assume that the dependent

variable (i.e., the mean scores in this study) is approximately normally distributed (Dörnyei, 2007).

Scatterplots. Since among the major assumptions of the *Pearson correlation coefficient* is the presence of a linear relationship between the two target variables (Dörnyei, 2007; Hinton et al., 2014), multiple scatterplots were generated to visually inspect the relationship between each pair of dependent and independent variables prior to running correlations (see Appendix G).

Chapter Five: Findings

The fifth chapter sets forth the findings that were obtained from the statistical analyses and answers the five research questions that guided the present study. To this end, this chapter is divided into four major sections, each of which focuses on the descriptive and inferential statistical tests that were multiply carried out on different sets of quantitative variables that were elicited from the textual data during the data quantification stage and/or the textual evaluation that was assigned by the EAP judge.

The first three sections introduce the results of the *measures of central tendency* and *measures of variability* that were run on several quantitative variables prior to adjusting outlying scores in order to summarize them. They then discuss the major findings obtained from computing multiple *mixed factorial ANOVAs* to highlight any statistically significant differences within and/or between different groups and respectively answer the first three research questions that mainly focused on the effectiveness of different presentation methods of formulaic sequences in promoting EAP students' writing proficiency. It is worth reiterating here that the results of the *mixed factorial ANOVAs* and the follow-up *simple-effects analyses* were considered statistically significant at $p \leq .05$ when the data met the homogeneity of variance assumption and at $p \leq .025$ and $p \leq .01$ when the data moderately and severely violated that assumption (Fidell & Tabachnick, 2003; Tabachnick & Fidell, 2001, 2007a). Moreover, results of within-subjects effects were reported unless the data violated the sphericity assumption, in which case the multivariate results were reported (Pallant, 2016; Tabachnick & Fidell, 2007a).

The fourth section presents the results of the *Pearson correlation coefficients* that were computed on the mean scores for the performance of the six groups combined at the

three production stages in order to reveal any statistically significant positive and/or negative associations between different pairs of dependent and independent variables and answer the fourth and fifth research questions that guided this study. When interpreting the *Pearson correlation coefficient* results, an alpha level of $p \leq .05$ exhibited a statistically significant correlation.

Research Question One

In order to summarize part of the textual data and answer the first research question that guided this study—*Will both explicit instruction and saliency of formulaic sequences help EAP students (a) acquire and properly utilize a wider range of formulaic sequences when approaching different writing tasks and (b) outperform EAP students who practice other writing materials?—descriptive statistics* and *mixed factorial ANOVAs* were computed on the quantitative variables that represented frequency, occurrence, correct, wrong, and correct% at the three productions stages for both graphical summaries (i.e., task 1) and essay responses (i.e., task 2).

Frequency (task 1). *Descriptive statistics* were computed on the raw scores for the frequency of the target formulaic sequences to summarize the scores and detect any possible differences in the total number of the formulaic sequences the six groups produced when summarizing graphical information at different time points. They were followed by *mixed factorial ANOVAs* to determine whether the detected differences within and between groups were statistically significant.

According to Table 6, the upper-intermediate focused instruction, saliency, and control groups used more formulaic sequences in their posttests (M=15.38, SD=4.61, n=13; M=14.80, SD=4.82, n=5; M=12.80, SD=2.15, n=10) than they did in their pretests

(*M*=13.92, *SD*= 3.25, *n*=13; *M*=10.80, *SD*=6.91, *n*=5; *M*=11.20, *SD*=4.02, *n*=10). However, the number of formulaic sequences the three upper-intermediate groups used in their delayed posttests dropped in dissimilar ways. While the saliency group used more formulaic sequences in the delayed posttest (*M*=13.80, *SD*=5.76, *n*=5) than they did in the pretest, the focused instruction and control groups used fewer formulaic sequences in the delayed posttest (*M*=11.69, *SD*=6.60, *n*=13 and *M*=10.02, *SD*=4.46, *n*=10) than they did in the pretest.

Table 6

Descriptive Statistics: Frequency (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	13.92	15.38	11.69	10.80	14.80	13.80	11.20	12.80	10.02
Median	14.00	16.00	10.00	14.00	16.00	14.00	12.50	13.0	11.00
Mode	16.00	10.00[a]	9.00	2.00[a]	8.00[a]	14.00	8.00	13.00	11.00
SD	3.25	4.61	6.60	6.91	4.82	5.76	4.02	2.15	4.46
Minimum	7.00	9.00	4.00	2.00	8.00	8.00	4.00	9.00	1.00
Maximum	19.00	25.00	26.00	18.00	20.00	23.00	17.00	16.00	17.00

a. Multiple modes exist. The smallest value is shown

The three advanced groups, on the other hand, had dissimilar patterns of using formulaic sequences at the three production stages, as shown in Table 7 below. The frequency of the target formulaic sequences in the advanced focused instruction group increased constantly from the pretest (*M*=14.61, *SD*=5.50, *n*=19) to the posttest (*M*=18.16, *SD*=5.67, *n*=19) and then to the delayed posttest (*M*=21.15, *SD*=5.55, *n*=19). The advanced saliency group used the highest number of formulaic sequences in the posttest (*M*=17.56, *SD*=6.03, *n*=13), which dropped slightly in the delayed posttest (*M*=15.02, *SD*=5.58, *n*=13) but was still higher than the pretest (*M*=12.85, *SD*=5.57, *n*=13). The advanced control group, in turn, used more formulaic sequences in the pretest

(*M*=15.33, *SD*= 3.68, *n*=7) and the delayed posttest (*M*=19.14, *SD*=4.10, *n*=7) than they did in the posttest (*M*=14.86, *SD*=3.58, *n*=7).

Table 7

Descriptive Statistics: Frequency (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	14.61	18.16	21.15	12.85	17.56	15.02	15.33	14.86	19.14
Median	14.00	18.00	21.00	10.00	18.00	16.00	15.33	12.00	21.00
Mode	10.00[a]	22.00	15.00[a]	10.00	14.00	12.00[a]	11.00[a]	12.00	15.00
SD	5.50	5.67	5.55	5.57	6.03	5.58	3.68	3.58	4.10
Minimum	5.00	5.00	13.03	6.00	6.00	4.00	11.00	12.00	15.00
Maximum	27.00	27.00	32.00	24.00	28.00	26.00	21.00	19.00	25.00

a. Multiple modes exist. The smallest value is shown

In order to determine whether the detected differences in Table 6 and Table 7 within and between different groups were statistically significant, a *mixed factorial ANOVA* was computed on the variables representing the total number of the formulaic sequences the six groups used in their graphical summaries after screening these variables for outliers, assessing the distribution of the sum of scores, and checking the assumptions of sphericity and homogeneity of variance. An inspection of the *boxplots* and *z scores* generated to identify outlying scores suggested that there were no significant outliers within any of the six groups at the three production stages (see Table E1 and Table E2). The normality tests calculated on the summative variables that included the total number of formulaic sequences the six groups utilized at the three production stages confirmed that the distribution of scores within each group was approximately normal (see Table F1). Mauchly's test and Levene's test results, in turn, indicated that these variables met the sphericity (W=1.00, x^2=.021, *df*=2, *p*=.990) and homogeneity of variance [T1: $F(5, 61)=1.88$, $p=.111$; T2: $F(5, 61)=1.96$, $p=.098$; T3: $F(5, 61)=.39$, $p=.851$] assumptions.

According to Table 8, almost all the within-subjects effects, except for the time×group interaction (p>.05), were statistically significant. The statistically significant main effect of time $F(2, 122)=6.78$, $p=.002$, partial $\eta^2=.10$ suggested that the overall use of formulaic sequences by the six groups combined differed over time. Another statistically significant difference was the time×proficiency interaction $F(2, 122)=5.74$, $p=.004$, partial $\eta^2=.086$, which indicated that the upper-intermediate and advanced EAP students, irrespective of their groups, used different numbers of formulaic sequences at the three production stages. The time×group×proficiency interaction was also statistically significant $F(4, 122)=2.75$, $p=.031$, partial $\eta^2=.083$, implying that different groups at different proficiency levels used different numbers of formulaic sequences at the three productions stages, which could also be visually depicted in Figure 6 and Figure 7 below.

Table 8

Frequency (Task 1): Tests of Within-Subjects Effects

Source		Type III SS.	df	Mean Sq.	F	Sig	Partial Eta Sq.
Tm.	Sphericity As.	192.312	2	96.156	6.780*	.002	.100
Tm.*Gr.	Sphericity As.	57.911	4	14.478	1.021	.399	.032
Tm.* Pro.	Sphericity As.	162.750	2	81.375	5.737*	.004	.086
Tm.*Gr.*Pro.	Sphericity As.	156.062	4	39.016	2.751*	.031	.083
Error	Sphericity As.	1730.365	122	14.183			

*$p \leq .05$

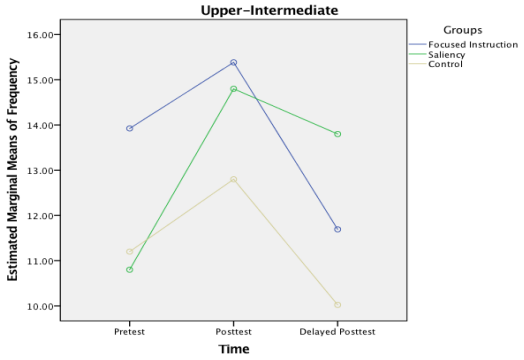


Figure 6. Frequency task 1 inter.

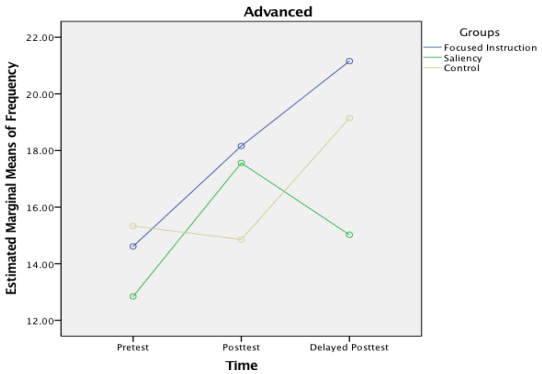


Figure 7. Frequency task 1 adv.

The results of the between-subjects effects, presented in Table 9 below, showed that the only significant main effect was that of proficiency $F(1, 61)=11.96$, $p=.001$ with a very large effect size (partial $\eta^2=.164$), suggesting that the overall performance of the upper-intermediate and advanced EAP students, regardless of their groups, was different. Neither the effect of groups alone nor the group×proficiency interaction was statistically significant ($p>.05$), as shown in Table 9 below.

Table 9

Frequency (Task 1): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig	Partial Eta Sq.
Intercept	35537.575	1	35537.575	705.353	.000	.920
Groups	153.864	2	76.932	1.527	.225	.048
Proficiency	602.782	1	602.782	11.964*	.001	.164
Groups*Proficiency	60.155	2	30.078	.597	.554	.019
Error	3073.342	61	50.383			

*$p\leq.05$

Given the significant three-way interaction and the nonsignificant time×group interaction, a *simple-effect analysis* was performed after splitting the data based on proficiency in order to reanalyze the nonsignificant time×group interaction at each proficiency level. The data met the sphericity assumption at the upper-intermediate level ($W=.92$, $x^2=2.04$, $df=2$, $p=.360$) and the advanced level ($W=.97$, $x^2=.96$, $df=2$, $p=.619$). The data also met the homogeneity of variance assumption at the advanced level [T1: $F(2, 36)=.80$, $p=.458$; T2: $F(2, 36)=.62$, $p=.543$; T3: $F(2, 36)=.26$, $p=.776$] but moderately violated it at the upper-intermediate level [T1: $F(2, 25)=5.09$, $p=.014$; T2: $F(2, 25)=3.11$, $p=.062$; T3: $F(2, 25)=.63$, $p=.539$]. A more conservative alpha level ($p\leq.025$) was used for the interpretation of all the effects for the upper-intermediate groups in order to reduce the inflated *Type I error*.

The results of the *simple-effect analysis*, presented in Table 10 below, revealed a statistically significant main effect of time at the advanced proficiency level $F(2, 72)=10.30$, $p=.001$, partial $\eta^2=.222$, which suggested that the number of formulaic sequences used by the three advanced groups combined significantly differed over time. Another statistically significant effect was the time by group interaction in the three advanced groups $F(4, 72)=3.43$, $p=.013$, partial $\eta^2=.160$, implying that the performance of the three advanced groups was different at the three production stages. None of the within-subjects effects for the three upper-intermediate groups approached statistical significance ($p>.025$), nor did any of the between-subjects effects (see Table 11 below) for the three upper-intermediate groups ($p>.025$) or the three advanced groups ($p>.05$).

Table 10

Frequency (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	93.598	2	46.799	3.326	.044	.117
	Time*Group	50.507	4	12.627	.897	.473	.067
	Error	703.460	50	14.069			
Advanced	Time	293.664	2	146.832	10.295*	.000	.222
	Time*Group	195.768	4	48.942	3.432*	.013	.160
	Error	1026.905	72	14.263			

*$p\leq.05$ for the advanced groups

Table 11

Frequency (Task 1): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	11578.308	1	11578.308	294.181	.000	.922
	Groups	94.554	2	47.277	1.201	.318	.088
	Error	983.945	25	39.358			
Advanced	Intercept	27052.668	1	27052.668	466.114	.000	.928
	Groups	188.285	2	94.142	1.622	.212	.083
	Error	2089.396	36	58.039			

In light of the significant effect of time and the time×group interaction, the simple comparisons results were checked to identify any significant differences within and/or

between the three advanced groups. The simple comparisons results demonstrated that the increase in the mean scores for the advanced focused instruction group from (M=14.61, SD=5.50, n=19) in the pretest to (M=18.16, SD=5.67, n=19) in the posttest was statistically significant (MD=-3.54, SE=1.31, p=.031). The advanced focused instruction group's use of the target formulaic sequences in the delayed posttest (M=21.15, SD=5.55, n=19) was also statistically significantly higher than the pretest (MD=6.54, SE=1.22, p=.001) and the posttest (MD=3, SE=1.14, p=.037).

Similarly, the increase in the mean scores for the advanced saliency group from (M=12.85, SD=5.57, n=13) in the pretest to (M=17.56, SD=6.03, n=13) in the posttest was statistically significant (MD=-4.72, SE=1.59, p=.016). The mean score for the advanced saliency group's delayed posttest (M=15.02, SD=5.58, n=13) was not significantly higher than that for the pretest or the posttest (p>.05), nor were the mean differences identified within the advanced control group (p>.05). As for the simple comparisons between the three advanced groups at each production stage, the only statistically significant difference was between the delayed posttest of the advanced focused instruction group (M=21.15, SD=5.55, n=19) and that of the advanced saliency group (M=15.02, SD=5.58, n=13, MD=6.13, SE=1.92, p=.009). It is worth noting that the simple comparisons results for the upper-intermediate groups were checked and confirmed that none of the differences within each upper-intermediate group or between the three upper-intermediate groups at each production stage approached statistical significance (p>.025).

Based on these results, focused instruction of formulaic sequences significantly affected the advanced EAP students' acquisition and retention of formulaic sequences

and helped them outperform the other experimental group (i.e., saliency) at the third production stage. Making formulaic sequences salient in input, in turn, affected the advanced EAP students' immediate learning gains of this language phenomenon with no consequences on retention. These two presentation methods, however, did not affect the performance of the upper-intermediate students whose performance remained relatively similar before and after the training period, so did the performance of the upper-intermediate and advanced control groups.

Occurrence (task 1). *Measures of central tendency* and *measures of variability* were also computed on the raw scores for the distinct types the six groups used in their graphical summaries at the three production stages. The mean scores for the occurrence of the target formulaic sequences in the three upper-intermediate groups at the three production stages had different trends, as shown in Table 12 below.

Table 12

Descriptive Statistics: Occurrence (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	8.85	11.00	8.46	6.00	9.60	10.80	7.50	7.50	6.29
Median	9.00	11.00	8.00	5.00	10.00	9.00	8.00	7.00	7.00
Mode	8.00	13.00	5.00[a]	2.00[a]	10.00	9.00	8.00	7.00	7.00[a]
SD	1.91	3.21	3.62	4.06	1.82	4.02	2.12	1.84	2.31
Minimum	6.00	5.00	3.00	2.00	7.00	9.00	3.00	4.00	1.00
Maximum	12.00	18.00	15.00	12.00	12.00	18.00	10.00	10.00	9.00

a. Multiple modes exist. The smallest value is shown

While the mean scores for the distinct types used by the upper-intermediate saliency group increased constantly from the pretest (M=6, SD=4.06, n=5) to the posttest (M=9.60, SD=1.82, n=5) and then again to the delayed posttest (M=10.80, SD=4.02, n=5), the mean scores for the upper-intermediate control group were identical for both the pretest (M=7.5, SD=2.12, n=10) and the posttest (M=7.5, SD=1.84, n=10) and then

dropped slightly in the delayed posttest (*M*=6.29, *SD*=2.31, *n*=10). The upper-intermediate focused instruction group, in turn, used the highest number of distinct types in the posttest (*M*=11, *SD*=3.21, *n*=13) and the lowest number in the delayed posttest (*M*=8.46, *SD*=3.62, *n*=13).

As for the three advanced groups, Table 13 below shows that both the advanced focused instruction and saliency groups utilized more distinct types in their posttests (*M*=11.21, *SD*=3.01, *n*=19 and *M*=11.23, *SD*=3.85, *n*=13) and their delayed posttests (*M*=11.91, *SD*=2.86, *n*= 19 and *M*=10.20, *SD*=4.31, *n*=13) than they did in their pretests (*M*=8.73, *SD*=2.28, *n*=19 and *M*=8.23, *SD*=3.39, *n*=13). The advanced control group, on the other hand, utilized the highest number of distinct types in the delayed posttest (*M*=11, *SD*=2.31, *n*=7) and the lowest number in the posttest (*M*=8.43, *SD*=2.94, *n*=7).

Table 13

Descriptive Statistics: Occurrence (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	8.73	11.21	11.91	8.23	11.23	10.20	9.96	8.43	11.00
Median	9.00	11.00	11.00	7.00	10.00	10.00	8.75	9.00	11.00
Mode	7.00[a]	11.00	10.00	5.00	9.00	10.00	7.00	6.00[a]	12.00
SD	2.28	3.01	2.86	3.39	3.85	4.31	3.71	2.94	2.31
Minimum	5.00	5.00	8.19	4.00	4.00	3.00	7.00	5.00	8.00
Maximum	13.00	16.00	18.00	14.00	16.00	20.00	17.00	13.00	15.00

a. Multiple modes exist. The smallest value is shown

A *mixed factorial ANOVA* was run to determine whether the detected differences in the mean scores of the distinct types within and between groups were statistically significant. These variables met the assumptions of sphericity (*W*=1.00, x^2=.16, *df*= 2, *p*=.924) and homogeneity of variance in the posttest [*F*(5, 61)=1.87, *p*=.113] and the delayed posttest [*F*(5, 61)=.80, *p*=.558], but not the pretest [*F*(5, 61)=2.45, *p*=.044]. These variables had no significant outliers (see Table E3 and Table E4), and the

distribution of the sum of distinct types within each group was approximately normally (see Table F2). Because of the moderate violation of the homogeneity of variance assumption, an alpha level of $p \leq .025$ was used to interpret all the effects.

The results summarized in Table 14 and Table 15 below indicated that the within-subjects effects of time and time×group interaction and the between-subjects effects of proficiency were statistically significant ($p<.025$). The statistically significant effect of time $F(2, 122)=7.79$, $p=.001$, partial $\eta^2=.113$ indicated that the six groups combined used different numbers of distinct types at different production stages. Another statistically significant effect was the time×group interaction $F(4, 122)=3.77$, $p=.006$, partial $\eta^2=.110$, suggesting that the use of distinct types at different time points was different for different groups, which can also be seen in the visual depiction in Figure 8 and Figure 9 below.

Table 14

Occurrence (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig	Partial Eta Sq.
Tm.	Sphericity As.	93.492	2	46.746	7.791*	.001	.113
Tm.*Gr.	Sphericity As.	90.583	4	22.646	3.774*	.006	.110
Tm.* Pro.	Sphericity As.	17.995	2	8.997	1.500	.227	.024
Tm.*Gr.*Pro.	Sphericity As.	66.965	4	16.741	2.790	.029	.084
Error	Sphericity As.	732.014	122	6.000			

*$p \leq .025$

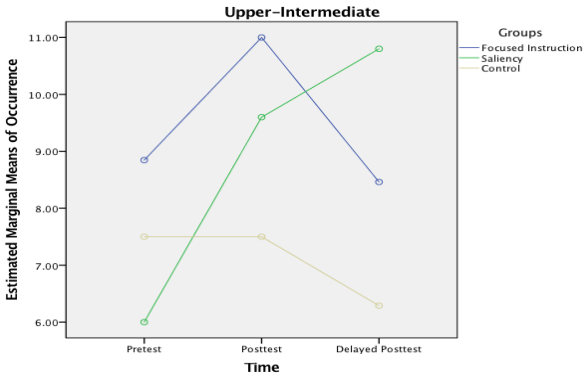


Figure 8. Occurrence task 1 inter.

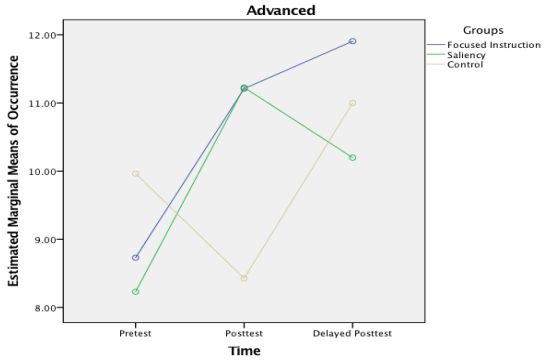


Figure 9. Occurrence task 1 adv.

Table 15

Occurrence: Tests of Between-Subjects Effects

	Type III	df	Mean Sq.	F	Sig	Partial Eta Sq.
Intercept	14298.595	1	14298.595	903.902	.000	.937
Groups	80.928	2	40.464	2.558	.086	.077
Proficiency	113.983	1	113.983	7.206*	.009	.106
Groups*Proficiency	21.870	2	10.935	.691	.505	.022
Error	964.943	61	15.819			

*$p \leq .025$

The statistically significant effect of proficiency $F(1, 61)=7.21$, $p=.009$, partial $\eta^2=.106$ suggested that the overall use of distinct types by the upper-intermediate and advanced EAP students, irrespective of their groups, was statistically significantly different. All other within- and between-subjects effects were statistically non-significant ($p>.025$).

The significant effect of proficiency and time×group interaction, together with the patent differences in Figure 8 and Figure 9 above within and between different groups at different proficiency levels, suggested the need for splitting the data based on proficiency and running *simple-effects analysis* to identify any statistically significant differences within groups over time and/or between groups at each point in time. The *simple-effects analysis* results were all interpreted at an alpha level of $p \leq .05$ because the data, after being split, met the homogeneity of variance assumption at both the upper-intermediate [T1: $F(2, 25)=3.23$, $p=.056$; T2: $F(2, 25)=1.12$, $p=.344$; T3: $F(2, 25)=1.18$, $p=.325$] and advanced [T1: $F(2, 36)=2.18$, $p=.128$; T2: $F(2, 36)=1.39$, $p=.262$; T3: $F(2, 36)=.87$, $p=.429$] proficiency levels. Mauchly's test results for the upper-intermediate ($W=.98$, $x^2=.47$, $df=2$, $p=.790$) and advanced ($W=.97$, $x^2=.91$, $df=2$, $p=.635$) groups also confirmed that the data met the sphericity assumption at both proficiency levels.

The results of the separate *mixed ANOVA*s listed in Table 16 and Table 17 below indicated that all the within-subjects and between-subjects effects for the three upper-

intermediate groups were statistically significant (p.<05). The statistically significant effect of time $F(2, 50)=4.15$, $p=.022$, partial $\eta^2=.142$ and time by group interaction $F(4, 50)=3.61$, $p=.012$, partial $\eta^2=.224$ respectively suggested that the three upper-intermediate groups combined utilized different numbers of distinct types over time and the different numbers of distinct types were different for different groups. The statistically significant effect of groups $F(2, 25)=3.67$, $p=.040$, partial $\eta^2=.227$, in turn, indicated that the overall use of distinct types was different for different groups.

Table 16

Occurrence (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	44.107	2	22.053	4.145*	.022	.142
	Time*Group	76.878	4	19.220	3.612*	.012	.224
	Error	266.025	50	5.320			
Advanced	Time	71.844	2	35.922	5.550*	.006	.134
	Time*Group	64.194	4	16.048	2.480	.051	.121
	Error	465.989	72	6.472			

*$p \leq .05$

Table 17

Occurrence (Task 1): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	5107.583	1	5107.583	396.668	.000	.941
	Groups	94.604	2	47.302	3.674*	.040	.227
	Error	321.906	25	12.876			
Advanced	Intercept	10110.168	1	10110.168	566.010	.000	.940
	Groups	17.040	2	8.520	.477	.625	.026
	Error	643.038	36	17.862			

*$p \leq .05$

The only statistically significant effect in the advanced groups, on the other hand, was that of time $F(2, 72)=5.55$, $p=.006$, partial $\eta^2=.134$ indicating that the three advanced groups combined used different numbers of distinct types at different production stages. All other differences for the advanced groups were statistically nonsignificant (p>.05).

The results of the pairwise comparisons indicated that the upper-intermediate focused instruction group used a statistically significantly higher number of distinct types in the posttest (*M*=11, *SD*=3.21, *n*=13) than they did in the pretest (*M*=8.85, *SD*=1.91, *n*=13, *MD*=2.15, *SE*=.84, *p*=.050); however, their use of distinct types in the delayed posttest (*M*=8.46, *SD*=3.62, *n*=13) dropped to a very similar point to that of the pretest, which was inferred from the statistically significant difference between their posttest and delayed posttest (*MD*=2.54, *SE*=.94, *p*=.036) and the absence of significant difference between their pretest and delayed posttest (*p*>.05). The upper-intermediate saliency group, in turn, used a statistically significantly lower number of distinct types in their pretest (*M*=6, *SD*=4.06, *n*=5) than they did in the posttest (*M*=9.60, *SD*=1.82, *n*=5, *MD*=-3.60, *SE*=1.35, *p*=.040) and the delayed posttest (*M*=10.80, *SD*=4.02, *n*=5, *MD*=-4.80, *SE*=1.51, *p*=.012).

None of the differences within the upper-intermediate control group, on the other hand, was statistically significant; nor were any of the differences between the three upper-intermediate groups at each stage of data collection except for the upper-intermediate focused-instruction group's statistically significantly higher use of distinct types at the posttest stage (*M*=11, *SD*=3.21, *n*=13) than their peers in the control group (*M*=7.50, *SD*=1.84, *n*=10, *MD*=3.50, *SE*=1.09, *p*=.011). As for the multiple comparisons between different groups, the *Tukey-Kramer* post hoc test results indicated that the significant effect of group at the upper-intermediate proficiency level was due to the focused instruction group's higher overall use of distinct types (*M*=9.44, *SE*=.58, *n*=13) than that of the control group (*M*=7.10, *SE*=.66, *n*=10, *MD*=2.34, *SE*=.87, *p*=.033). All

other overall differences between the three upper-intermediate groups were not large enough to approach statistical significance ($p>.05$).

As for the three advanced groups, the simple comparisons indicated that the statistically significant main effect of time was the result of the higher delayed posttest mean score for the three advanced groups combined (M=11.04, SE=.58, n=39) than their pretest mean score (M=8.98, SE=.51, n=39, MD=2.06, SE=.60, p=.005). Because of the apparent increase in the mean scores within the two advanced experimental groups (see Figure 9 above), the pairwise comparisons that were produced as part of the *simple-effects analysis* were also of interest. These results confirmed that the increase in the use of distinct types by the advanced focused instruction group from (M=8.73, SD=2.28, n=19) in the pretest to (M=11.21, SD=3.01, n=19, MD=-2.48, SE=.89, p=.025) in the posttest and to (M=11.91, SD=2.86, n=19, MD=-3.18, SE=.79, p=.001) in the delayed posttest was statistically significant. The pairwise comparisons results also showed that the use of distinct types by the advanced saliency group in the posttest (M=11.23, SD=3.85, n=13) was significantly higher than their use of distinct types in the pretest (M=8.23, SD=3.39, n=13, MD=3, SE=1.08, p=.025). All other differences within the advanced saliency and control groups and between the three advanced groups at each time point were found to be statistically nonsignificant ($p>.05$).

In a nutshell, the results of the *mixed ANOVAs* suggested that focused instruction of formulaic sequences encouraged the upper-intermediate group to use more distinct types of formulaic sequences at the end of the training period without any effects on retention; it also helped them outperform their counterparts in the control group not only at the second production stage but also in their overall use of distinct types. The same

presentation method helped the advanced group acquire and retain more distinct types of formulaic sequences. Making formulaic sequences salient in input, in turn, helped the two saliency groups use a significantly higher number of distinct types at the end of the training period. This presentation method seemed to have a stronger effect on the upper-intermediate students than the advanced students in that they could retain distinct types. Unlike the experimental groups, the use of distinct types by the upper-intermediate and advanced control groups remained relatively stable over time.

Correct (task 1). *Descriptive statistics* were computed on the raw scores for the total number of correct formulaic sequences that the six groups produced in their graphical summaries at the three production stages in order to summarize these scores. Based on the results presented in Table 18 below, the three upper-intermediate groups had very similar trends. The upper-intermediate focused instruction, saliency, and control groups had higher mean scores for their posttests (*M*=10.31, *SD*=5.89, *n*=13; *M*=7, *SD*=4.47, *n*=5; and *M*=8.20, *SD*=3.01, *n*=10) than those for their pretests (*M*=8.38, *SD*=3.10, *n*=13; *M*=5.40, *SD*=4.28, *n*=5; and *M*=6.20, *SD*=4.21, *n*=10). The mean scores

Table 18

Descriptive Statistics: Correct (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	8.38	10.31	7.31	5.40	7.00	5.70	6.20	8.20	6.78
Median	8.00	11.00	6.00	3.00	6.00	5.00	5.50	9.50	8.00
Mode	8.00	4.00	6.00	2.00	2.00[a]	5.00	4.00	10.00	8.00
SD	3.10	5.89	4.77	4.28	4.47	3.83	4.21	3.01	3.59
Minimum	3.00	4.00	0	2.00	2.00	1.50	1.00	3.00	.78
Maximum	14.00	23.00	16.00	11.00	14.00	12.00	14.00	11.00	11.00

a. Multiple modes exist. The smallest value is shown

for the focused instruction, saliency, and control groups' delayed posttests (*M*=7.31, *SD*=4.77, *n*=13; *M*=5.70, *SD*=3.83, *n*=5; and *M*=6.78, *SD*=3.59, *n*=10 respectively) decreased to similar points to those of their pretests.

Table 19

Descriptive Statistics: Correct (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	9.31	13.26	16.50	7.85	11.79	10.27	10.83	11.00	12.00
Median	9.00	13.00	16.00	7.00	12.00	11.00	12.00	10.00	12.00
Mode	8.00	5.00[a]	20.00	6.00[a]	13.00	8.00	8.00[a]	7.00[a]	5.00[a]
SD	4.45	5.54	5.13	4.24	4.12	3.76	4.42	4.04	5.72
Minimum	1.00	5.00	8.00	1.00	4.00	3.00	4.00	7.00	5.00
Maximum	22.00	24.00	28.00	16.00	19.00	17.00	16.00	19.00	22.00

a. Multiple modes exist. The smallest value is shown

The three advanced groups, on the other hand, had different trends, as shown in Table 19 above. The advanced focused instruction, saliency and control groups produced more correct formulaic sequences in their posttests (*M*=13.26, *SD*=5.54, *n*=19; *M*=11.79, *SD*=4.12, *n*=13; *M*=11.00, *SD*=4.04, *n*=7) and their delayed posttests (*M*=16.50, *SD*=5.13, *n*=19; *M*=10.27, *SD*=3.76, *n*=13; *M*=12.00, *SD*=5.72, *n*=7) than they did in their pretests (*M*=9.31, *SD*=4.45, *n*=19; *M*=7.85, *SD*=4.24, *n*=13; *M*=10.83, *SD*=4.42, *n*=7).

A *mixed factorial ANOVA* was computed on the raw scores for the total number of correct formulaic sequences the six groups produced in their graphical summaries at three time points in order to identify any statistically significant differences between the mean scores presented in Tables 18 and 19 above. An inspection of the *boxplots* and *z scores* generated to identify outliers revealed the presence of one outlying score in the pretest of the advanced focused instruction group (see Table E5 and Table E6). Since *mixed ANOVA* is sensitive to outliers (Tabachnick & Fidell, 2007a), the outlying score

was replaced with the second highest score within that variable, resulting in an adjustment for the maximum score presented in Table 19 above (*Max*=13). The results of normality tests, Mauchly's test, and Levene's test confirmed that the data were approximately normally distributed (see Table F3) and met the sphericity (*W*=.97, x^2=2.04, *df*=2, *p*=.361) and the homogeneity of variance [T1: *F*(5, 61)=.70, *p*=.627; T2: *F*(5, 61)=1.33, *p*=.262; T3: *F*(5, 61)=.61, *p*=.695] assumptions.

The *mixed factorial ANOVA* results presented in Table 20 below indicated that all the within-subjects effects, except for the time×group interaction (*p*>.05), were statistically significant. The statistically significant effects of time *F*(2, 122)=9.07, *p*=.001, partial η^2=.129, time×proficiency *F*(2, 122)=5.83, *p*=.004, partial η^2=.087, and time×group× proficiency *F*(4, 122)=3.45, *p*=.010, partial η^2=.102 respectively suggested that the six groups combined used different numbers of correct formulaic sequences at the three production stages, the different numbers of correct formulaic sequences were different for the upper-intermediate and advanced groups, and the different numbers of correct formulaic sequences generated at the three time points were different for different groups at different proficiency levels.

Table 20

Correct (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig	Partial Eta Sq.
Tm.	Sphericity As.	168.962	2	84.481	9.066*	.000	.129
Tm.*Gr.	Sphericity As.	49.221	4	12.305	1.321	.266	.041
Tm.* Pro.	Sphericity As.	108.701	2	54.351	5.833*	.004	.087
Tm.*Gr.*Pro.	Sphericity As.	128.540	4	32.135	3.449*	.010	.102
Error	Sphericity As.	1136.847	122	9.318			

*$p \leq .05$

The results summarized in Table 21 below showed that the only statistically significant between-subjects effect was that of proficiency *F*(1, 61)=17.43, *p*=.001,

partial η^2=.222, indicating that the overall performance of the upper-intermediate and advanced EAP students, irrespective of their groups, was significantly different. All other between-subjects effects were found to be statistically nonsignificant (p>.05).

Table 21

Correct (Task 1): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig	Partial Eta Sq.
Intercept	14422.458	1	14422.458	356.481	.000	.854
Groups	244.334	2	122.167	3.020	.056	.090
Proficiency	705.039	1	705.039	17.426*	.000	.222
Groups*Proficiency	.609	2	.305	.008	.993	.000
Error	2467.931	61	40.458			

*$p \leq .05$

Given the statistically significant highest order interaction, a *simple-effects analysis* was carried out after splitting the data based on proficiency in order to reanalyze the nonsignificant time by group interaction at each proficiency level. Mauchly's tests results confirmed that the data met the sphericity assumption at the upper-intermediate level (W=.88, x^2=2.99, df=2, p=.225) and at the advanced level (W=.96, x^2=1.49, df=2, p=.476). The results of Levene's test also confirmed that the data met the homogeneity of variance assumption at both the upper-intermediate [T1: F(2, 25)=1.38, p=.270; T2: F(2, 25)=2.15, p=.138; T3: F(2, 25)=.82, p=.452] and advanced [T1: F(2, 36)=.58, p=.568; T2: F(2, 36)=1.28, p=.291; T3: F(2, 36)=.58, p=.565] proficiency levels.

The *mixed ANOVA* results, presented in Table 22 below, revealed a statistically significant main effect of time at both the upper-intermediate F(2, 50)=3.70, p=.032, partial η^2=.129 and the advanced F(2, 72)=12.04, p=.001, partial η^2=.251 proficiency levels, which indicated that the three upper-intermediate groups combined and the three advanced groups combined utilized different numbers of correct formulaic sequences over time. Another statistically significant effect was the time×group interaction in the

advanced groups $F(4, 72)=4.70$, $p=.002$, partial $\eta^2=.207$ but not the three upper-intermediate groups ($p>.05$), suggesting that the different numbers of correct formulaic sequences at the three production stages were different for different advanced groups, which can also be seen in Figures 10 and 11 below. None of the between-subjects effects presented in Table 23 below was statistically significant ($p>.05$).

Table 22

Correct (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	55.967	2	27.983	3.695*	.032	.129
	Time*Group	12.036	4	3.009	.397	.810	.031
	Error	378.687	50	7.574			
Advanced	Time	253.488	2	126.744	12.036*	.000	.251
	Time*Group	197.883	4	49.471	4.698*	.002	.207
	Error	758.160	72	10.530			

*$p \leq .05$

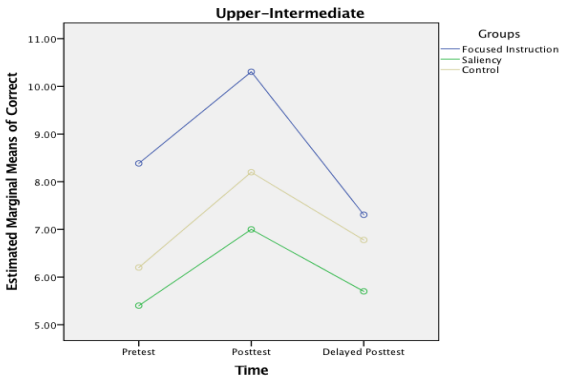


Figure 10. Correct task 1 inter.

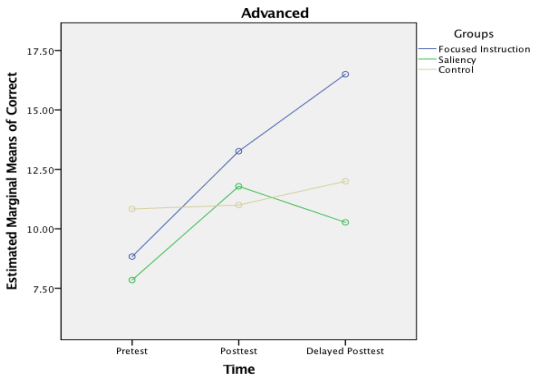


Figure 11. Correct task 1 adv.

Table 23

Correct (Task 1): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	3768.423	1	3768.423	94.653	.000	.791
	Groups	89.918	2	44.959	1.129	.339	.083
	Error	995.324	25	39.813			
Advanced	Intercept	12815.157	1	12815.157	313.285	.000	.897
	Groups	197.257	2	98.629	2.411	.104	.118
	Error	1472.607	36	40.906			

An inspection of the results of the simple comparisons for the three upper-intermediate groups combined indicated that none of differences in the mean scores for their pretests, posttests, and the delayed posttests approached statistical significance ($p>.05$) despite the statistically significant effect of time within these groups. An inspection of the simple comparisons produced as part of the *simple-effects analysis* for the three upper-intermediate groups, however, revealed a statistically significant difference between the posttest (M=10.31, SD=5.89, n=13) and the delayed posttest of the upper-intermediate focused instruction group (M=7.31, SD=4.77, n=13, MD=3, SE=1.10, p=.035), which may explain the statistically significant main effect of time in the three upper-intermediate groups.

The pairwise comparison results for the three advanced groups, on the other hand, revealed several statistically significant differences within and between the two advanced experimental groups. The differences between the mean scores for the focused instruction group's pretest (M=8.83, SD=3.38, n=19) and posttest (M=13.26, SD=5.54, n=19, MD=-4.43, SE=1.12, p=.001), pretest and delayed posttest (M=16.50, SD=5.13, n=19, MD=-7.67, SE=1.09, p=.001), and posttest and delayed posttest (MD=-3.24, SE=.94, p=.004) were all statistically significant. The advanced saliency group, in turn, used a statistically significant higher number of correct formulaic sequences in the posttest (M=11.79, SD=4.12, n=13) than they did in the pretest (M=7.85, SD=4.24, n=13, MD=3.94, SE=1.36, p=.018). Their use of correct formulaic sequences in the delayed posttest was not statistically significantly different from that in the pretest or the posttest ($p>.05$).

The multiple comparisons between the three advanced groups at each point in time indicated that the three advanced groups utilized a similar number of formulaic sequences in their pretests and posttests, but the number of correct formulaic sequences that were produced by the advanced focused instruction and saliency groups differed significantly at the delayed posttest stage; that is, the advanced focused instruction group produced a statistically significantly higher number of correct formulaic sequences in their delayed posttest (M=16.50, SD=5.13, n=19) than did the advanced saliency group (M=10.27, SD=3.76, n=13, MD=6.23, SE=1.74, p=.003). All other differences between the three advanced groups and within the advanced control group were not found to be statistically significant (p>.05).

In brief, neither presenting formulaic sequences through a focused instructional or a saliency approach nor focusing on syntactic aspects increased the three upper-intermediate groups' accurate production of formulaic sequences. Different instructional methods affected the three advanced groups differently, however. Explicitly teaching formulaic sequences helped the advanced focused instruction group produce more correct formulaic sequences at both production stages after the training period. It also helped them outperform the advanced saliency group at the delayed posttest stage. Making formulaic sequences salient in input, in turn, affected the advanced saliency group's immediate learning gains. The advanced control group's use of correct formulaic sequences remained relatively stable at the three production stages.

Wrong (task 1). *Descriptive statistics* were carried out on the raw scores for the total number of incorrect formulaic sequences the six groups produced when summarizing graphical information at different time points. The mean scores for the

upper-intermediate focused instruction and the upper-intermediate control groups' posttests (*M*=5.08, *SD*=2.29, *n*=13 and *M*=4.60, *SD*=2.80, *n*=10) and delayed posttests (*M*=4.38, *SD*=3.25, *n*=13 and *M*=3.25, *SD*=2.58, *n*=10) were lower than those for their pretests (*M*=5.54, *SD*=3.86, *n*=13 and *M*=5, *SD*=2.40, *n*=10), as shown in Table 24 below. The mean scores for the upper-intermediate saliency group, on the other hand, increased constantly from the pretest (*M*=5.40, *SD*=4.51, *n*=5) to the posttest (*M*=7.80, *SD*=4.02, *n*=5), and then to the delayed posttest (*M*=8.10, *SD*=2.36, *n*=5).

Table 24

Descriptive Statistics: Wrong (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	5.54	5.08	4.38	5.40	7.80	8.10	5.00	4.60	3.25
Median	5.00	5.00	4.00	4.00	10.00	9.00	5.00	5.00	3.45
Mode	3.00[a]	5.00[a]	3.00[a]	0[a]	10.00	9.00	7.00	3.00[a]	0
SD	3.86	2.29	3.25	4.51	4.02	2.36	2.40	2.80	2.58
Minimum	1.00	2.00	0	0	3.00	5.00	0	0	0
Maximum	16.00	11.00	12.00	11.00	12.00	11.00	8.00	8.00	7.00

a. Multiple modes exist. The smallest value is shown

The mean scores for the incorrect formulaic sequences the three advanced groups produced were also different at the three production stages (see Table 25 below).

Table 25

Descriptive Statistics: Wrong (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	5.31	4.89	4.65	5.00	5.77	4.75	4.50	3.86	7.14
Median	4.00	4.00	3.28	4.00	5.00	4.00	5.00	4.00	6.00
Mode	2.00[a]	3.00[a]	2.00	1.00[a]	2.00[a]	4.00	5.00	0[a]	6.00[a]
SD	4.08	3.13	3.20	3.14	3.06	3.12	2.78	2.41	4.49
Minimum	1.00	0	1.00	1.00	2.00	0	-.50	0	1.00
Maximum	16.00	12.00	12.00	10.00	10.04	12.00	8.00	7.00	14.00

a. Multiple modes exist. The smallest value is shown

Similar to the upper-intermediate focused instruction group, the advanced focused instruction group produced fewer incorrect formulaic sequences in the posttest (*M*=4.89, *SD*=3.13, *n*=19) and the delayed posttest (*M*=4.65, *SD*=3.20, *n*=19) than they did in the pretest (*M*=5.31, *SD*=4.08, *n*=19). The advanced saliency and control groups, in turn, respectively had the highest number of incorrect formulaic sequences at the posttest stage (*M*=5.77, *SD*=3.06, *n*=13) and the delayed posttest stage (*M*=7.14, *SD*=4.49, *n*=7).

It is worth mentioning here that an inspection of the mean scores for the types of errors made by the six groups (see Table 26 and Table 27 below) indicated that the majority of incorrect formulaic sequences made by the six groups had structural errors (e.g. subject/verb agreement, tenses, word form, etc.), with very few usage or spelling errors at the three production stages. The upper-intermediate saliency group was the only group that produced more incorrectly formed formulaic sequences due to spelling errors than structural errors at the posttest stage, as shown in Table 26 below.

Table 26

Mean Scores: Types of Errors (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Use	.38	.38	.08	0	1.00	.25	.50	.40	.14
Structure	4.69	4.31	3.77	4.60	2.60	6.60	4.00	3.90	2.77
Spelling	.46	.38	.54	.80	4.20	1.25	.50	.30	.33

Table 27

Mean Scores: Types of Errors (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Use	.37	.37	.24	.85	.36	.45	-.06	.29	.43
Structure	4.27	3.63	3.07	3.69	4.38	3.48	4.29	3.43	6.14
Spelling	.67	.89	1.35	.46	1.04	.82	.27	.14	.57

A *mixed factorial ANOVA* was also computed on the raw scores for the number of incorrect formulaic sequences to determine whether the mean differences presented in Tables 24 and 25 were statistically significant. The results of the Mauchly's test (W=.99, x^2=.43, df= 2, p=.808), Levene's test [T1: F(5, 61)=1.04, p=.405; T2: F(5, 61)=1.62, p=.167; T3: F(5, 61)=.76, p=.579], and normality tests (see Table F4) confirmed that the data met the sphericity and homogeneity of variance assumptions and the sum of scores within each of the six groups was approximately normally distributed. These variables were also checked for outliers, and the preliminary screening revealed the presence of three outlying scores in the pretest and posttest variables for the upper-intermediate focused instruction group and the pretest variable for the advanced focused instruction group (see Table E7 and Table E8). The outliers in the pretest variables were replaced with the next highest number of incorrect formulaic sequences within each group, which resulted in adjusting the maximum score in Table 24 (*Max*=9) and Table 25 (*Max*=14) above. The third outlier, on the other hand, was not altered because it only exceeded the *z score* threshold (i.e., +/-2.58) by .01, and, therefore, it was not expected to affect the analysis.

According to Table 28 and Table 29, none of the within-subjects or the between-subjects effects was statistically significant (p>.05), which indicated that the differences in the mean scores presented in Tables 24 and 25 above and graphically plotted in Figure 12 and Figure 13 below were not large enough to reach statistical significance. These results demonstrated that the six groups' production of inaccurate formulaic sequences was not only similar but also remained relatively stable over time despite being involved in different types of pedagogical interventions.

Table 28

Wrong (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	4.319	2	2.159	.297	.744	.005
Tm.*Gr.	Sphericity As.	32.548	4	8.137	1.119	.351	.035
Tm.* Pro.	Sphericity As.	11.082	2	5.541	.762	.469	.012
Tm.*Gr.*Pro.	Sphericity As.	62.558	4	15.640	2.150	.079	.066
Error	Sphericity As.	887.397	122	7.274			

Table 29

Wrong (Task 1): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	4572.559	1	4572.559	309.257	.000	.835
Groups	58.412	2	29.206	1.975	.147	.061
Proficiency	4.124	1	4.124	.279	.599	.005
Groups*Proficiency	48.790	2	24.395	1.650	.201	.051
Error	901.924	61	14.786			

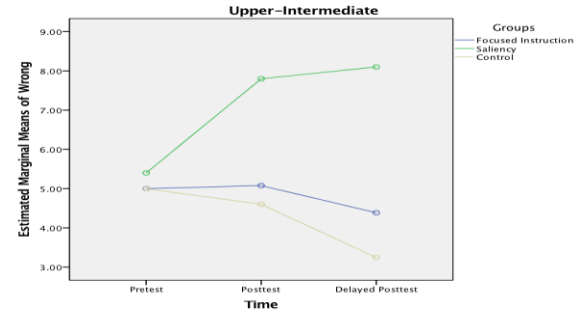


Figure 12. Wrong task 1 inter.

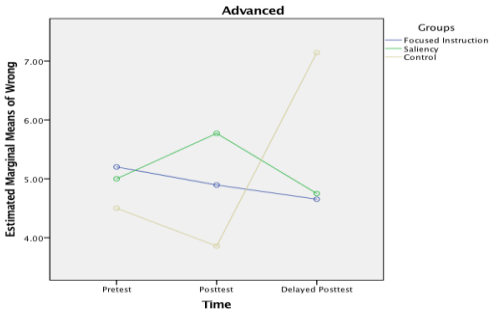


Figure 13. Wrong task 1 adv.

Correct% (task 1). To summarize the overall percentage of the correct formulaic sequences the six groups produced in the summary task, *measures of central tendency* and *measures of variability* were run on these variables. As shown in Table 30 below, the mean scores for the percentage of correct formulaic sequences that were utilized by the upper-intermediate focused instruction group at the three production stages were very similar, with the lowest percentage at the delayed posttest stage (*M*=59.13, *SD*=23.45, *n*=13). On the other hand, the percentage of correct formulaic sequences decreased gradually over time in the upper-intermediate saliency group (T1: *M*=56.95, *SD*=30.49,

n=5; T2: *M*=46.89, *SD*=23.70, *n*=5; T3: *M*=38.47, *SD*=13.46, *n*=5) but increased constantly in the upper-intermediate control group (T1: *M*=35.71, *SD*=28.50, *n*=10; T2: *M*=64.30, *SD*=24.05, *n*=10; T3: *M*=70.02, *SD*=27.54, *n*=10).

Table 30

Descriptive Statistics: Correct% (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	61.59	62.66	59.13	56.95	46.89	38.47	53.71	64.30	70.02
Median	61.53	64.70	60.00	50.00	40.00	35.71	55.77	67.71	70.00
Mode	61.53	40.00	66.66	21.42[a]	16.66[a]	35.71	12.50[a]	76.92	100.00
SD	19.13	19.80	23.45	30.49	23.70	13.46	28.50	24.05	27.54
Minimum	15.78	31.25	0	21.42	16.66	18.75	12.50	27.27	12.32
Maximum	93.33	92.00	100.00	100.00	77.77	52.17	100.00	100.00	100.00

a. Multiple modes exist. The smallest value is shown

The three advanced groups, as shown in Table 31 below, had different trends from those for the three upper-intermediate groups. While both the advanced focused instruction and saliency groups had higher percentages of correct formulaic sequences in their posttests (*M*=72.93, *SD*=17.05, *n*=19 and *M*=67.90, *SD*=10.98, *n*=13) and delayed posttests (*M*=77.94, *SD*=14.18, *n*=19 and *M*=70.17, *SD*=14.33, *n*=13) than those in their pretests (*M*=63.62, *SD*=21.95, *n*=19 and *M*=61.50, *SD*=21.54, *n*=13), the advanced

Table 31

Descriptive Statistics: Correct% (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	63.62	72.93	77.94	61.50	67.90	70.17	69.27	73.26	61.99
Median	66.66	75.00	80.38	66.66	66.66	75.00	72.72	68.42	60.00
Mode	20.00[a]	72.72[a]	90.90	66.66	47.36[a]	75.00	33.33[a]	58.33[a]	33.33
SD	21.95	17.05	14.18	21.54	10.98	14.33	20.89	14.47	23.98
Minimum	20.00	35.71	44.44	10.00	47.36	42.85	33.33	58.33	33.33
Maximum	91.66	100.00	95.23	87.50	85.71	100.00	103.00	100.00	93.33

a. Multiple modes exist. The smallest value is shown

control group had the highest percentage of correct formulaic sequences at the posttest stage (*M*=73.26, *SD*=14.47, *n*=7) and the lowest one at the delayed posttest stage (*M*=61.99, *SD*=23.98, *n*=7).

A *mixed factorial ANOVA* was computed on the same variables to identify any significant differences between the mean scores summarized in Table 30 and Table 31 above after screening these variables for outliers, assessing their distribution, and checking the assumptions of sphericity and homogeneity of variance. The *boxplots* and *z scores* generated to identify outliers within each of the six groups at the three production stages showed no significant outliers (see Tables E9 and E10). The skewness and kurtosis *z scores* and the *Shapiro-Wilk* test results confirmed that the distribution of scores within the summative variables for almost all the groups was approximately normal (see Table F5). Although the scores within the upper-intermediate saliency group were slightly skewed and kurtotic and had statistically significant *Shapiro-Wilk* results ($p<.01$), no data transformation was performed on these variables because of the robustness of *mixed ANOVA* to slight departures from normality (Tabachnick & Fidell, 2001, 2007a). The results of Mauchly's test ($W=.95$, $x^2=3.03$, $df=2$, $p=.220$) and Levene's test [T1: $F(5, 61)=.83$, $p=.536$; T2: $F(5, 61)=1.62$, $p=.169$; T3: $F(5, 61)=1.21$, $p=.316$] confirmed that the data met the sphericity and homogeneity of variance assumptions.

The results of the *mixed factorial ANOVA*, displayed in Tables 32 and 33 below, revealed two statistically significant effects, that is, the effects of the three-way interaction $F(4, 122)=2.90$, $p=.025$, partial $\eta^2=.087$ which suggested that different groups at different proficiency levels produced different percentages of correct formulaic at the three production stages and the effect of proficiency level $F(1, 61)=7.86$, $p=.007$, partial

η^2=.114, which indicated that the overall percentage of correct formulaic sequences was different for the EAP students at different proficiency levels. All other within- and between-subjects effects did not approach statistical significance (*p*>.05).

Table 32

Correct% (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	349.555	2	174.777	.661	.518	.011
Tm.*Gr.	Sphericity As.	708.508	4	177.127	.670	.614	.021
Tm.* Pro.	Sphericity As.	382.754	2	191.377	.724	.487	.012
Tm.*Gr.*Pro.	Sphericity As.	3066.278	4	766.570	2.899*	.025	.087
Error	Sphericity As.	32260.237	122	264.428			

*$p \leq .05$

Table 33

Correct% (Task 1): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	658174.010	1	658174.010	916.528	.000	.938
Groups	2729.963	2	1364.981	1.901	.158	.059
Proficiency	5647.060	1	5647.060	7.864*	.007	.114
Groups*Proficiency	1090.859	2	545.429	.760	.472	.024
Error	43805.124	61	718.117			

*$p \leq .05$

Because of the statistically significant three-way interaction and the significant main effect of the proficiency level, a *simple-effects analysis* was computed after splitting the data based on proficiency. The data, after being split, met the sphericity assumption at the upper-intermediate (*W*=.97, x^2=.69, *df*=2, *p*=.710) and the advanced (*W*=.92, x^2=2.81, *df*=2, *p*=.246) proficiency levels. The results of Levene's tests also confirmed that the homogeneity of variance assumption was met at the upper-intermediate [T1: *F*(2, 25)=1.56, *p*=.229; T2: *F*(2, 25)=.21, *p*=.815; T3: *F*(2, 25)=.64, *p*=.534] and the advanced [T1: *F*(2, 36)=.16, *p*=.849; T2: *F*(2, 36)=.58, *p*=.566; T3: *F*(2, 36)=2.20, *p*=.126] proficiency levels.

The *mixed ANOVA* results summarized in Tables 34 and 35 below showed that none of the mean differences presented in Table 30 and Table 31 above and graphically plotted in Figures 14 and 15 below were statistically significant, which was demonstrated by the statistically nonsignificant within-subjects and between-subjects effects (p>.05).

Table 34

Correct% (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	55.519	2	27.759	.090	.914	.004
	Time*Group	2224.074	4	556.019	1.811	.141	.127
	Error	15348.169	50	306.963			
Advanced	Time	795.966	2	397.983	1.694	.191	.045
	Time*Group	1309.763	4	327.441	1.394	.245	.072
	Error	16912.068	72	234.890			

Table 35

Correct% (Task 1): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	233382.227	1	233382.227	221.155	.000	.898
	Groups	2583.480	2	1291.740	1.224	.311	.089
	Error	26382.202	25	1055.288			
Advanced	Intercept	468239.407	1	468239.407	967.497	.000	.964
	Groups	602.034	2	301.017	.622	.543	.033
	Error	17422.922	36	483.970			

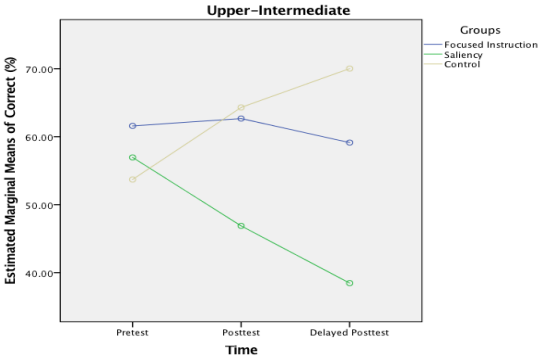


Figure 14. Correct% task 1 inter.

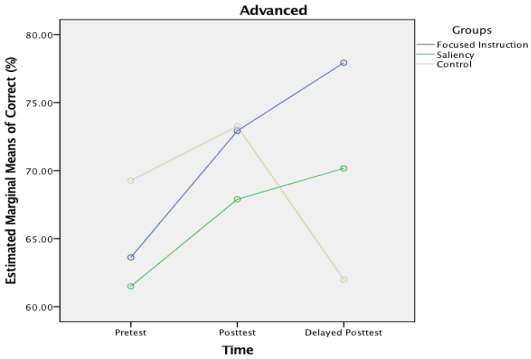


Figure 15. Correct task 1 adv.

A inspection of the pairwise comparisons results, however, revealed that the three-way significant interaction identified in Table 32 was due to the statistically significant

difference in the mean scores for the advanced focused instruction group's delayed posttest (*M*=77.94, *SD*=14.18, *n*=19) when compared to their pretest (*M*=63.62, *SD*=21.95, *n*=19, *MD*=14.32, *SE*=5.52, *p*=.041), a difference that can be seen in Figure 15 above. All other differences within each of the six groups or between the three upper-intermediate and three advanced groups at each time point were statistically nonsignificant (*p*>.05). Based on the results proposed above, the three presentation methods had no effects on the overall accuracy of the formulaic sequences produced by the three upper-intermediate groups or the advanced saliency or control groups. Focused instruction of formulaic sequences, on the other hand, positively affected the advanced focused instruction group's accurate production of formulaic sequences one month after the training period.

Frequency (task 2). *Descriptive statistics* were run on the raw scores for the total number of formulaic sequences the six groups utilized in their essay responses at the three production stages. The results presented in Table 36 below indicated that both the upper-intermediate saliency group and the upper-intermediate control group used more formulaic sequences in their posttests (*M*=14.60, *SD*=7.54, *n*=5 and *M*=8.10, *SD*=3.57, *n*=10) and delayed posttests (*M*=12.78, *SD*=5.10, *n*=5 and *M*=8.01, *SD*=2.84, *n*=10) than they did in their pretests (*M*=11.20, *SD*=4.87, *n*=5 and *M*=6.80, *SD*=2.74, *n*=10). The upper-intermediate focused instruction group, in turn, produced the highest number of formulaic sequences in their posttest (*M*=9.46, *SD*=7.74, *n*=13) and the lowest number of formulaic sequences in their delayed posttest (*M*=8.15, *SD*=4.04, *n*=13), which was very similar to the one they utilized in their pretest (*M*=8.46, *SD*=6.45, *n*=13).

Table 36

Descriptive Statistics: Frequency (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	8.46	9.46	8.15	11.20	14.60	12.78	6.80	8.10	8.01
Median	8.00	9.00	9.00	11.00	13.00	12.00	6.50	7.50	8.20
Mode	5.00	2.00[a]	9.00	16.00	7.00[a]	7.88[a]	4.00	5.00[a]	5.00[a]
SD	6.45	7.74	4.04	4.87	7.54	5.10	2.74	3.57	2.84
Minimum	0	2.00	1.00	5.00	7.00	7.88	4.00	4.00	4.00
Maximum	26.00	31.00	16.00	16.00	27.00	19.00	12.00	15.00	12.00

a. Multiple modes exist. The smallest value is shown

The three advanced groups had different patterns, as shown in Table 37 below. While the advanced focused instruction group and the advanced control group used more formulaic sequences in their posttests (*M*=13.89, *SD*=7.15, *n*=19 and *M*=13.57, *SD*=4.12, *n*=7) and delayed posttests (*M*=12.48, *SD*=5.48, *n*=19 and *M*=12, *SD*=7.85, *n*=7) than they did in their pretests (*M*=9.53, *SD*=5.36, *n*=19 and *M*=9.79, *SD*=6.30, *n*=7), the advanced saliency group utilized very similar numbers of formulaic sequences in their pretest (*M*=10.69, *SD*=5.69, *n*=13) and posttest (*M*=10.61, *SD*=4.86, *n*=13) and the lowest number in their delayed posttest (*M*=8.46, *SD*=4.16, *n*=13).

Table 37

Descriptive Statistics: Frequency (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	9.53	13.89	12.48	10.69	10.61	8.46	9.79	13.57	12.00
Median	9.00	13.00	12.00	10.00	9.00	8.00	9.00	14.00	11.00
Mode	9.00	5.00[a]	19.00	8.00[a]	6.00[a]	7.00	3.00[a]	14.00	1.00[a]
SD	5.36	7.15	5.48	5.69	4.86	4.16	6.30	4.12	7.85
Minimum	3.00	5.00	3.00	2.00	4.92	0	3.00	6.00	1.00
Maximum	25.00	33.00	21.00	21.00	22.00	14.00	22.00	19.00	25.00

a. Multiple modes exist. The smallest value is shown

In order to determine whether the mean differences presented in Table 36 and Table 37 above were statistically significant, a *mixed factorial ANOVA* was performed on

the frequency variables for the second task. Prior to running the *mixed factorial ANOVA*, the maximum scores for the pretest and posttest of the upper-intermediate and advanced focused instruction groups were altered because they were outliers (see Tables E11 and E12). Altering these four scores resulted in adjusting the maximum pretest (*Max*=13) and posttest (*Max*=15) scores of the upper-intermediate focused instruction group and the pretest (*Max*=16) and posttest (*Max*=24) maximum scores of the advanced focused instruction group. The results of normality tests (see Table F6), Mauchly's test (W=1.00, x^2=.29, df= 2, p=.865), and Levene's test [T: F(5, 61)=1.19, p=.327; T2: F(5, 61)=.85, p=.519; T3: F(5, 61)= 2.30, p=.056] were also checked, and they confirmed that the distribution of the summative scores within each group was approximately normal, and the data met the assumptions of sphericity and homogeneity of variance.

According to Table 38, the only statistically significant within-subjects effect was the effect of time F(2, 122)=4.10, p=.019, partial η^2=.063, which suggested that the six groups combined used different numbers of formulaic sequences at the three different production stages. An inspection of the pairwise comparisons results indicated that the significant effect of time was due to the statistically significantly higher mean scores for

Table 38

Frequency (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	141.189	2	70.594	4.097*	.019	.063
Tm.*Gr.	Sphericity As.	29.591	4	7.398	.429	.787	.014
Tm.* Pro.	Sphericity As.	7.147	2	3.573	.207	.813	.003
Tm.*Gr.*Pro.	Sphericity As.	78.506	4	19.626	1.139	.341	.036
Error	Sphericity As.	2102.395	122	17.233			

*$p \leq .05$

the posttest of the six groups combined (M=11.42, SE=.70, n=67) when compared to their pretest (M=9.17, SE=.62, n=67, MD=2.26, SE=.82, p=.022). The differences between the

mean scores for the six groups' pretest and delayed posttest as well as their posttest and delayed posttest were not statistically significant (p>.05).

The results of the between-subjects effects, shown in Table 39 below, revealed a statistically significant group×proficiency interaction $F(2, 61)=5.02$, $p=.010$, partial $\eta^2=.141$, which suggested that different instructional methods and proficiency levels had moderating effects on the overall performance of the 67 EAP students. The visual representation of the six groups' mean scores in Figure 16 and Figure 17 below suggested that the statistically significant group by proficiency interaction might be due to the upper-intermediate saliency group's higher overall mean score than those of their peers in the focused instruction and control groups.

Table 39

Frequency (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	17646.060	1	17646.060	460.609	.000	.883
Groups	87.502	2	43.751	1.142	.326	.036
Proficiency	111.497	1	111.497	2.910	.093	.046
Groups*Proficiency	384.844	2	192.422	5.023*	.010	.141
Error	2336.927	61	38.310			

*$p \leq .05$

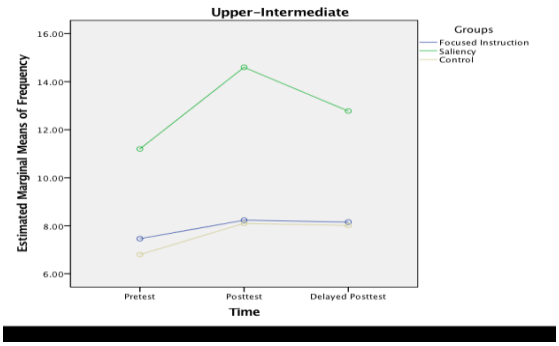


Figure 16. Frequency task 2 inter.

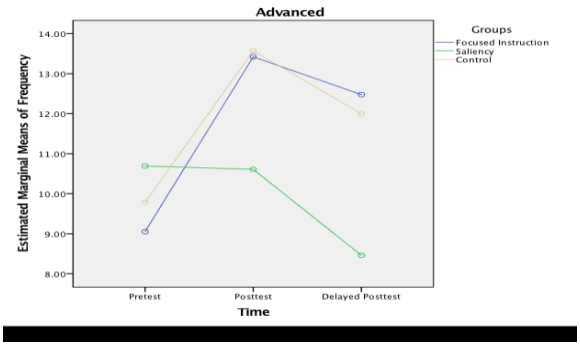


Figure 17. Frequency task 2 adv.

A *simple-effects analysis* was, therefore, computed after splitting the data based on proficiency to determine which groups at which proficiency levels differed

significantly from one another. The results of Mauchly's tests confirmed that the data met the sphericity assumption at the upper-intermediate level (*W*=.98, x^2=.47, *df*=2, *p*=.789) and the advanced level (*W*=.99, x^2=.27, *df*=2, *p*=.876). The results of Levene's tests, in turn, demonstrated that the variance of scores at the upper-intermediate level [T1: *F*(2, 25)=1.53, *p*=.235; T2: *F*(2, 25)=1.23, *p*=.308; T3: *F*(2, 25)=1.39, *p*=.269] and the advanced level [T1: *F*(2, 36)=.94, *p*=.402; T2: *F*(2, 36)=.89, *p*=.419; T3: *F*(2, 36)=1.99, *p*=.151] was homogeneous.

The results of the *simple-effects analysis*, summarized in Table 40 and Table 41 below, indicated that the only statistically significant main effect was that of groups at the upper-intermediate level *F*(2, 25)=5.35, *p*=.012, partial η^2= .300. The significant effect of groups at upper-intermediate level was, as expected from Figure 16 above, due to the statistically significantly higher mean score for the overall number of formulaic sequences the upper-intermediate saliency group used (*M*=12.86, *SE*=1.40, *n*=5) when compared to their counterparts in the focused instruction group (*M*=7.95, *SE*=.87, *n*=13, *MD*=4.91, *SE*=1.65, *p*=.017) and in the control group (*M*=7.64, *SE*=.99, *n*=10, *MD*=5.22, *SE*=1.72, *p*=.015). All other within- and between-subjects effects, together with the pairwise comparisons results, were statistically nonsignificant (*p*>.05).

Table 40

Frequency (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	40.661	2	20.330	1.680	.197	.063
	Time*Group	13.670	4	3.418	.282	.888	.022
	Error	605.071	50	12.101			
Advanced	Time	120.530	2	60.265	2.898	.062	.074
	Time*Group	142.167	4	35.542	1.709	.157	.087
	Error	1497.324	72	20.796			

Table 41

Frequency (Task 2): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	6439.640	1	6439.640	218.319	.000	.897
	Groups	315.312	2	157.656	5.345*	.012	.300
	Error	737.413	25	29.497			
Advanced	Intercept	12253.696	1	12253.696	275.792	.000	.885
	Groups	81.337	2	40.668	.915	.410	.048
	Error	1599.514	36	44.431			

*$p \leq .05$

Because of the apparent difference between the pretest and posttest scores of the advanced focused instruction and control groups and between the delayed posttest scores of these group and that of the saliency group (see Figure 17 above), the pairwise comparisons results that were produced as part of the *simple-effects analysis* were also checked in order to detect any significant differences within and between the three advanced groups. The pairwise comparisons results revealed a statistically significant difference between the mean pretest (M=9.05, SD=4.18, n=19) and posttest (M=13.42, SD=6.02, n=19, MD=-4.37, SE=1.54, p=.022) scores of the advanced focused instruction group. Although the pretest and posttest mean scores of the advanced control group seemed to be very similar to those of the focused instruction groups (see Figure 17 above), the differences between the mean scores of the former did not achieve statistical significance (p>.05). Nor did any of the other differences within the three upper-intermediate and the three advanced groups or between the three upper-intermediate groups and the three advanced groups at each stage of data collection (p>.05).

Based on these results, explicitly teaching formulaic sequences and making them salient respectively helped the advanced focused instruction group use higher numbers of formulaic sequences after the training period and helped the upper-intermediate saliency group outperform the upper-intermediate focused instruction and control groups in their

overall use of formulaic sequences. On the other hand, the performance of the upper-intermediate and advanced control groups remained relatively stable.

Occurrence (task 2). *Descriptive statistics* were also calculated for the distinct types used by the six groups in their essay responses at the three production stages. According to Table 42, the upper-intermediate focused instruction group had very similar mean scores for the distinct types they utilized in the pretest (*M*=5.38, *SD*=3.04, *n*=13) and the posttest (*M*=5.31, *SD*=2.43, *n*=13), which then dropped in the delayed posttest (*M*=4.85, *SD*=2.44, *n*=13). The mean scores for the distinct types utilized by the upper-intermediate control group in the pretest (*M*=5.20, *SD*=1.48, *n*=10), the posttest (*M*=5.60, *SD*=2.63, *n*=10), and the delayed posttest (*M*=5.52, *SD*=1.45, *n*=10) were also similar. The upper-intermediate saliency group, on the other hand, had higher mean scores for the number of distinct types they used in the posttest (*M*=8.80, *SD*=3.90, *n*=5) and the delayed posttest (*M*=8.58, *SD*=3.08, *n*=5) when compared to their pretest mean score (*M*=6.60, *SD*=2.41, *n*=5).

Table 42

Descriptive Statistics: Occurrence (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	5.38	5.31	4.85	6.60	8.80	8.58	5.20	5.60	5.52
Median	5.00	6.00	6.00	8.00	9.00	9.00	4.50	5.50	5.84
Mode	5.00	2.00[a]	6.00	4.00[a]	9.00	5.00[a]	4.00	6.00	6.00
SD	3.04	2.43	2.44	2.41	3.90	3.08	1.48	2.63	1.45
Minimum	0	2.00	1.00	4.00	5.00	5.00	4.00	3.00	3.00
Maximum	12.00	8.00	9.00	9.00	15.00	12.00	8.00	12.00	8.00

a. Multiple modes exist. The smallest value is shown

As for the three advanced groups, Table 43 below shows that the focused instruction and control groups had higher mean scores for their posttests (*M*=8.58, *SD*=3.82, *n*=19 and *M*=8.29, *SD*=2.93, *n*=7) and delayed posttests (*M*=7.53, *SD*=2.60,

n=19 and *M*=6.57, *SD*=3.55, *n*=7) when compared to their pretests (*M*=5.58, *SD*=2.71, *n*=19 and *M*=5.85, *SD*=3.86, *n*=7). The advanced saliency group, in turn, had the highest mean score at the posttest stage (*M*=7.84, *SD*=2.62, *n*=13) and the lowest mean score at the delayed posttest stage (*M*=5.62, *SD*=3.36, *n*=13).

Table 43

Descriptive Statistics: Occurrence (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	5.58	8.58	7.53	5.69	7.84	5.62	5.85	8.29	6.57
Median	5.00	8.00	7.00	6.00	7.00	6.00	5.00	8.00	8.00
Mode	4.00[a]	5.00[a]	11.00	3.00[a]	5.00[a]	9.00	5.00	8.00	4.00[a]
SD	2.71	3.82	2.60	2.72	2.62	3.36	3.86	2.93	3.55
Minimum	2.00	4.00	2.83	2.00	4.92	0	2.92	5.00	1.00
Maximum	14.00	17.00	11.00	11.00	13.00	10.00	14.00	13.00	10.00

a. Multiple modes exist. The smallest value is shown

To identify any statistically significant differences between the mean scores for the distinct types used by the six groups in their essay responses, a *mixed factorial ANOVA* was computed on the occurrence variables after screening these variables for outliers, assessing their distribution, and checking two major assumptions for *mixed factorial ANOVAs*. The initial screening revealed the presence of one outlying score within the advanced focused instruction group's pretest variable (see Table E13 and Table E14). This outlier was replaced with the second highest number of distinct types within that variable, which resulted in adjusting the maximum number of distinct types for the advanced focused instruction group's pretest (*Max*=8). The normality tests, in turn, demonstrated that the distribution of scores within each summative variable was approximately normal (see Table F7). The results of Mauchly's test (*W*=.98, x^2=.94, *df*=2, *p*=.625) and Levene's test [T1: *F*(5, 61)=1.15, *p*=.346; T2: *F*(5, 61)=.85, *p*=.517; T3: *F*(5, 61)=2.50, *p*=.040] confirmed that the data met the sphericity assumption but

moderately violated the assumption of homogeneity of variance; a more conservative alpha level of $p \leq .025$ was, therefore, utilized in the interpretation of the results.

The results of the *mixed factorial ANOVA*, shown in Table 44 and Table 45 below, revealed a statistically significant effect of time $F(2, 122)=7.00$, $p=.001$, partial $\eta^2=.103$ and group×proficiency interaction $F(2, 61)=4.13$, $p=.021$, partial $\eta^2=.119$. Similar to the frequency of the target formulaic sequences, the statistically significant effect of time was due to the higher numbers of the distinct types the six groups combined utilized in their posttest ($M=7.40$, $SE=.42$, $n=67$) when compared to their pretest ($M=5.66$, $SE=.34$, $n=67$, $MD=1.74$, $SE=.48$, $p=.002$). Figure 18 and Figure 19 below suggested that the statistically significant group by proficiency interaction might be the result of the higher overall mean scores for the upper-intermediate saliency group than those for the upper-intermediate focused instruction and control groups.

Table 44

Occurrence (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	84.008	2	42.004	7.000*	.001	.103
Tm.*Gr.	Sphericity As.	2.856	4	.714	.119	.976	.004
Tm.* Pro.	Sphericity As.	24.712	2	12.356	2.059	.132	.033
Tm.*Gr.*Pro.	Sphericity As.	31.813	4	7.953	1.325	.264	.042
Error	Sphericity As.	732.047	122	6.000			

*$p \leq .025$

Table 45

Occurrence (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	7034.341	1	7034.341	598.926	.000	.908
Groups	35.221	2	17.611	1.499	.231	.047
Proficiency	14.884	1	14.884	1.267	.265	.020
Groups*Proficiency	96.915	2	48.458	4.126*	.021	.119
Error	716.441	61	11.745			

*$p \leq .025$

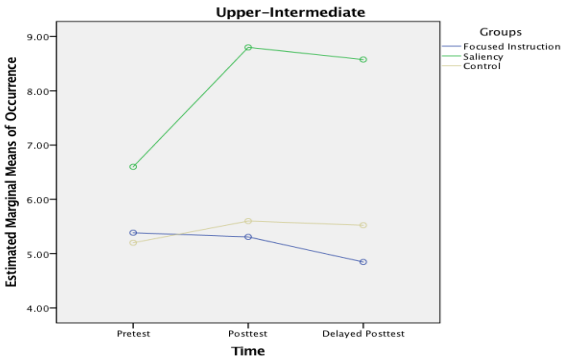


Figure 18. Occurrence task 2 inter.

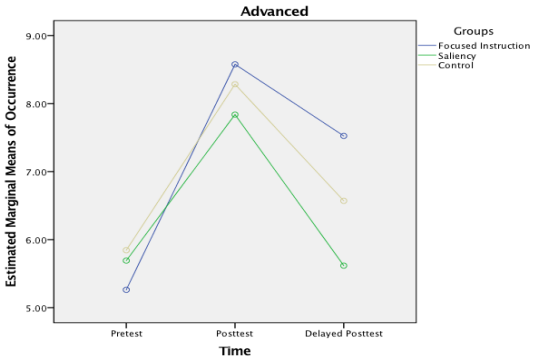


Figure 19. Occurrence task 2 adv.

A *simple-effects analysis* was computed after splitting the data based on proficiency in order to identify any statistically significant differences between different groups at different proficiency levels. A more liberal alpha level ($p \leq .05$) was used to interpret the results because the data met the homogeneity of variance assumption at the upper-intermediate [T1: $F(2, 25)=1.38$, $p=.269$; T2: $F(2, 25)=.42$, $p=.663$; T3: $F(2, 25)=3.07$, $p=.064$] and advanced [T1: $F(2, 36)=1.46$, $p=.246$; T2: $F(2, 36)=1.07$, $p=.353$; T3: $F(2, 36)=1.28$, $p=.291$] levels. The data also met the sphericity assumption at both proficiency levels: upper-intermediate ($W=.98$, $x^2=.53$, $df=2$, $p=.766$) and advanced ($W=.98$, $x^2=.74$, $df=2$, $p=.690$).

The *simple-effects analysis* results revealed a statistically significant main effect of groups at the upper-intermediate level $F(2, 25)=4.56$, $p=.021$, partial $\eta^2= .267$ and a statistically significant main effect of time at the advanced level $F(2, 72)=8.36$, $p=.001$, partial $\eta^2= .188$ (see Table 46 and Table 47 below). The statistically significant effect of groups, as expected from Figure 18 above, was the result of the higher overall mean score for the upper-intermediate saliency group ($M=7.99$, $SE=.82$, $n=5$) than those for their peers in the focused instruction group ($M=5.18$, $SE=.51$, $n=13$, $MD=2.81$, $SE=.96$, $p=.019$) and the control group ($M=5.44$, $SE=.58$, $n=10$, $MD=2.55$, $SE=1.00$, $p=.044$).

Table 46

Occurrence (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	8.885	2	4.442	.979	.383	.038
	Time*Group	14.245	4	3.561	.785	.541	.059
	Error	226.954	50	4.539			
Advanced	Time	117.282	2	58.641	8.359*	.001	.188
	Time*Group	22.115	4	5.529	.788	.537	.042
	Error	505.093	72	7.015			

*$p \leq .05$

Table 47

Occurrence (Task 2): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	2757.258	1	2757.258	276.470	.000	.917
	Groups	90.904	2	45.452	4.557*	.021	.267
	Error	249.327	25	9.973			
Advanced	Intercept	4586.364	1	4586.364	353.467	.000	.908
	Groups	12.760	2	6.380	.492	.616	.027
	Error	467.114	36	12.975			

*$p \leq .05$

The statistically significant main effect of time, in turn, was due to the higher overall mean score that the three advanced groups combined had for their posttest (M=8.24, SE=.58, n=39) than the one they had for their pretest (M=5.60, SE=.45, n=39, MD=2.63, SE=.66, p=.001).

Given the patent differences between the three upper-intermediate groups at each time point and within each of the three advanced groups (see Figures 18 and 19 above), the results of the pairwise comparisons were also of interest. These results confirmed that the upper-intermediate saliency group utilized statistically significantly more distinct types in their delayed posttest (M=8.58, SD=3.08, n=5) than the focused instruction group (M=4.85, SD=2.44, n=13, MD=3.73, SE=1.19, p=.013). The results also confirmed that the differences between the mean scores of the advanced focused instruction group's

pretest (*M*=5.26, *SD*=1.91, *n*=19) and posttest (*M*=8.58, *SD*=3.82, *n*=19, *MD*=-3.32, *SE*=.87, *p*=.002) and their pretest and delayed posttest (*M*=7.53, *SD*=2.60, *n*=19, *MD*=-2.26, *SE*=.80, *p*=.022) were statistically significant. None of the other differences between and/or within the three upper-intermediate and the three advanced groups approached statistical significance (*p*>.05).

These results indicated that focused instruction of formulaic sequences helped the advanced students acquire and retain a wider range of formulaic sequences and utilize them in their essay responses after the training period. Making formulaic sequences salient in input, in turn, helped the upper-intermediate EAP students outperform their peers in the focused instruction and control groups and use significantly more distinct types in their delayed posttest than the upper-intermediate focused instruction group. The performance of the two control groups, on the other hand, remained relatively the same.

Correct (task 2). *Measures of central tendency* and *measures of variability* were computed on the raw scores for the total number of correct formulaic sequences the six groups produced in their essays. According to Table 48, the upper-intermediate focused instruction, saliency, and control groups had higher mean scores for their posttests (*M*=5.92, *SD*=4.72, *n*=13; *M*=8.60, *SD*=4.83, *n*=5; *M*=4.10, *SD*=2.81, *n*=10) and delayed posttests (*M*=5.38, *SD*=3.48, *n*=13; *M*=8.40, *SD*=2.07, *n*=5; *M*=4.77, *SD*=1.55, *n*=10) when compared to their pretests (*M*=4.38, *SD*=4.11, *n*=13; *M*=5.20, *SD*=3.27, *n*=5; *M*=3.50, *SD*=2.46, *n*=10). Similar results were obtained for the advanced focused instruction and control groups: higher mean scores for their posttests (*M*=8.68, *SD*=5.77, *n*=19 and *M*=8.29, *SD*=4.39, *n*=7) and their delayed posttests (*M*=8.48, *SD*=3.73, *n*=19 and *M*=7.57, *SD*=5.44, *n*=7) than those for their pretests (*M*=5.26, *SD*=2.79, *n*=19

Table 48

Descriptive Statistics: Correct (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	4.38	5.92	5.38	5.20	8.60	8.40	3.50	4.10	4.77
Median	3.00	4.00	5.00	5.00	7.00	8.00	2.50	3.50	4.84
Mode	3.00	3.00	4.00	2.00	7.00	6.00[a]	2.00	3.00	3.00
SD	4.11	4.72	3.48	3.27	4.83	2.07	2.46	2.81	1.55
Minimum	0	1.00	1.00	2.00	3.00	6.00	1.00	0	3.00
Maximum	15.00	16.00	13.00	9.00	16.00	11.00	8.00	9.00	7.00

a. Multiple modes exist. The smallest value is shown

and *M*=6.93, *SD*=6.00, *n*=7), as shown Table 49 below. The advanced saliency group, in turn, had the highest mean score at the posttest stage (*M*=7.04, *SD*=3.15, *n*=13) and the lowest mean score at the delayed posttest stage (*M*=5.46, *SD*=3.10, *n*=13).

Table 49

Descriptive Statistics: Correct (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	5.26	8.68	8.48	6.15	7.04	5.46	6.93	8.29	7.57
Median	5.00	8.00	9.00	6.00	6.00	6.00	4.50	7.00	6.00
Mode	3.00[a]	8.00	9.00	4.00[a]	5.00	9.00	3.00[a]	7.00	6.00
SD	2.79	5.77	3.73	3.98	3.15	3.10	6.00	4.39	5.44
Minimum	1.00	3.00	3.00	0	4.00	0	3.00	2.00	1.00
Maximum	11.00	29.00	16.00	14.00	14.00	9.00	20.00	15.00	17.00

a. Multiple modes exist. The smallest value is shown

A *mixed factorial ANOVA* was computed on the raw scores for the number of correct formulaic sequences the six groups produced at the three production stages in order to detect any significant differences between the mean scores listed in Table 48 and Table 49 above. Prior to the analysis, several statistical tests were computed on these three variables to identify outliers, assess the distribution of the sum of scores, and check two major *mixed ANOVA* assumptions. The initial screening for outliers revealed the presence of nine potential outlying scores, of which one score in the advanced saliency

group's posttest was an actual outlier (see Tables E15 and E16). This score (i.e., 29) was replaced with the second highest score within the posttest of the advanced focused group (i.e., 15). The statistical assessment of the distribution of the sum of scores indicated that the data were approximately normally distributed (see Table F8). The results of Mauchly's test (W=1.00, x^2=.06, df=2, p=.971) and Levene's test [T1: F(5, 61)=.76, p=.579; T2: F(5, 61)=1.00, p=.425; T3: F(5, 61)= 1.97, p=.097] confirmed that the data met both the sphericity and homogeneity of variance assumptions.

The results of the *mixed factorial ANOVA*, displayed in Table 50 and Table 51 below, indicated that the only statistically significant main effect was that of time F(2, 122)=5.74, p=.004, partial η^2=.086, which implied that the use of the correct formulaic sequences by the six groups combined differed over time. The statistically significant effect of time was the result of the lower number of correct formulaic sequences that the six groups combined produced in their pretests (M=5.24, SE=.50, n= 67) than the ones they produced in their posttests (M=6.98, SE=.51, n= 67, MD=-1.74, SE=.56, p=.008) and their delayed posttests (M=6.68, SE=.46, n=67, MD=-1.44, SE=.55, p=.032). All other within- and between-subjects effects were statistically nonsignificant (p>.05), suggesting that the differences in the mean scores presented in Table 48 and 49 above and

Table 50

Correct (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	96.172	2	48.086	5.743*	.004	.086
Tm.*Gr.	Sphericity As.	13.585	4	3.396	.406	.804	.013
Tm.* Pro.	Sphericity As.	4.361	2	2.181	.260	.771	.004
Tm.*Gr.*Pro.	Sphericity As.	50.677	4	12.669	1.513	.203	.047
Error	Sphericity As.	1021.457	122	8.373			

*$p \leq .05$

graphically plotted in Figure 20 and Figure 21 below within and between groups were minor and, hence, did not achieve statistical significance.

Table 51

Correct (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	6599.921	1	6599.921	280.694	.000	.821
Groups	21.088	2	10.544	.448	.641	.014
Proficiency	85.011	1	85.011	3.616	.062	.056
Groups*Proficiency	130.753	2	65.377	2.780	.070	.084
Error	1434.286	61	23.513			

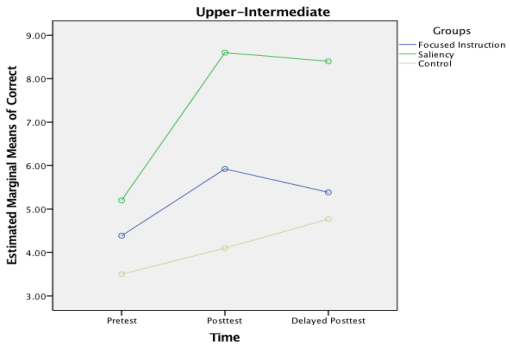


Figure 20. Correct task 2 inter.

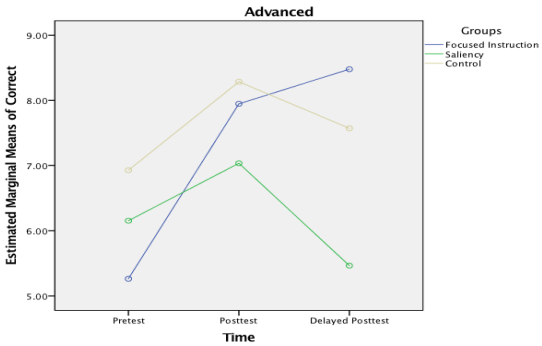


Figure 21. Correct task 2 adv.

Accordingly, the six groups combined used more accurate formulaic sequences after the training period, but there were no differences between and/or within different groups at different proficiency levels that could be attributed to a single pedagogical intervention.

Wrong (task 2). The results of the *descriptive statistics* carried out on the raw scores for the total number of incorrect formulaic sequences that the six groups produced at the three production stages and summarized in Tables 52 and 53 below showed that the upper-intermediate focused instruction group and the advanced saliency group produced fewer incorrect formulaic sequences in the posttest (*M*=3.54, *SD*=3.97, *n*=13 and *M*=3.57,

SD=2.35, *n*=13) and the delayed posttest (*M*=2.77, *SD*=1.64, *n*=13 and *M*=3.00, *SD*=1.63, *n*=13) than they did in the pretest (*M*=4.08, *SD*=2.87, *n*=13 and *M*=4.54, *SD*=2.37, *n*=13).

Table 52

Descriptive Statistics: Wrong (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	4.08	3.54	2.77	6.00	6.00	4.38	3.30	4.00	3.25
Median	4.00	3.00	3.00	7.00	5.00	4.00	3.00	3.00	2.53
Mode	5.00	3.00	1.00[a]	3.00	4.00	1.00[a]	3.00	3.00[a]	1.00[a]
SD	2.87	3.97	1.64	2.83	2.92	3.07	1.42	2.62	2.03
Minimum	0	0	0	3.00	4.00	1.00	1.00	1.00	1.00
Maximum	11.00	15.00	5.00	9.00	11.00	8.00	6.00	9.00	7.00

a. Multiple modes exist. The smallest value is shown

Table 53

Descriptive Statistics: Wrong (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	4.26	5.21	4.00	4.54	3.57	3.00	2.86	5.29	4.43
Median	3.00	6.00	3.00	4.00	2.00	3.95	3.00	4.00	5.00
Mode	3.00	6.00	1.31[a]	4.00[a]	2.00	4.00	0	4.00	3.00[a]
SD	3.59	3.39	3.21	2.37	2.35	1.63	2.34	2.81	2.70
Minimum	0	0	0	1.00	.46	0	0	1.00	0
Maximum	16.00	14.00	12.00	8.00	8.00	5.00	6.00	9.00	8.00

a. Multiple modes exist. The smallest value is shown

The upper-intermediate control group and the advanced focused instruction group had the highest numbers of incorrect formulaic sequences in the posttest (*M*=4, *SD*=2.62, *n*=10 and *M*=5.21, *SD*=3.39, *n*=19) and the lowest ones in the delayed posttest (*M*=3.25, *SD*=2.03, *n*=10 and *M*=4.00, *SD*=3.21, *n*=19). The upper-intermediate saliency group, in turn, had identical mean scores for their pretest (*M*=6, *SD*=2.83, *n*=5) and posttest (*M*=6, *SD*=2.92, *n*=5), which dropped slightly at the delayed posttest stage (*M*=4.38, *SD*=3.07, *n*=5). The advanced control group, on the other hand, had higher mean scores at the

posttest (*M*=5.29, *SD*=2.81, *n*=7) and the delayed posttest stages (*M*=4.43, *SD*=2.70, *n*=7) than the one they did at the pretest stage (*M*=2.86, *SD*=2.34, *n*=7).

It is worth noting here that similar to the first task, most of the incorrectly constructed formulaic sequences by the six groups in their essay responses at the three production stages had structural errors (e.g. subject-verb agreement, tenses, word form, missing article etc.), with very few errors that relate to an inaccurate use of a formulaic sequence or a misspelled one; this was inferred from the mean scores for the types of errors that the six groups had in their essay responses and summarized in Table 54 and Table 55 below.

Table 54

Mean Scores: Types of Errors (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Use	.46	.54	.23	1.00	1.40	1.08	.30	.30	.22
Structure	3.00	2.62	2.23	3.80	3.00	2.78	2.70	2.80	2.50
Spelling	.62	.38	.31	1.20	1.60	.52	.30	.90	.52

Table 55

Mean Scores: Types of Errors (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Use	.58	.58	.62	.38	.97	.28	.25	.57	.43
Structure	3.00	4.05	2.57	3.69	2.38	2.50	2.18	4.14	3.71
Spelling	.68	.58	.82	.46	.22	.22	.43	.57	.29

To determine whether the differences in the mean scores presented in Table 52 and 53 above were statistically significant, a *mixed factorial ANOVA* was conducted on the variables that included the total number of incorrect formulaic sequences. These variables contained three outlying scores (see Tables E17 and E18). Two outliers were

replaced with the next highest score, which resulted in adjusting the posttest maximum score presented in Table 52 for the upper-intermediate focused instruction group (*Max*=6) and the pretest maximum score listed in Table 53 for the advanced focused instruction group (i.e., 8). The third outlier was not replaced because its *z score* exceeded the threshold by .01 and, hence, was not expected to affect the analysis. The results of normality tests (see Table F9), Mauchly's test (W=1, x^2=.11, df=2, p=.947), and Levene's test [T1: $F(5, 61)$=1.01, p=.418; T2: $F(5, 61)$=.58, p=.712; T3: $F(5, 61)$=1.56, p=.186] confirmed that the data within each group were approximately normally distributed and met the sphericity and homogeneity of variance assumptions.

The *mixed factorial ANOVA* results reported in Table 56 and Table 57 below showed that all the within- and between-subjects effects were statistically nonsignificant (p>.05). These results indicated that none of the mean differences listed in Table 52 and Table 53 above or graphically plotted in Figure 22 and Figure 23 below were large enough to achieve statistical significance; therefore, neither focusing on formulaic sequences nor teaching structural materials had any significant effects on the production of the upper-intermediate or advanced EAP students.

Table 56

Wrong (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	20.106	2	10.053	2.113	.125	.033
Tm.*Gr.	Sphericity As.	27.285	4	6.821	1.433	.227	.045
Tm.* Pro.	Sphericity As.	10.994	2	5.497	1.155	.318	.019
Tm.*Gr.*Pro.	Sphericity As.	16.434	4	4.108	.863	.488	.028
Error	Sphericity As.	580.580	122	4.759			

Table 57

Wrong (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	2761.561	1	2761.561	273.393	.000	.818
Groups	19.730	2	9.865	.977	.382	.031
Proficiency	.007	1	.007	.001	.979	.000
Groups*Proficiency	63.057	2	31.528	3.121	.051	.093
Error	616.166	61	10.101			

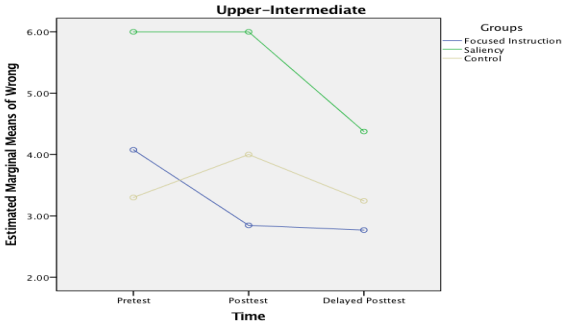


Figure 22. Wrong task 2 inter.

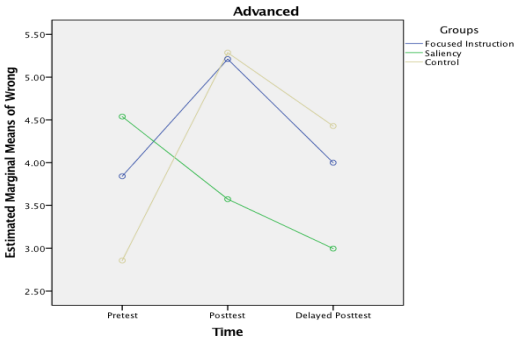


Figure 23. Wrong task 2 adv.

Correct% (task 2). Since raw scores for correct formulaic sequences cannot independently reflect increased accuracy, *measures of central tendency* and *measures of variability* were also computed on the percentage of correct formulaic sequences within each of the six groups at the three production stages. According to Table 58, the upper-intermediate focused instruction, saliency, and control groups had higher percentages of

Table 58

Descriptive Statistics: Correct% (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	45.06	64.50	65.69	45.39	57.25	69.40	48.80	50.68	62.08
Median	50.00	66.66	66.66	50.00	59.25	66.66	53.57	55.49	64.75
Mode	37.50[a]	100.00	44.44	18.18[a]	42.85[a]	57.89[a]	25.00	60.00	75.00
SD	20.94	23.76	20.80	17.33	9.38	12.50	20.97	25.76	14.69
Minimum	0	33.33	25.00	18.18	42.85	57.89	16.66	0	35.76
Maximum	81.81	100.00	100.00	62.50	66.66	87.50	75.00	87.50	80.00

a. Multiple modes exist. The smallest value is shown

correct formulaic sequences for their posttests (*M*=64.50, *SD*=23.76, *n*=13; *M*=57.25, *SD*=9.39, *n*=5; *M*=50.68, *SD*=25.76, *n*=10) and their delayed posttests (*M*=65.69, *SD*=20.80, *n*=13; *M*=69.40, *SD*=12.50, *n*=5; *M*=62.08, *SD*=14.69, *n*=10) than those for their pretests (*M*=45.06, *SD*=20.94, *n*=13; *M*=45.39, *SD*=17.33, *n*=5; *M*=48.80, *SD*=20.97, *n*=10).

Similar results were obtained for both the advanced focused instruction group and the advanced saliency group, that is, higher mean scores for their posttests (*M*=63.47, *SD*=16.30, *n*=19 and *M*=68.10, *SD*=13.40, *n*=13) and delayed posttests (*M*=70.84, *SD*=16.78, *n*=19 and *M*=57.97, *SD*=22.20, *n*=13) than those for their pretests (*M*=57.99, *SD*=19.35, *n*=19 and *M*=52.69, *SD*=23.59, *n*=13), as shown in Table 59 below. The advanced control group, in turn, had opposite trends in that the mean scores for the percentages of the correct formulaic sequences they had in their posttest (*M*=58.24, *SD*=21.41, *n*=7) and delayed posttest (*M*=66.16, *SD*=16.98, *n*=7) were lower than that for their pretest (*M*=71.90, *SD*=25.49, *n*=7).

Table 59

Descriptive Statistics: Correct% (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	57.99	63.47	70.84	52.69	68.10	57.97	71.90	58.24	66.16
Median	66.66	62.50	71.42	57.89	68.75	64.28	63.63	63.15	66.66
Mode	66.66	50.00	66.66[a]	60.00	66.66[a]	66.66	100.00	33.33[a]	50.00[a]
SD	19.35	16.30	16.78	23.59	13.40	22.20	25.49	21.41	16.98
Minimum	20.00	36.36	42.85	0	41.66	0	33.33	33.33	50.00
Maximum	100.00	100.00	100.00	85.71	90.65	85.71	100.00	93.75	100.00

a. Multiple modes exist. The smallest value is shown

A *mixed factorial ANOVA* was computed on the percentage of correct formulaic sequences in order to determine whether the differences in the mean scores presented in Table 58 and Table 59 above were statistically significant. Prior to computing the *mixed*

factorial ANOVA, the delayed posttest minimum score for the advanced saliency group (i.e., 0) was replaced with 37.5 because its *z score* was greater than -2.58 (see Table E19 and Table E20). The results of the normality tests (see Table F10), Mauchly's test (W=.99, x^2=.49, df= 2, p=.784), and Levene's test [T1: F(5, 61)=.38, p=.863; T2: F(5, 61)= 1.78, p=.130; T3: F(5, 61)=.40, p=.850] confirmed that the data were approximately normally distributed and met the sphericity and homogeneity of variance assumptions.

The results of the tests of within-subjects effects summarized in Table 60 below indicated that the only statistically significant main effect was that of time F(2, 122)=6.59, p=.002, partial η^2=.097. Based on the pairwise comparisons results, the significant effect of time was due to the statistically significantly higher percentage of correct formulaic sequences that the six groups combined had at the delayed posttest stage (M=65.84, SE=2.27, n=67) than the one they did at the pretest stage (M=53.64, SE=2.86, n=67, MD=12.20, SE=3.21, p=.001). The differences between the pretest and the posttest and the posttest and the delayed posttest were statistically nonsignificant (p>.05).

Table 60

Correct% (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	4141.551	2	2070.775	6.588*	.002	.097
Tm.*Gr.	Sphericity As.	2228.485	4	557.121	1.772	.139	.055
Tm.* Pro.	Sphericity As.	1423.748	2	711.874	2.265	.108	.036
Tm.*Gr.*Pro.	Sphericity As.	970.305	4	242.576	.772	.546	.025
Error	Sphericity As.	38347.759	122	314.326			

*$p \leq .05$

Tests of the between-subjects effects, presented in Table 61 below, showed that the only statistically significant effect was that of proficiency level F(1, 61)=3.99, p=.050, partial η^2=.061, which was attributed to the higher overall mean score for the

three advanced groups combined (*M*=63.36, *SE*=2.21, *n*=39) than that for the three upper-intermediate groups combined (*M*=56.54, *SE*=2.60, *n*=28, *MD*=6.82, *SE*=3.42, *p*=.050). All other within-subjects and between-subjects effects did not approach statistical significance, which indicated that the mean differences within and between different groups, also visually depicted in Figure 24 and Figure 25 below, were not large enough to approach statistical significance.

Table 61

Correct% (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	597755.951	1	597755.951	1231.848	.000	.953
Groups	185.793	2	92.897	.191	.826	.006
Proficiency	1936.296	1	1936.296	3.990*	.050	.061
Groups*Proficiency	450.404	2	225.202	.464	.631	.015
Error	29600.335	61	485.251			

*$p \leq .05$

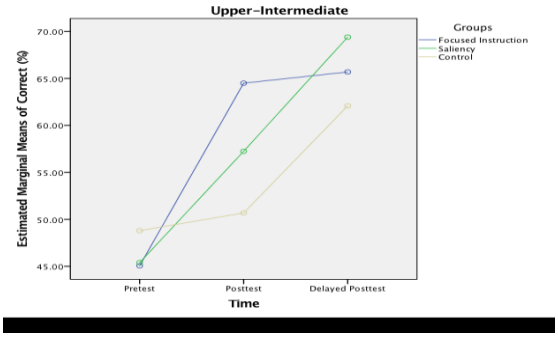


Figure 24. Correct% task 2 inter.

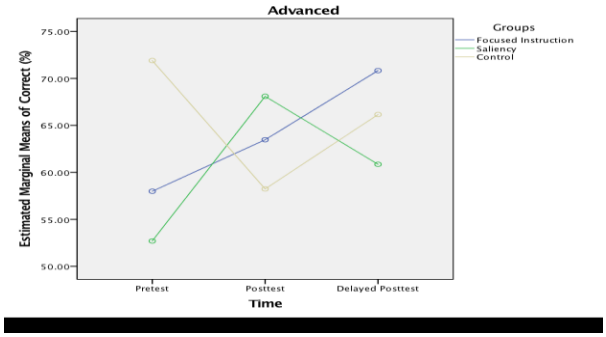


Figure 25. Correct% task 2 adv.

These results indicate that the accurate production of formulaic sequences by the six groups combined improved one month after the training period, which may be attributed to their natural progression as a result of being registered in an EAP program and/or the similar positive effects of the three pedagogical interventions rather than a single presentation method they received during the training period.

Results summary. The results obtained from the multiple *mixed factorial ANOVAs* along with the *simple-effects analyses* and simple comparisons to answer the first research question revealed different effects of different instruction methods on the upper-intermediate and advanced EAP students' use of formulaic sequences in graphical summaries and essay responses.

Focused instruction of formulaic sequences. An explicit instructional approach to formulaic sequences had significant effects on the advanced EAP students' graphical summaries but very minimal effects on their essay responses; the same instructional method only minimally affected the upper-intermediate students' use of formulaic sequences in both writing tasks. In other words, focused instruction of formulaic sequences helped the advanced EAP students acquire, retain, and utilize higher numbers and more versatile types of formulaic sequences when summarizing graphical information. It also seemed to help the advanced EAP students produce more accurate formulaic sequences in their graphical summaries after the training period and outperform their counterparts in the saliency group in both the use and accuracy of formulaic sequences in graphical summaries at the delayed posttest stage. Focused instruction of formulaic sequences also helped the advanced EAP students use more formulaic sequences in their essay responses at the end of the training period and more versatile types at both production stages after the training period. This method of instruction, however, did not help the advanced EAP students produce more accurate or fewer inaccurate formulaic sequences in their essay responses.

As for the upper-intermediate students, an explicit instructional approach did not promote their use of formulaic sequences, increase the number or percentage of correct

formulaic sequences, or decrease the number of incorrect formulaic sequences they produced in their essay responses. On the other hand, it helped them utilize more distinct types in their graphical summaries at the end of the training period and outperform the upper-intermediate students in the control group in their higher use of distinct types in the first writing task (both overall and at the second production stage).

Salient formulaic sequences. Making formulaic sequences salient in input had different effects on the upper-intermediate and advanced EAP students' use of formulaic sequences in their graphical summaries and essay responses. That is, this presentation method did not promote the upper-intermediate EAP students' overall or accurate use of formulaic sequences in their graphical summaries. However, it helped them augment their summaries with more distinct types after the training period and outperform the upper-intermediate focused instruction and control groups in the overall number and distinct types used in essay responses. Making formulaic sequences salient in input for the advanced EAP students did not have any effects on their tendency to use more formulaic sequences, both overall and distinct types, or their ability to produce more correct and fewer incorrect formulaic sequences in their essay responses. The same presentation method, on the other hand, helped the advanced EAP students use more formulaic sequences, both repeated and distinct ones, and more correct ones in their graphical summaries at the end of the training period without any effects on retention.

Writing materials. Although the upper-intermediate and advanced control groups might have been incidentally exposed to formulaic sequences as a result of dealing with classroom and other academic materials, their use of formulaic sequences remained relatively stable in both writing tasks. In other words, they did not use more formulaic

sequences, utilize higher numbers of correct formulaic sequences, or produce fewer incorrect ones whether in their graphical summaries or essay responses despite being taught some materials that focus on form.

In a nutshell, while both an explicit instructional approach and a saliency approach to presenting formulaic sequences affected, at least to some extent, the participants' performance and helped them occasionally outperform their peers in the control groups, a focused instructional approach seemed to have stronger effects, at least in this study, on the EAP students' delayed learning gains than a saliency approach. It also seemed to effect more changes in the EAP students' utilization of formulaic sequences, especially at the advanced proficiency level, than the other presentation method (i.e., saliency).

Research Question Two

To answer the second research question—*Will both saliency and explicit instruction of formulaic sequences (a) affect the number of words EAP students generate in different timed written texts as well as the evaluation they receive and (b) help them outdo EAP students who do not practice formulaic sequences?—descriptive statistics* and *mixed factorial ANOVAs* were run on the variables representing the raw scores for the total number of words generated in and the evaluation assigned to graphical summaries and essay responses to respectively summarize the data and highlight any statistically significant differences within and/or between different groups.

Words (task 1). The *descriptive statistics* calculated for the total number of words that the three upper-intermediate groups generated at the three production stages indicated that the upper-intermediate focused instruction and saliency groups produced

fewer words at the posttest (*M*=114.85, *SD*=23.21, *n*=13 and *M*=90.80, *SD*=29.96, *n*=5) and delayed posttest stages (*M*=100.69, *SD*=36.67, *n*=13 and *M*=97.63, *SD*=28.18, *n*=5) than they did in their pretests (*M*=138, *SD*=40.03, *n*=13 and *M*=131.20, *SD*=28.60, *n*=5), as shown in Table 62 below. The upper-intermediate control group, in turn, produced the highest number of words in the posttest (*M*=125, *SD*=31.99, *n*=10) and the lowest number of words in the delayed posttest (*M*=99.47, *SD*=41.85, *n*=10).

Table 62

Descriptive Statistics: Words (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	138.00	114.85	100.69	131.20	90.80	97.63	112.20	125.00	99.47
Median	133.00	116.00	109.00	128.00	82.00	88.00	108.50	129.00	97.34
Mode	133.00	149.00	83.00	102.00[a]	61.00[a]	68.13[a]	69.00[a]	77.00[a]	87.00
SD	40.03	23.21	36.67	28.60	29.96	28.18	33.08	31.99	41.85
Minimum	65.00	62.00	30.00	102.00	61.00	68.13	69.00	77.00	3.00
Maximum	203.00	149.00	162.00	170.00	123.00	129.00	177.00	166.00	157.00

a. Multiple modes exist. The smallest value is shown

The three advanced groups had different trends, as displayed in Table 63 below. While the advanced saliency group generated fewer words in their posttest (*M*=104.21, *SD*=21.81, *n*=13) and their delayed posttest (*M*=106.05, *SD*=20.25, *n*=13) than they did in their pretest (*M*=113.54, *SD*=30.47, *n*=13), the mean score for the number of words produced by the advanced control group in the pretest (*M*=118.05, *SD*=21.99, *n*=7) was lower than the mean scores for their posttest (*M*=134.14, *SD*=23.53, *n*=7) and their delayed posttest (*M*=167.29, *SD*=18.11, *n*=7). The advanced focused instruction group, on the other hand, generated the lowest number of words in their posttest (*M*=115.58, *SD*=29.35, *n*=19) and the highest number of words in their delayed posttest (*M*=127.40, *SD*=25.47, *n*=19).

Table 63

Descriptive Statistics: Words (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	117.78	115.58	127.40	113.54	104.21	106.05	118.05	134.14	167.29
Median	112.00	116.00	121.00	105.00	99.00	107.00	122.00	130.00	157.00
Mode	126.00	116.00	153.00	95.00[a]	97.00	129.00	91.00[a]	110.00[a]	189.00
SD	27.07	29.35	25.47	30.47	21.81	20.25	21.99	23.53	18.11
Minimum	74.00	29.00	88.00	75.00	68.00	65.00	91.00	110.00	146.00
Maximum	167.00	157.00	175.00	169.00	141.00	129.00	156.00	168.00	189.00

a. Multiple modes exist. The smallest value is shown

In order to determine whether the differences in the mean scores listed in Tables 62 and 63 above were statistically significant, a *mixed factorial ANOVA* was computed on the raw scores for the total number of words the six groups generated when summarizing graphical information at the three production stages. The initial screening for outliers revealed the presence of five potential outlying scores, of which one case in the posttest of the advanced focused instruction group had a *z score* that exceeded -2.58 (see Table E21 and Table E22). This score was replaced with the second lowest score within the advanced focused instruction group's posttest variable, which resulted in altering the minimum score presented in Table 63 above (*Min*=82).

The results of Mauchly's test of sphericity (W=.92, x^2=4.91, df= 2, p=.086) and Levene's test of homogeneity of variance [T1: $F(5, 61)$=.70, p=.623; T2: $F(5, 61)$=.82, p=.541; T3: $F(5, 61)$=1.50, p=.204] were also inspected and showed that the data met the sphericity and homogeneity of variance assumptions. As for the distribution of the sum of words within each of the six groups, the results of the *Shapiro-Wilk* normality tests were nonsignificant, but the skewness and kurtosis *z scores* indicated that the summative scores within the upper-intermediate focused instruction and saliency groups were

slightly negatively skewed and the summative scores of the latter were also slightly kurtotic (see Table F11). However, these variables were still deemed appropriate for the analysis because of the very minor normality violation and the robustness of *mixed ANOVA* to slight departures from normality (Tabachnick & Fidell, 2007a).

The *mixed factorial ANOVA* results, presented in Table 64 and Table 65 below, revealed two significant interactions, that is, time×group $F(4, 122)=5.17$, $p=.001$, partial $\eta^2=.145$ and time×proficiency $F(2, 122)=15.64$, $p=.001$, partial $\eta^2=.204$. While the former suggested that different groups generated different numbers of words at the three production stages, the latter implied that the different numbers of the words generated at the three production stages were moderated by the EAP students' proficiency level. All other within- and between-subjects effects were statistically nonsignificant ($p>.05$).

Table 64

Words (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	1564.932	2	782.466	1.733	.181	.028
Tm.*Gr.	Sphericity As.	9334.557	4	2333.639	5.169*	.001	.145
Tm.* Pro.	Sphericity As.	14120.610	2	7060.305	15.640*	.000	.204
Tm.*Gr.*Pro.	Sphericity As.	4036.727	4	1009.182	2.236	.069	.068
Error	Sphericity As.	55074.085	122	451.427			

*$p\leq.05$

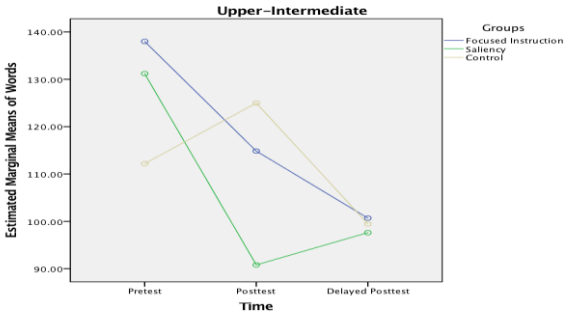
Table 65

Words (Task 1): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	2299906.68	1	2299906.68	1470.227	.000	.960
Groups	8386.403	2	4193.202	2.681	.077	.081
Proficiency	4829.293	1	4829.293	3.087	.084	.048
Groups*Proficiency	5639.127	2	2819.564	1.802	.174	.056
Error	95423.566	61	1564.321			

The visual presentation of the mean scores in Figure 26 and Figure 27 below suggested that the detected statistically significant interactions were likely to be the result

of the decrease in the mean scores for the posttest and delayed posttest of the upper-intermediate focused instruction and saliency groups and the lower mean scores of the same groups at the advanced level than their counterparts in the control group.



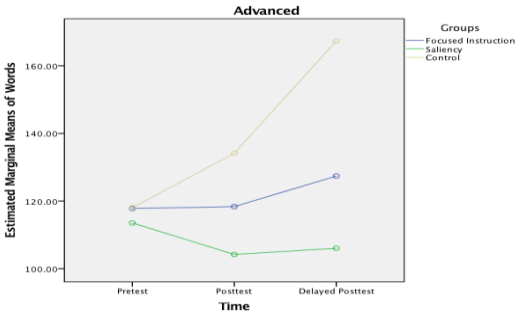


Figure 26. Words task 1 inter. *Figure 27*. Words task 1 adv.

In order to decompose the statistically significant interactions and identify any statistically significant differences between the mean scores plotted in Figure 26 and Figure 27 above, a *simple-effects analysis was* run after splitting the data based on proficiency. The results of Mauchly's and Levene's tests confirmed that the data met the sphericity and homogeneity of variance assumptions at both the upper-intermediate proficiency level [W=.94, x^2=1.41, df=2, p=.495; T1: F(2, 25)=.27, p=.769; T2: F(2, 25)=1.11, p=.344; T3: F(2, 25)=.14, p=.871] and the advanced proficiency level [W=.87, x^2=5.09, df=2, p=.078; T1: F(2, 36)=.72, p=.492; T2: F(2, 36)=.17, p=.841; T3: F(2, 36)=1.39, p=.262].

The results of the within-subjects effects tests summarized in Tables 66 below revealed a statistically significant main effect of time for the upper-intermediate groups F(2, 50)=8.45, p=.001, partial η^2=.253 and the advanced groups F(2, 72)=7.50, p=.001, partial η^2=.172 as well as a statistically significant time by group interaction for the three upper-intermediate groups F(4, 50)=2.82, p=.035, partial η^2=.184 and the three advanced groups F(4, 72)=4.91, p=.001, partial η^2=.214. While the statistically significant effect of

time at both proficiency levels suggested that the three upper-intermediate groups combined and the three advanced groups combined generated different numbers of words at different production stages, the statistically significant interactions indicated that different groups at each proficiency level generated different numbers of words at the three time points.

Table 66

Words: Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	9415.895	2	4707.948	8.448*	.001	.253
	Time*Group	6283.620	4	1570.905	2.819*	.035	.184
	Error	27863.604	50	557.272			
Advanced	Time	5666.114	2	2833.057	7.496*	.001	.172
	Time*Group	7420.001	4	1855.000	4.908*	.001	.214
	Error	27210.481	72	377.923			

*$p \leq .05$

The tests of the between-subjects effects, in turn, revealed a statistically significant main effect of group for the three advanced groups $F(2, 36)=6.89$, $p=.003$, partial $\eta^2=.227$ (see Table 67 below), implying that the overall performance of some advanced groups was different from other groups.

Table 67

Words (Task 1): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	901828.201	1	901828.201	383.280	.000	.939
	Groups	1503.302	2	751.651	.319	.729	.025
	Error	58823.103	25	2352.924			
Advanced	Intercept	1499027.82	1	1499027.82	1474.435	.000	.976
	Groups	14011.212	2	7005.606	6.891*	.003	.277
	Error	36600.463	36	1016.680			

*$p \leq .05$

In light of the multiple statistically significant interactions, the pairwise comparison results at each proficiency level were inspected. These results confirmed that the upper-intermediate focused instruction group generated statistically significantly

fewer words in the posttest (*M*=114.85, *SD*=23.21, *n*=13, *MD*=-23.15, *SE*=8.53, *p*=.036) and the delayed posttest (*M*=100.69, *SD*=36.67, *n*=13, *MD*=-37.31, *SE*=10.29, *p*=.004) than they did in the pretest (*M*=138, *SD*=40.03, *n*=13). The upper-intermediate saliency group also generated more words in the pretest (*M*=131.20, *SD*=28.60, *n*=5) than they did in the posttest (*M*=90.80, *SD*=29.96, *n*=5, *MD*=40.40, *SE*=13.75, *p*=.021) but not the delayed posttest (*M*=97.63, *SD*=28.18, *n*=5, *p*>.05). None of the differences that were visually depicted in Figure 26 above within the upper-intermediate control group over time or between the three upper-intermediate groups at each time point approached statistical significance (*p*>.05).

The pairwise comparisons for the three advanced groups, on the other hand, revealed the two experimental groups' relatively stable performance over time; that is, the increase in the mean scores for the focused instruction group's and saliency group's posttest and delayed posttest seemed to be minimal and, hence, did not approach statistical significance (*p*>.05). However, these results confirmed that the noticeable higher mean score for the advanced control group's delayed posttest (*M*=167.29, *SD*=18.11, *n*=7) than the ones for their pretest (*M*=118.05, *SD*=21.99, *n*=7, *MD*=49.24, *SE*=11.84, *p*=.001) and for their posttest (*M*=134.14, *SD*=23.53, *n*=7, *MD*=33.14, *SE*=8.47, *p*=.001) was statistically significant.

As for the differences between the three advanced groups at each time point, the pairwise comparisons results indicated that the number of words the three advanced groups generated was very similar at the pretest stage (*p*>.05) but not at the posttest or the delayed posttest stages (*p*<.05). To elaborate, the saliency group generated statistically significantly fewer words in the posttest (*M*=104.21, *SD*=21.81, *n*=13) than did the

control group (*M*=134.14, *SD*=23.53, *n*=7, *MD*=-29.94, *SE*=10.49, *p*=.021). The former could also produce more concise summaries at the delayed posttest stage (*M*=106.05, *SD*=20.25, *n*=13) than their peers in the focused instruction group (*M*=127.40, *SD*=25.47, *n*=19, *MD*=-21.34, *SE*=8.17, *p*=.039) and the control group (*M*=167.29, *SD*=18.11, *n*=7, *MD*=-61.23, *SE*=10.65, *p*=.001). Moreover, the summaries the advanced focused instruction produced at the delayed posttest stage (*M*=127.40, *SD*=25.47, *n*=19) were statistically significantly more concise than the ones produced by the advanced control group (*M*=167.29, *SD*=18.11, *n*=7, *MD*=-39.89, *SE*=10.04, *p*=.001). The *Tukey-Kramer* post hoc test, in turn, showed that the statistically significant effect of group at the advanced level was due to the lower overall mean score for the saliency group (*M*=107.93, *SE*=5.11, *n*=13) than that for the advanced control group (*M*=139.83, *SE*=6.96, *n*=7, *MD*=-31.89, *SE*=8.63, *p*=.002).

Based on the above-mentioned results, different types of pedagogical interventions seemed to have different effects on the performance of the upper-intermediate and advanced EAP students. Both a focused instructional approach and a saliency approach to presenting formulaic sequences had significant effects on the number of words that the upper-intermediate groups produced after the training period with the former approach having delayed effects on their performance. On the other hand, although these two presentation methods had no observable effects on the performance of the advanced experimental groups over time, they could outperform the advanced control group in their overall performance and/or their production at a single or both stages after the training period.

Moreover, while a focused instructional approach seemed to have more substantial effects than a saliency approach on the performance of the upper-intermediate students, the opposite was true at the advanced level. In other words, not only did the advanced saliency group produce more concise summaries (i.e., overall and at both production stages) after the training period than the advanced control group, but their summaries were also more concise than the ones produced by the advanced focused instruction group at the third production stage. The relatively stable performance of the upper-intermediate control group and the sharp increase in the number of words the advanced control group generated at the third production stage implied that presenting formulaic sequences had more positive effects on the EAP students' ability to produce more concise summaries than not focusing on formulaic sequences.

Evaluation (task 1). *Measures of central tendency* and *measures of variability* were computed on the raw scores for the evaluation that the six groups received for their pretest, posttest, and delayed posttest. According to Table 68, the mean scores for the upper-intermediate focused instruction and control groups' posttests (*M*=68.69, *SD*=8.20,

Table 68

Descriptive Statistics: Evaluation (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	70.23	68.69	61.77	56.00	54.40	55.58	66.00	65.60	58.20
Median	70.00	67.00	62.00	58.00	57.00	55.00	66.00	64.00	65.00
Mode	65.00[a]	63.00[a]	60.00	40.00[a]	60.00	47.88[a]	55.00[a]	60.00[a]	65.00
SD	7.47	8.20	12.46	9.97	8.32	7.05	8.94	5.17	23.35
Minimum	57.00	55.00	40.00	40.00	40.00	47.88	55.00	60.00	0
Maximum	85.00	83.00	83.00	67.00	60.00	65.00	77.00	75.00	83.00

a. Multiple modes exist. The smallest value is shown

n=13; *M*=65.60, *SD*=5.17, *n*=10) and delayed posttests (*M*=61.77, *SD*=12.46, *n*=13; *M*=58.20, *SD*=23.35, *n*=10) were lower than those for their pretests (*M*=70.23, *SD*=7.47,

n=13; *M*=66, *SD*=8.94, *n*=10). Similarly, the mean scores for the upper-intermediate saliency group's evaluation decreased from (*M*=56, *SD*=9.97, *n*=5) in the pretest to (*M*=54.40, *SD*=8.32, *n*=5) in the posttest but increased minimally to (*M*=55.58, *SD*=7.05, *n*=5) in the delayed posttest.

The mean scores for the three advanced groups differed over time in dissimilar ways (see Table 69 below). While the advanced focused instruction group had the highest mean score for their delayed posttest (*M*=72.11, *SD*=8.55, *n*=19) and the lowest mean score for their posttest (*M*=69.26, *SD*=13.37, *n*=19), the advanced saliency group had the highest mean score at the posttest stage (*M*=69.45, *SD*=7.11, *n*=13) and the lowest mean score at the delayed posttest stage (*M*=64.32, *SD*=17.96, *n*=13). The advanced control group, in turn, had higher mean scores for their posttest (*M*=77.29, *SD*=9.07, *n*=7) and delayed posttest (*M*=77, *SD*=6.86, *n*=7) than that for their pretest (*M*=76.64, *SD*=6.92, *n*=7).

Table 69

Descriptive Statistics: Evaluation (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	70.09	69.26	72.11	68.15	69.45	64.32	76.64	77.29	77.00
Median	70.00	68.00	70.00	70.00	68.00	70.00	77.00	75.00	76.00
Mode	65.00[a]	65.00[a]	70.00	70.00[a]	60.00[a]	65.00[a]	68.00[a]	67.00[a]	70.00
SD	6.77	13.37	8.55	8.38	7.11	17.96	6.92	9.07	6.86
Minimum	55.00	25.00	60.00	50.00	60.00	10.00	68.00	67.00	70.00
Maximum	82.00	87.00	90.00	83.00	81.79	78.00	88	95.00	90.00

a. Multiple modes exist. The smallest value is shown

A *mixed factorial ANOVA* was computed on the raw scores for the evaluation that the EAP students received for their summary tasks at the three production stages to determine whether the mean differences presented in Tables 68 and 69 within and/or between groups were statistically significant. These variables were screened for outliers

(see Table E23 and E24). The screening stage revealed the presence of two outliers—one in the advanced focused instruction group's posttest and one in the advanced saliency group's delayed posttest. Since they had *z scores* that exceeded -2.58, they were replaced with the second lowest score within each variable (i.e., 60 and 50 respectively).

The distribution of the sum of scores within each group was also assessed, and the nonsignificant *Shapiro-Wilk* normality test results confirmed that data were approximately normally distributed despite the slight negative skewness of the summative scores for the upper-intermediate control group (see Table F12). The sphericity and homogeneity of variance assumptions were evaluated and indicated that the data violated the sphericity assumption (W=.67, x^2=24.17, df= 2, p=.001) and the assumption of homogeneity of variance in the delayed posttest $F(5, 61)$=2.54, p=.038 but not in the pretest $F(5, 61)$=.38, p=.862 or the posttest $F(5, 61)$=.53, p=.756. Therefore, the multivariate results were reported instead of the results of the within-subjects effects, and a more stringent alpha level of $p \leq .025$ was used in the interpretation of the results in order to counter the possible inflation of *Type I error*.

The results of the multivariate tests, reported in Table 70 below, indicated that none of the within-subjects effects was statistically significant (p>.025). These results implied that the mean differences identified in Table 68 and Table 69 above and graphically plotted in Figure 28 and Figure 29 below were statistically negligible.

Table 70

Evaluation (Task 1): Multivariate Tests

Source		Value	F	df	Error df	Sig.	Partial Eta Sq.
Tm.	Wilks' Lambda	.952	1.524	2.000	60.000	.226	.048
Tm.*Gr.	Wilks' Lambda	.986	.209	4.000	120.000	.933	.007
Tm.* Pro.	Wilks' Lambda	.918	2.685	2.000	60.000	.076	.082
Tm.*Gr.*Pro.	Wilks' Lambda	.955	.695	4.000	120.000	.597	.023

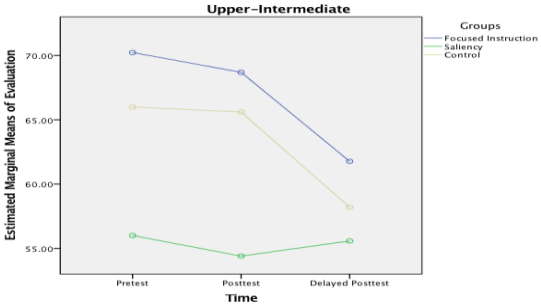


Figure 28. Evaluation task 1 inter.

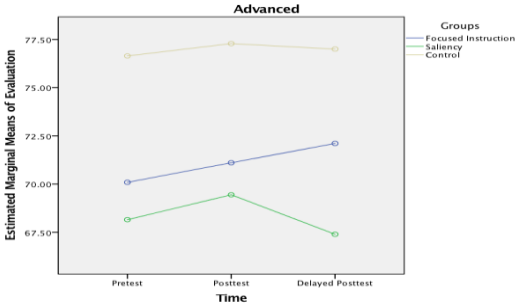


Figure 29. Evaluation task 1 adv.

The results of the tests of between-subjects effects, on the other hand, revealed two statistically significant effects, that is, the effects of groups $F(2, 61)=5.39$, $p=.007$, partial $\eta^2=.150$ and the effects of proficiency level $F(1, 61)=24.56$, $p=.001$, partial $\eta^2=.287$ (see Table 71 below), with the former implying that the overall evaluation of different groups, irrespective of their proficiency levels, was different and the latter suggesting that the overall evaluation of students at different proficiency levels was different.

Table 71

Evaluation (Task 1): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	746251.839	1	746251.839	4148.993	.000	.986
Groups	1938.723	2	969.361	5.389*	.007	.150
Proficiency	4416.951	1	4416.951	24.557*	.000	.287
Groups*Proficiency	978.712	2	489.356	2.721	.074	.082
Error	10971.666	61	179.863			

*$p \leq .025$

An inspection of the simple comparisons results confirmed that the significant effect of proficiency was due to the statistically significantly lower oveall evaluation of the upper-intermediate EAP students (M= 61.83, SE= 1.59, n=28) when compared to the advanced students' overall evaluation (M=72.14, SE=1.35, n=39, MD=-10.31, SE=2.08, p=.001). The *Tukey-Kramer* post hoc test results, on the other hand, indicated that none of the mean differences between the focused instruction (M=69, SE=1.39, n=32), the

saliency (*M*=61.83, *SD*=2.04, *n*=18), and/or the control (*M*=70.12, *SE*=1.91, *n*=17) groups was large enough to achieve statistical significance (*p*>.025). While the higher evaluation scores for the advanced students was expected because higher proficiency usually means better evaluation, the *ANOVA* results showed that neither focusing on formulaic sequences nor teaching linguistic aspects had any significant effects on the evaluation that the six groups received for their graphical summaries.

Words (task 2). The results of the *descriptive statistics* calculated for the numbers of words the three upper-intermediate groups generated in their essay responses suggested that both the upper-intermediate focused instruction and control groups had lower mean scores for their posttests (*M*=193.08, *SD*=61.41, *n*=13 and *M*=229.70, *SD*=76.52, *n*=10) and their delayed posttests (*M*=190, *SD*=33.62, *n*=13 and *M*=210.42, *SD*=39.23, *n*=10) than those for the pretests (*M*=247.92, *SD*=86.46, *n*=13 and *M*=236.60, *SD*=74.22, *n*=10), as can be seen in Table 72 below. The upper-intermediate saliency group, on the other hand, generated more words at the posttest (*M*=224, *SD*=29.33, *n*=5) and the delayed posttest stages (*M*=243.25, *SD*=67.08, *n*=5) than they did at the pretest stage (*M*=195, *SD*=59.38, *n*=5).

Table 72

Descriptive Statistics: Words (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	247.92	193.08	190.00	195.00	224.00	243.25	236.60	229.70	210.42
Median	236.00	196.00	196.00	201.00	225.00	218.00	218.50	216.00	207.50
Mode	31.00[a]	62.00[a]	122.00[a]	127.00[a]	186.00[a]	175.00[a]	150.00[a]	141.00[a]	153.28[a]
SD	86.46	61.41	33.62	59.38	29.33	67.08	74.22	76.52	39.23
Minimum	31.00	62.00	122.00	127.00	186.00	175.00	150.00	141.00	153.28
Maximum	369.00	305.00	236.00	256.00	266.00	349.00	387.00	348.00	267.00

a. Multiple modes exist. The smallest value is shown

Table 73

Descriptive Statistics: Words (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	201.84	249.11	267.47	217.31	202.64	179.73	203.32	227.43	229.71
Median	199.00	246.00	282.17	206.00	196.00	195.00	222.00	220.00	207.00
Mode	51.00[a]	254.00	287.00	80.00[a]	60.00[a]	21.00[a]	92.25[a]	177.00[a]	58.00[a]
SD	65.34	58.97	52.42	68.42	82.11	72.37	53.87	53.43	104.97
Minimum	51.00	166.00	161.31	80.00	60.00	21.00	92.25	177.00	58.00
Maximum	299.00	422.00	383.00	339.00	327.00	268.00	246.00	337.00	394.00

a. Multiple modes exist. The smallest value is shown

The three advanced groups had opposite trends to those of the three upper-intermediate groups, as shown in Table 73 above. While the mean scores for the advanced focused instruction and control groups' posttests (*M*=249.11, *SD*=58.97, *n*=19 and 227.43, *SD*=53.43, *n*=7) and delayed posttests (*M*=267.47, *SD*=52.42, *n*=19 and *M*=229.71, *SD*=104.97, *n*=7) were higher than those for their pretests (*M*=201.84, *SD*=65.34, *n*=19 and *M*=203.32, *SD*=53.87, *n*=7), the advanced saliency group generated fewer words in the posttest (*M*=202.64, *SD*=82.11, *n*=13) and the delayed posttest (*M*=179.73, *SD*=72.37, *n*=13) than they did in the pretest (*M*=217.31, *SD*=68.42, *n*=13).

A *mixed factorial ANOVA* was run on the words variables to identify any statistically significant differences in the mean scores reported in Tables 72 and 73 above. The initial screening for the presence of outliers revealed one outlying score whose *z score* exceeded +2.58 within the advanced focused instruction group's posttest variable (see Table E25 and Table E26). This value (i.e., 422) was replaced with the second highest number of words generated by the same group at the posttest stage (i.e., 319), and, hence, the maximum score for the advanced focused instruction group's posttest was altered (*Max*=319). The normality tests revealed the approximately normal distribution of the summative scores within each group (see Table F13). The results of Mauchly's test

and Levene's test were also checked and confirmed that the data met the assumptions of sphericity (*W*=.99, x^2=.88, *df*=2, *p*=.643) and homogeneity of variance [T1: *F*(5, 61)=.30, *p*=.914; T2: *F*(5, 61)=2.05, *p*=.084; T3: *F*(5, 61)=2.18, *p*=.068].

The results of the *mixed factorial ANOVA* provided in Tables 74 and 75 below showed that the only statistically significant effect was the time×group×proficiency interaction *F*(4, 122)=4.96, *p*=.001, partial η^2=.140. The highest order statistically significant interaction suggested that different groups at different proficiency levels generated different numbers of words at the three production stages, which can also be seen in the visual depiction of the mean scores in Figures 30 and 31 below. None of the other within-subjects and between-subjects effects was found to be statistically significant (*p*>.05).

Table 74

Words (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	353.802	2	176.901	.062	.940	.001
Tm.*Gr.	Sphericity As.	2545.930	4	636.483	.222	.925	.007
Tm.* Pro.	Sphericity As.	7830.037	2	3915.018	1.369	.258	.022
Tm.*Gr.*Pro.	Sphericity As.	56753.284	4	14188.321	4.960*	.001	.140
Error	Sphericity As.	349011.753	122	2860.752			

*$p \leq .05$

Table 75

Words (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	7981559.56	1	7981559.56	1172.993	.000	.951
Groups	5900.431	2	2950.215	.434	.650	.014
Proficiency	5.145	1	5.145	.001	.978	.000
Groups*Proficiency	19880.396	2	9940.198	1.461	.240	.046
Error	415070.883	61	6804.441			

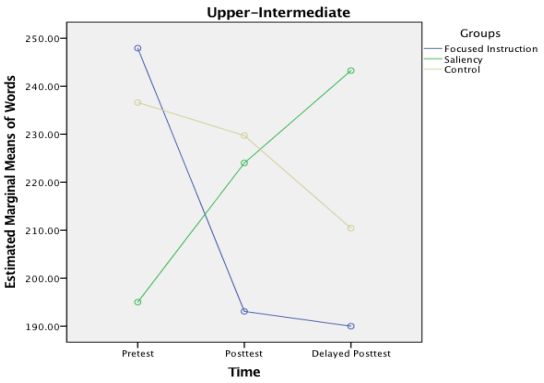


Figure 30. Words task 2 inter.

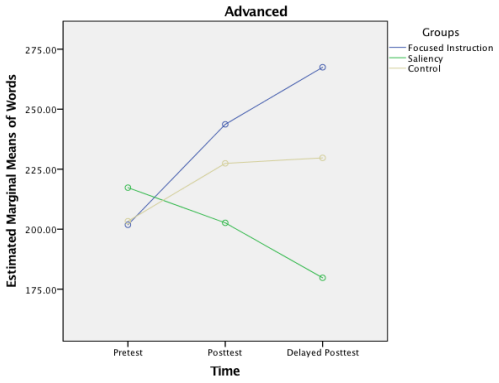


Figure 31. Words task 2 adv.

A *simple-effects analysis* was computed after splitting the data based on proficiency in order to identify any statistically significant differences within and between different groups at different proficiency levels. Since the data, after being split, met the assumption of sphericity at both the upper-intermediate (W=.92, x^2=2.13, df=2, p=.344) and advanced levels (W=1.00, x^2=.02, df=2, p=.989), the results of the tests of within-subjects effects were reported. An alpha level of $p \leq .05$ was used in the interpretation of all effects because the data did not violate the homogeneity of variance assumption at the upper-intermediate level [T1: F(2, 25)=.15, p=.862; T2: F(2, 25)=2.41, p=.111; T3: F(2, 25)=2.13, p=.140] or the advanced level [T1: F(2, 36)=.21, p=.810; T2: F(2, 36)=2.70, p=.081; T3: F(2, 36)=1.90, p=.164].

The results of the *simple-effects analysis*, however, showed that almost all the within- and between-subjects effects were not statistically significant (p>.05), except for the time by group interaction for the three advanced groups F(4, 72)=3.43, p=.013, partial η^2=.160 (see Table 76 and Table 77 below). The statistically significant interaction at the advanced proficiency level suggested that the different numbers of words generated at different time points were different for different groups.

Table 76

Words: Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
	Time	2093.487	2	1046.744	.397	.675	.016
Intermediate	Time*Group	25020.300	4	6255.075	2.371	.065	.159
	Error	131906.204	50	2638.124			
	Time	6857.054	2	3428.527	1.137	.326	.031
Advanced	Time*Group	41310.329	4	10327.582	3.425*	.013	.160
	Error	217105.548	72	3015.355			

*$p \leq .05$

Table 77

Words: Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
	Intercept	3431990.87	1	3431990.87	509.901	.000	.953
Intermediate	Groups	4115.639	2	2057.819	.306	.739	.024
	Error	168267.373	25	6730.695			
	Intercept	4763955.19	1	4763955.19	694.894	.000	.951
Advanced	Groups	33121.416	2	16560.708	2.416	.104	.118
	Error	246803.510	36	6855.653			

An inspection of the pairwise comparisons results indicated that the statistically significant interaction at the advanced level was due to the higher mean score for the number of words the advanced focused instruction group generated in the delayed posttest (M=267.47, SD= 52.42, n=19) than the one they did in the pretest (M=201.84, SD=65.34, n=19, MD=65.63, SE=18.03, p=.003) as well as the number of words the advanced saliency group generated at the same production stage (M=179.73, SD=72.37, n=13, MD=87.74, SE=25.34, p=.004). Given the seemingly significant differences within and between the three upper-intermediate groups (see Figure 30 above), the pairwise comparisons results that were produced for the three upper-intermediate groups as part of the *simple-effect analysis* were also checked. These results demonstrated that the upper-intermediate focused instruction group produced statistically significantly fewer words in the delayed posttest (M=190, SD=33.62, n=13) than they did in the pretest (M=247.92,

SD=86.46, *n*=13, *MD*=-57.92, *SE*=21.36, *p*=.036). All other differences within and between the three upper-intermediate groups and/or the three advanced groups did not approach statistical significance (*p*>.05).

These results suggested that focused instruction of formulaic sequences might have helped the advanced EAP students produce longer essays and outdo the advanced saliency group one month after the training period. However, it seemed to have no positive effect on the performance of the upper-intermediate students, whose delayed posttest essays were significantly shorter than the pretest ones.

Evaluation (task 2). *Measures of central tendency* and *dispersion* were obtained for the evaluation of the six groups' essays. The results presented in Table 78 below showed that the mean scores for the evaluation of the three upper-intermediate groups had opposite trends. That is, the mean scores for the upper-intermediate focused instruction group's posttest (*M*=55.69, *SD*=16.63, *n*=13) and delayed posttest (*M*=58.62, *SD*=11.72, *n*=13) were lower than that for their pretest (*M*=59.38, *SD*=17.53, *n*=13); the

Table 78

Descriptive Statistics: Evaluation (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	59.38	55.69	58.62	63.60	65.40	61.90	58.10	60.10	63.00
Median	65.00	57.00	60.00	65.00	67.00	63.00	60.00	59.50	64.00
Mode	65.00	50.00[a]	65.00	52.00[a]	55.00[a]	42.00[a]	60.00	50.00[a]	46.22[a]
SD	17.53	16.63	11.72	8.62	6.35	12.65	6.97	10.20	10.53
Minimum	5.00	10.00	40.00	52.00	55.00	42.00	45.00	45.00	46.22
Maximum	73.00	80.00	80.00	73.00	72.00	75.00	68.00	75.00	80.00

a. Multiple modes exist. The smallest value is shown

upper-intermediate saliency group had the highest mean score in the posttest (*M*=65.40, *SD*=6.35, *n*=5) and the lowest mean score in the delayed posttest (*M*=61.90, *SD*=12.65, *n*=5). The evaluation the upper-intermediate control group received, in turn, increased

from (*M*=58.10, *SD*=6.97, *n*=10) in the pretest to (*M*=60.10, *SD*=10.20, *n*=10) in the posttest and again to (*M*=63, *SD*=10.53, *n*=10) in the delayed posttest.

Table 79

Descriptive Statistics: Evaluation (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	62.89	68.53	66.62	66.69	63.36	59.73	68.87	76.00	66.57
Median	65.00	67.00	65.00	67.00	70.00	67.00	72.00	72.00	73.00
Mode	60.00[a]	67.00[a]	57.00[a]	55.00	70.00	75.00	39.08[a]	72.00	15.00[a]
SD	10.85	8.02	10.09	9.37	20.14	23.98	15.35	7.72	24.31
Minimum	30.00	55.00	43.78	55.00	10.00	5.00	39.08	68.00	15.00
Maximum	75.00	85.00	86.00	85.00	85.00	80.00	90.00	88.00	85.00

a. Multiple modes exist. The smallest value is shown

Similar to the three upper-intermediate groups, the three advanced groups, as displayed in Table 79 above, had dissimilar mean scores over time. While the advanced focused instruction group received higher evaluation for their posttest (*M*=68.53, *SD*=8.02, *n*=19) and their delayed posttest (*M*=66.62, *SD*=10.09, *n*=19) than they did for their pretest (*M*=62.89, *SD*=10.85, *n*=19), the advanced saliency group received lower evaluation for their posttest (*M*=63.36, *SD*=20.14, *n*=13) and their delayed posttest (*M*=59.73, *SD*=23.98, *n*=13) than they did for their pretest (*M*=66.69, *SD*=9.37, *n*=13). The advanced control group, in turn, had a higher mean score for their posttest (*M*=76, *SD*=7.72, *n*=7) than their pretest (*M*=68.87, *SD*=15.35, *n*=7); their delayed posttest mean score (*M*=66.57, *SD*=24.31, *n*=7), however, was lower than their pretest and posttest mean scores.

A *mixed factorial ANOVA* was performed on the evaluation variables to determine whether the mean differences presented in Table 78 and Table 79 above were statistically significant. The *boxplots* and *z scores* generated in order to identify and adjust outlying scores revealed the presence of four values whose *z scores* were lower

than -2.58 (see Table E27 and Table E28). These values were replaced with the second lowest evaluation score within each variable, resulting in an adjustment of four minimum scores in Table 78 and Table 79. The upper-intermediate focused instruction group's pretest and posttest minimum scores were respectively replaced with 50 and 45. The pretest of the advanced focused instruction group and the posttest of the advanced saliency group, in turn, were changed to 40 and 32.62.

The distribution of the data was also assessed, and the normality tests results confirmed that the distribution of the sum of scores was for the most part approximately normal, with slight negative skewness in the advanced saliency group and positive kurtosis in the advanced control group (see Table F14). No data transformation was performed because of the nonsignificant results of *Shapiro-Wilk* normality tests, the very minor departures of the above-mentioned variables from normality, and the robustness of *mixed ANOVA* to slight normality violations (Tabachnick & Fidell, 2007a). The results of Mauchly's test and Levene's test indicated that the data violated the assumption of sphericity (W=.90, x^2=6.33, df=2, p=.042) but met the assumption of homogeneity of variance [T1: F(5, 61)=.52, p=.757; T2: F(5, 61)=1.11, p=.364; T3: F(5, 61)=1.63, p=.165]. The multivariate results were, thus, reported.

The *mixed factorial ANOVA* results reported in Table 80 and Table 81 below indicated that the only statistically significant effect was that of proficiency level F(1, 61)=4.43, p=.039, partial η^2=.068, which was due to the statistically significantly lower overall evaluation that the upper-intermediate EAP students received (M=61.33, SE=1.99, n=28) when compared to the advanced EAP students (M=66.84, SE= 1.70, n=39, MD=-5.51, SE=2.62, p=.039). None of the other differences between the mean scores that are

summarized in Table 78 and Table 79 above and graphically plotted in Figure 32 and Figure 33 below achieved statistical significance. These results indicated that none of the target instructional methods had any significant effects on the evaluation of upper-intermediate or advanced EAP students.

Table 80

Evaluation (Task 2): Multivariate Tests

Source		Value	F	df	Error df	Sig.	Partial Eta Sq.
Tm.	Wilks' Lambda	.963	1.152	2.000	60.000	.323	.037
Tm.*Gr.	Wilks' Lambda	.955	.706	4.000	120.000	.589	.023
Tm.* Pro.	Wilks' Lambda	.962	1.184	2.000	60.000	.313	.038
Tm.*Gr.*Pro.	Wilks' Lambda	.903	1.573	4.000	120.000	.186	.050

Table 81

Evaluation (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	683006.663	1	683006.663	2398.918	.000	.975
Groups	181.916	2	90.958	.319	.728	.010
Proficiency	1261.934	1	1261.934	4.432*	.039	.068
Groups*Proficiency	569.711	2	284.856	1.000	.374	.032
Error	17367.584	61	284.714			

*$p \leq .05$

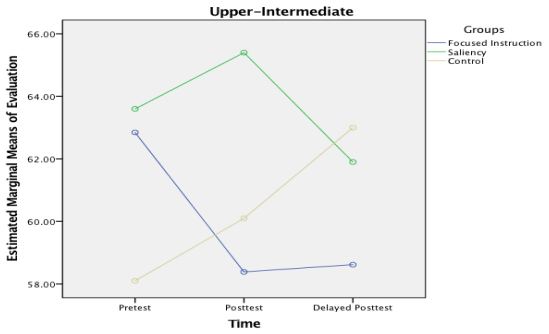


Figure 32. Evaluation task 2 inter.

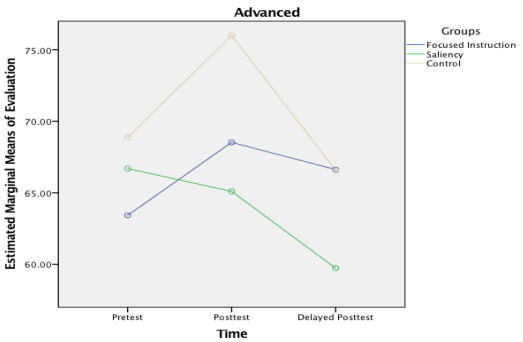


Figure 33. Evaluation task 2 adv.

Results summary. The results showed that the upper-intermediate EAP students who received focused instruction of formulaic sequences wrote more concise summaries at both production stages after the training period, but the number of words they generated in their essay responses also decreased significantly one month after the

training period. The advanced EAP students who were assigned the same instructional approach produced more concise summaries one month after the training period than their counterparts in the control group. The advanced focused instruction group also generated longer essays at the third production stage than they did in their pretest and than their counterparts in the advanced saliency group did at the same production stage.

Making formulaic sequences salient in input helped the upper intermediate EAP students produce shorter summaries at the end of the training period but had no effects on the essays they completed after the training period. The advanced EAP students who practiced salient formulaic sequences, in turn, produced more concise summaries than the advanced control group (both overall and at the posttest and the delayed posttest stages) as well as than the advanced focused instruction group at the delayed posttest stage. The essays the advanced saliency group wrote at the third stage of data collection were, however, significantly shorter than the ones produced by the advanced EAP students in the focused instruction group.

The upper-intermediate EAP students who were taught some writing materials did not write more concise summaries or longer essays after the training period. Similar results were obtained for the advanced control group, whose essay length was relatively the same at the three production stages. The graphical summaries that the advanced control group produced at the delayed posttest stage, on the other hand, were longer than the ones they produced before the training period and those that were produced by their peers in the advanced experimental groups. As for textual evaluation, none of the target pedagogical interventions had any significant effects on the evaluation the upper-intermediate or advanced EAP students received for their summaries or essays.

In brief, both a focused instructional and a saliency approach affected, at least to some extent, the number of words the EAP students generated in their summary tasks at different time points and helped them outperform their counterparts in the control groups in one or multiple aspects. Teaching formulaic sequences explicitly also seemed to have affected the number of words the advanced EAP students generated in their essay responses. These two approaches, however, did not have any effects on the evaluation the upper-intermediate or advanced EAP students received after the training period, nor did a focus on writing materials have any positive effects on the numbers of words generated or the evaluation received by the upper-intermediate and advanced control groups.

Research Question Three

To answer the third research question—*Will both saliency and explicit instruction of formulaic sequences help EAP students (a) reduce the number of linguistic errors they make in different writing tasks and (b) write more accurately than their counterparts in control groups?—descriptive statistics* and *mixed factorial ANOVAs* were run on the variables that included the total number and the percentage of errors the six groups made in their summaries and essays at the three stages of data collection.

Errors (task 1). *Measures of central tendency* and *measures of variability* were computed on the raw scores for the number of errors the three upper-intermediate groups made at each stage of data collection. According to Table 82, both the upper-intermediate focused instruction and saliency groups made the highest number of errors in their pretests (M=27.54, SD=19.70, n=13 and M=29.20, SD=15.80, n=5) and the lowest number of errors in their posttests (M=17.85, SD=11.36, n=13 and M=16.20, SD=7.05, n=5). Their delayed posttest mean scores (M=18.62, SD=10.89, n=13 and M=23.95,

SD=6.79, *n*=5) were higher than their posttest mean scores but lower than those for their pretests. The upper-intermediate control group, in turn, had the highest mean score for their posttest (*M*=25.50, *SD*=11.09, *n*=10) and the lowest one for their delayed posttest (*M*=16.86, *SD*=9.01, *n*=10).

Table 82

Descriptive Statistics: Errors (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	27.54	17.85	18.62	29.20	16.20	23.95	22.10	25.50	16.86
Median	24.00	15.00	21.00	23.00	18.00	23.00	23.00	23.50	17.50
Mode	5.00[a]	8.00[a]	5.00	18.00[a]	7.00[a]	15.00[a]	2.00[a]	41.00	0[a]
SD	19.70	11.36	10.89	15.80	7.05	6.79	10.73	11.09	9.01
Minimum	5.00	3.00	5.00	18.00	7.00	15.00	2.00	7.00	0
Maximum	80.00	34.00	35.00	57.00	24.00	34.00	35.00	41.00	28.89

a. Multiple modes exist. The smallest value is shown

Similar results were obtained for the three advanced groups, as indicated in Table 83 below. Both the advanced focused instruction and saliency groups made fewer errors in their posttests (*M*=13.58, *SD*=6.36, *n*=19 and *M*=15.26, *SD*=5.15, *n*=13) and delayed posttests (*M*=12.88, *SD*=5.99, *n*=19 and *M*=13.65, *SD*=4.48, *n*=13) than they did in their

Table 83

Descriptive Statistics: Errors (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	19.01	13.58	12.88	20.77	15.26	13.65	17.06	18.00	21.29
Median	16.00	12.00	11.00	21.00	14.00	15.00	18.00	17.00	23.00
Mode	16.00	9.00	9.00[a]	15.00	14.00[a]	16.00	9.42[a]	15.00[a]	6.00[a]
SD	10.76	6.36	5.99	9.51	5.15	4.48	4.19	3.42	8.79
Minimum	8.00	1.00	4.00	7.00	8.00	7.50	9.42	15.00	6.00
Maximum	55.00	26.00	24.00	39.00	24.00	21.00	22.00	25.00	30.00

a. Multiple modes exist. The smallest value is shown

pretests (*M*=19.01, *SD*=10.76, *n*=19 and *M*=20.77, *SD*=9.51, *n*=13). The number of errors the advanced control group made in graphical summaries, on the other hand, increased

from (*M*=17.06, *SD*=4.19, *n*=7) in the pretest to (*M*=18, *SD*=3.42, *n*=7) in the posttest and then to (*M*=21.29, *SD*=8.79, *n*=7) in the delayed posttest.

To identify any significant differences between the mean scores that are summarized in Table 82 and Table 83 above, a *mixed factorial ANOVA* was computed on the variables representing the total number of errors. The *boxplots* and *z scores* generated for these variables revealed seven potential outliers, of which two cases (one in the upper-intermediate focused instruction group's pretest and one in the advanced focused instruction group's pretest) had *z scores* higher than +2.58 (see Tables E29 and E30). These scores were, therefore, replaced with the second highest number of errors within each variable, resulting in an adjustment of the pretest maximum scores presented in Table 82 and Table 83 for the upper-intermediate and advanced focused instruction groups (i.e., 44 and 31 respectively). The normality tests computed on the summative variables indicated that distribution of the sum of scores within each group was approximately normal (see Table F15). The results of Mauchly's and Levene's tests showed that the data met the sphericity assumption (W=.97, x^2=1.89, df= 2, p=.388) but severely violated the homogeneity of variance assumption (T:1 $F(5, 61)$=2.84, p=.023; T2: $F(5, 61)$=5.80, p=.001; T3: $F(5, 61)$= 2.92, p=.020). A very conservative alpha level (i.e., $p \leq .01$) was, therefore, used when interpreting the results.

The results of the *mixed factorial ANOVA* presented in Table 84 below demonstrated that the effects of time $F(2, 122)$=9.34, p=.001, partial η^2=.133 and time×group interaction $F(4, 122)$=4.13, p=.004, partial η^2=.119 were statistically significant, which respectively indicated that the total number of errors made by the six

groups combined differed over time and the different numbers of errors were different for different groups.

Table 84

Errors (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	633.815	2	316.908	9.337*	.000	.133
Tm.*Gr.	Sphericity As.	559.950	4	139.988	4.125*	.004	.119
Tm.* Pro.	Sphericity As.	72.298	2	36.149	1.065	.348	.017
Tm.*Gr.*Pro.	Sphericity As.	462.749	4	115.687	3.409	.011	.101
Error	Sphericity As.	4140.645	122	33.940			

*$p \leq .01$

None of the other within-subjects effects or between-subjects effects (see Table 85 below) approached statistical significance ($p > .01$). An inspection of the visual representation of the data in Figure 34 and Figure 35 below suggested that the statistically significant time by group interaction might be attributed to the performance of the three upper-intermediate groups and the three advanced groups.

Table 85

Errors (Task 1): Tests of Between-Subjects Effects

	Type III	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	61198.572	1	61198.572	383.920	.000	.863
Groups	273.755	2	136.878	.859	.429	.027
Proficiency	1030.220	1	1030.220	6.463	.014	.096
Groups*Proficiency	101.524	2	50.762	.318	.728	.010
Error	9723.675	61	159.405			

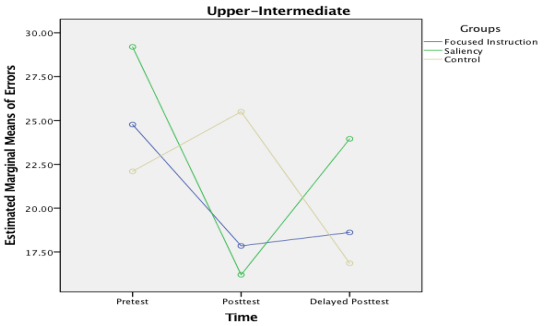


Figure 34. Errors task 1 inter.

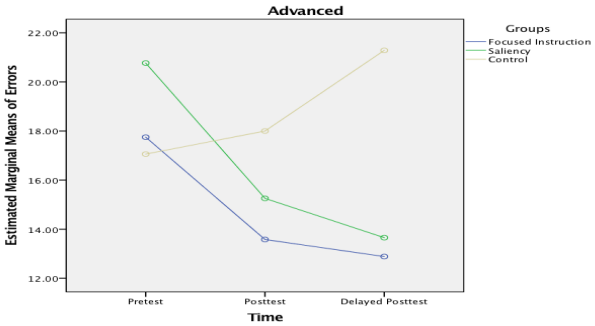


Figure 35. Errors task 1 adv.

Given the statistically significant time by group interaction and the objectives of this research study, the data were split based on proficiency and a *simple-effects analysis* was run in order to reanalyze the time×group interaction at each proficiency level. The results were interpreted using an alpha level of $p \leq .05$ because the data, after being split, met the homogeneity of variance assumption at both the upper-intermediate [T1: $F(2, 25)=.17$, $p=.849$; T2: $F(2, 25)=1.39$, $p=.267$; T3: $F(2, 25)=1.71$, $p=.201$] and the advanced [T1: $F(2, 36)=2.34$, $p=.111$; T2: $F(2, 36)=2.03$, $p=.147$; T3: $F(2, 36)=1.96$, $p=.156$] proficiency levels. The results of Mauchly's test of sphericity also confirmed that the data met the homogeneity of variance assumption at the upper-intermediate ($W=.95$, $x^2=1.16$, $df=2$, $p=.560$) and advanced ($W=.98$, $x^2=.86$, $df=2$, $p=.650$) levels.

The results of the within-subjects effects (see Table 86 below) revealed a statistically significant effect of time at both the upper-intermediate level $F(2, 50)=4.94$, $p=.011$, partial $\eta^2=.165$ and the advanced level $F(2, 72)=3.61$, $p=.032$, partial $\eta^2=.091$, suggesting that the three upper-intermediate groups combined and the three advanced groups combined made different numbers of errors at different time points. The results

Table 86

Errors: Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	486.552	2	243.276	4.938*	.011	.165
	Time*Group	690.328	4	172.582	3.503*	.013	.219
	Error	2463.127	50	49.263			
Advanced	Time	168.345	2	84.172	3.613*	.032	.091
	Time*Group	308.891	4	77.223	3.314*	.015	.156
	Error	1677.517	72	23.299			

*$p \leq .05$

also revealed a statistically significant time by group interaction at the upper-intermediate $F(4, 50)=3.50$, $p=.013$, partial $\eta^2=.219$ and the advanced $F(4, 72)=3.31$, $p=.015$, partial $\eta^2=.156$ levels, implying that different groups made different numbers of errors at the

three production stages. None of the between-subjects effects at either proficiency level approached statistical significance (*p*>.05), as shown in Table 87 below.

Table 87

Errors: Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	33640.234	1	33640.234	122.874	.000	.831
	Groups	81.366	2	40.683	.149	.863	.012
	Error	6844.473	25	273.779			
Advanced	Intercept	27619.516	1	27619.516	345.340	.000	.906
	Groups	264.944	2	132.472	1.656	.205	.084
	Error	2879.202	36	79.978			

The pairwise comparisons results for the three upper-intermediate groups revealed statistically significant differences within the saliency and control groups; that is, the decrease in the mean scores for the saliency group from (*M*=29.20, *SD*=15.80, *n*=5) in the pretest to (*M*=16.20, *SD*=7.05, *n*=5) in the posttest was statistically significant (*MD*=13, *SE*=4.48, *p*=.023). Moreover, the delayed posttest mean score for the control group (*M*=16.86, *SD*=9.01, *n*=10) was significantly lower than their posttest mean score (*M*=25.50, *SD*=11.09, *n*=10, *MD*=-8.64, *SE*=2.82, *p*=.015). All other differences within each of the three upper-intermediate groups or between the three upper-intermediate groups at each time point were statistically nonsignificant (*p*>.05).

The results of the pairwise comparisons for the three advanced groups, in turn, indicated that the statistically significant time by group interaction was attributed to the fewer errors the two experimental groups made after the training period and than their peers in the advanced control group at the third production stage. In other words, the advanced focused instruction group made statistically significantly more errors in their pretest (*M*=17.75, *SD*=7.07, *n*=19) than they did in their posttest (*M*=13.58, *SD*=6.36, *n*=19, *MD*=4.17, *SE*=1.48, *p*=.023) and their delayed posttest (*M*=12.88, *SD*=5.99, *n*=19,

MD=4.87, *SE*=1.68, *p*=.019). Similarly, the advanced saliency group's pretest mean score (*M*=20.77, *SD*=9.51, *n*=13) was statistically significantly higher than their posttest (*M*=15.26, *SD*=5.15, *n*=13, *MD*=5.51, *SE*=1.78, *p*=.012) and delayed posttest (*M*=13.65, *SD*=4.48, *n*=13, *MD*=7.12, *SE*=2.03, *p*=.004) mean scores. The pairwise comparisons results also revealed the statistically significantly higher number of errors the advanced control group made in their delayed posttest essays (*M*=21.29, *SD*=8.79, *n*=7) than their peers in the advanced focused instruction group (*M*=12.88, *SD*=5.99, *n*=19, *MD*=8.40, *SE*=2.71, *p*=.011) and the advanced saliency group (*M*=13.65, *SD*=4.48, *n*=13, *MD*=7.63, *SE*=2.87, *p*=.035) at the same production stage. All other differences between the three advanced groups at each time point and within the advanced control group were very minimal and, thus, did not approach statistical significance (*p*>.05).

These results suggested that focused instruction of formulaic sequences helped the advanced EAP students make fewer errors after the training period and retain a relatively stable level of performance; however, it had no effects on the performance of the upper-intermediate students. Making formulaic sequences salient, in turn, helped the five upper-intermediate EAP students make fewer errors at the end of the training period without any consequences on learning gains; the same presentation method seemed to be more effective for the advanced EAP students in that they produced more accurate summaries at both production stages after the training period. Presenting formulaic sequences, whether by teaching them explicitly or making them salient, helped the advanced students outperform their peers in the control group at the delayed posttest stage. Focusing on linguistic aspects, in turn, had delayed effects on the upper-intermediate EAP students' performance but not that of the advanced EAP students.

Errors% (task 1). Since the raw scores for errors do not reveal, on their own, increased accuracy, *descriptive statistics* were also calculated for the percentage of errors the six groups made in their graphical summaries to reveal any increased accuracy in their writing production. According to Table 88, the upper-intermediate focused instruction group had very similar mean scores for their pretest (*M*=18.86, *SD*=9.44, *n*=13) and delayed posttest (*M*=18.58, *SD*=9.40, *n*=13), which were both higher than their posttest mean score (*M*=14.85, *SD*=8.15, *n*=13). The upper-intermediate control group, on the other hand, had very similar mean scores for their pretest (*M*=21.50, *SD*=13.74, *n*=10) and posttest (*M*=21.26, *SD*=10.11, *n*=10) and the lowest mean score for their delayed posttest (*M*=16.07, *SD*=9.91, *n*=10). The upper-intermediate saliency group, in turn, had the lowest mean score for their posttest (*M*=18.77, *SD*=9.63, *n*=5) and the highest one for their delayed posttest (*M*=25.30, *SD*=6.81, *n*=5).

Table 88

Descriptive Statistics: Errors% (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	18.86	14.85	18.58	21.46	18.77	25.30	21.50	21.26	16.07
Median	18.51	16.30	16.78	17.96	16.66	26.35	17.93	20.19	16.89
Mode	6.75[a]	21.47	16.66	16.82[a]	8.53[a]	17.04[a]	2.53[a]	9.09[a]	0[a]
SD	9.44	8.15	9.40	6.99	9.63	6.81	13.74	10.11	9.91
Minimum	6.75	3.06	4.13	16.82	8.53	17.04	2.53	9.09	0
Maximum	39.40	29.82	38.55	33.52	34.42	33.39	44.92	33.60	33.21

a. Multiple modes exist. The smallest value is shown

The *descriptive statistics* computed on the percentage of errors made by the advanced students revealed decreasing trends within the three advanced groups, as shown in Table 89 below. The percentage of errors the advanced focused instruction, saliency, and control groups had decreased from (*M*=16.07, *SD*=7.44, *n*=19; *M*=18.30, *SD*=7.64, *n*=13; and *M*=14.92, *SD*=4.78, *n*=7) in their pretests to (*M*=11.68, *SD*=5.63, *n*=19;

M=15.08, *SD*=5.89, *n*=13; and *M*=13.66, *SD*=2.95, *n*=7) in their posttests and again to (*M*=10.26, *SD*=4.69, *n*=19; *M*=13.18, *SD*=4.79, *n*=13; and *M*=12.93, *SD*=5.97, *n*=7) in their delayed posttests.

Table 89

Descriptive Statistics: Errors% (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	16.07	11.68	10.26	18.30	15.08	13.18	14.92	13.66	12.93
Median	14.81	10.34	9.91	16.15	13.33	13.26	14.65	13.04	13.75
Mode	6.77[a]	3.44[a]	2.77[a]	9.33[a]	8.08[a]	14.28	7.70[a]	10.11[a]	3.82[a]
SD	7.44	5.63	4.69	7.64	5.89	4.79	4.78	2.95	5.97
Minimum	6.77	3.44	2.77	9.33	8.08	6.20	7.70	10.11	3.82
Maximum	36.91	27.65	21.81	35.05	25.88	24.61	21.97	19.23	20.54

a. Multiple modes exist. The smallest value is shown

A *mixed factorial ANOVA* was computed on the percentage of errors variables to detect any statistically significant differences between the mean scores listed in Tables 88 and 89 above. The generated *boxplots* and *z scores* revealed nine potential outliers, of which two cases within the advanced focused instruction group's pretest and posttest variables had *z scores* greater than +2.58 (see Table E31 and Table E32). These two scores were replaced with 31.31 and 18.54, the second highest percentage of errors within the pretest and posttest variables of that group. The *Shapiro-Wilk* normality test results indicated that the distribution of the summative variables of each group was approximately normal despite the slight negative skewness within the advanced saliency group and the very minimal kurtosis within the advanced control group (see Table F16). An inspection of the results of Mauchly's test and Levene's test indicated that the data met the assumption of sphericity (*W*=.96, x^2=2.45 *df*= 2, *p*=.294) but severely violated the homogeneity of variance assumption [T1: *F*(5, 61)=3.32, *p*=.010; T2: *F*(5, 61)=5.79,

p=.001; T3: F(5, 61)=3.02, p=.017]. A very stringent alpha level (i.e., $p \leq .01$) was applied to the interpretation of all effects to counter any possible *Type I error* inflation.

According to Table 90, almost all the within-subjects effects were statistically nonsignificant (p>.01), except for the effect of time F(2, 122)=5.10, p=.007, partial η^2=.077, which suggested that the overall percentage of errors made by the six groups combined differed over time. An inspection of the tests of the main effects of time showed that the six groups combined had a statistically significantly lower mean score for their posttest (M=15.80, SE=.93, n=67) than that for their pretest (M=18.47, SE=1.17, n=67, MD=-2.67, SE=.86, p=.009).

Table 90

Errors% (Task 1): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	240.262	2	120.131	5.103*	.007	.077
Tm.*Gr.	Sphericity As.	160.605	4	40.151	1.705	.153	.053
Tm.* Pro.	Sphericity As.	99.750	2	49.875	2.118	.125	.034
Tm.*Gr.*Pro.	Sphericity As.	228.479	4	57.120	2.426	.052	.074
Error	Sphericity As.	2872.255	122	23.543			

*$p \leq .01$

Table 91

Errors% (Task 1): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	46806.504	1	46806.504	373.059	.000	.859
Groups	430.193	2	215.096	1.714	.189	.053
Proficiency	1353.229	1	1353.229	10.786*	.002	.150
Groups*Proficiency	13.731	2	6.866	.055	.947	.002
Error	7653.467	61	125.467			

*$p \leq .01$

The between-subject effects results, in turn, revealed a statistically significant effect of proficiency F(1, 61)=10.79, p=.002, partial η^2=.150 (see Table 91 above), which was due to the lower overall mean score for the three advanced groups (M=13.92,

SE=1.13, *n*=39) than that for the three upper-intermediate groups (*M*=19.63, *SE*=1.32, *n*=28, *MD*=-5.71, *SE*=1.74, *p*=.002). All other between- and within-subjects effects were statistically nonsignificant. These results suggest that none of the mean differences presented in Table 88 and Table 89 above and visually depicted in Figure 36 and Figure 37 below was large enough to reach statistical significance. Accordingly, none of the three pedagogical interventions seemed to have a superior effect by itself on the EAP students' overall accuracy.

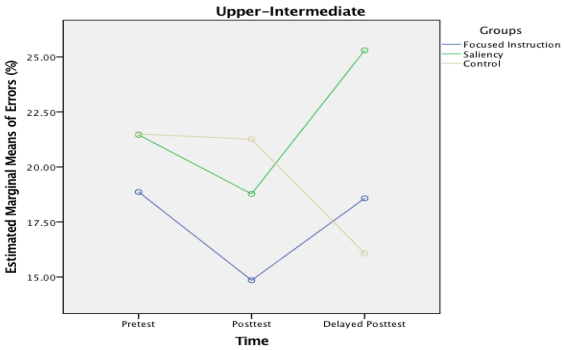


Figure 36. Error% task 1 inter.

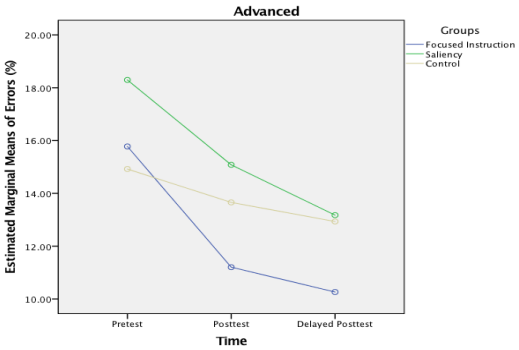


Figure 37. Error% task 1 adv.

Errors (task 2). *Measures of central tendency* and *variability* were run on the raw scores for the errors the six groups made in their essay responses. The *descriptive statistics* obtained for the three upper-intermediate groups and summarized in Table 92 below showed that the upper-intermediate focused instruction and control groups made fewer errors in their posttests (*M*=34.92, *SD*=15.22, *n*=13 and *M*=48.40, *SD*=20.86, *n*=10) and their delayed posttests (*M*=35.38, *SD*=12.76, *n*=13 and *M*=41.13, *SD*=11.98, *n*=10) than they did in their pretests (*M*=51.92, *SD*=26.15, *n*=13 and *M*=53.20, *SD*=20.93, *n*=10). The upper-intermediate saliency group, on the other hand, made more errors at the posttest (*M*=49.80, *SD*=15.48, *n*=5) and the delayed posttest stages (*M*=47.08, *SD*=20.72, *n*=5) than they did at the first stage of data collection (*M*=36.60, *SD*=6.69, *n*=5).

Table 92

Descriptive Statistics: Errors (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	51.92	34.92	35.38	36.60	49.80	47.08	53.20	48.40	41.13
Median	60.00	32.00	33.00	37.00	43.00	43.38	51.00	41.00	44.00
Mode	25.00[a]	32.00[a]	19.00[a]	28.00[a]	37.00[a]	19.00[a]	21.00[a]	41.00	47.00
SD	26.15	15.22	12.76	6.69	15.48	20.72	20.93	20.86	11.98
Minimum	8.00	9.00	19.00	28.00	37.00	19.00	21.00	12.00	12.00
Maximum	90.00	58.00	58.00	44.00	76.00	75.00	90.00	78.00	55.00

a. Multiple modes exist. The smallest value is shown

The three advanced groups had opposite trends to those of the three upper-intermediate groups, as shown in Table 93 below. The advanced saliency group was the only group whose mean scores for the posttest (*M*=29.28, *SD*=18.12, *n*=13) and the delayed posttest (*M*=26.31, *SD*=13.68, *n*=13) were lower than that for the pretest (*M*=37.23, *SD*=13.29, *n*=13). The advanced focused instruction and control groups, on the other hand, respectively had the lowest mean scores for their pretest (*M*=33.16, *SD*=20.91, *n*=19) and delayed posttest (*M*=29.14, *SD*=9.48, *n*=19) and for their pretest (*M*=35.12, *SD*=23.95, *n*=7) and posttest (*M*=28.71, *SD*=15.01, *n*=7).

Table 93

Descriptive Statistics: Errors (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	33.16	36.84	29.14	37.23	29.28	26.31	35.12	28.71	37.86
Median	29.00	32.00	29.00	36.00	29.00	29.00	27.00	25.00	40.00
Mode	17.00	25.00[a]	25.00	27.00	4.00[a]	3.00[a]	11.83[a]	12.00[a]	7.00[a]
SD	20.91	11.42	9.48	13.29	18.12	13.68	23.95	15.01	22.69
Minimum	9.00	25.00	10.58	13.00	4.00	3.00	11.83	12.00	7.00
Maximum	99.00	72.00	55.00	57.00	68.00	46.00	81.00	50.00	75.00

a. Multiple modes exist. The smallest value is shown

A *mixed factorial ANOVA* was computed on the errors variables in order to determine whether the differences in the mean scores listed in Table 92 and Table 93

above were statistically significant. Prior to running a *mixed factorial ANOVA*, the pretest, posttest, and delayed posttest maximum scores of the advanced focused instruction group were respectively replaced with 64, 51, and 41 because their *z scores* were greater than +2.58 (see Table E33 and Table E34). The normality tests computed on the summative variables demonstrated that the sum of errors within each group was approximately normally distributed (see Table F17). As for the assumptions of sphericity and homogeneity of variance, the data met the sphericity assumption (W=.94, x^2=3.59 df=2, p=.167) but violated the homogeneity of variance assumption at two production stages [T1: F(5, 61)=2.85, p=.022; T2: F(5, 61)=2.40, p=.048; T3: F(5, 61)=2.24, p=.062]. A more conservative alpha level ($p \leq .025$) was, thus, used to interpret the results.

The results of the *mixed factorial ANOVA* showed that almost all the within-subjects and between-subjects effects were nonsignificant except for the three-way interaction and the between-subjects effect of proficiency (see Table 94 and Table 95 below). The highest order time×group×proficiency interaction F(4, 122)=6.42, p=.001, partial η^2=.174 suggested that different numbers of errors were made by different groups at different stages of data collection. The statistically significant effect of proficiency F(1, 61)=11.36, p=.001, partial η^2=.157 indicated that the overall number of errors made in the essay responses was different for EAP students at different proficiency levels.

Table 94

Errors (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	673.254	2	336.627	2.653	.074	.042
Tm.*Gr.	Sphericity As.	663.847	4	165.962	1.308	.271	.041
Tm.* Pro.	Sphericity As.	62.308	2	31.154	.246	.783	.004
Tm.*Gr.*Pro.	Sphericity As.	3260.378	4	815.094	6.424*	.000	.174
Error	Sphericity As.	15479.713	122	126.883			

*$p \leq .025$

Table 95

Errors (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	243280.663	1	243280.663	457.417	.000	.882
Groups	640.115	2	320.057	.602	.551	.019
Proficiency	6039.557	1	6039.557	11.356*	.001	.157
Groups*Proficiency	255.609	2	127.805	.240	.787	.008
Error	32443.302	61	531.857			

*$p \leq .025$

A visual inspection of Figure 38 and Figure 39 below suggests that the statistically significant three-way interaction may be due to the decrease in the mean scores of the upper-intermediate focused instruction and advanced saliency groups.

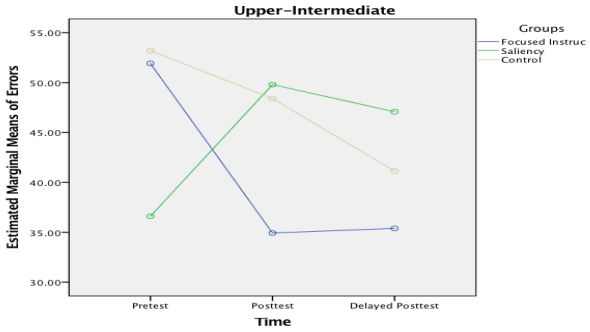


Figure 38. Errors task 2 inter.

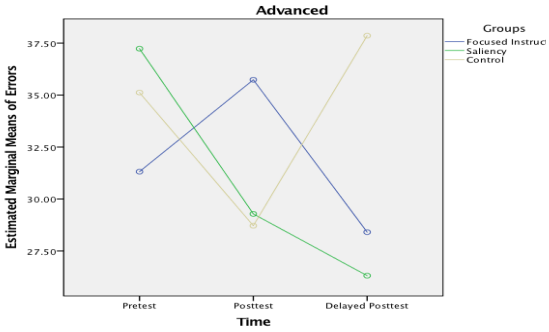


Figure 39. Errors task 2 adv.

Because of the statistically significant highest order interaction and the objectives of the study, *simple-effects analysis* was computed after splitting the data based on proficiency in order to reanalyze the performance of different groups at different proficiency levels. After splitting the data and rerunning the analysis, the data still met the sphericity assumption at the upper-intermediate (W=.90, x^2=2.55, df=2, p=.279) and advanced (W=.97, x^2=1.06, df=2, p=.590) levels. The same alpha level ($p \leq .025$) was used in the interpretation of the results because the homogeneity of variance assumption was moderately violated within the upper-intermediate group [T1: $F(2, 25)$=4.06, p=.030;

T2: $F(2, 25)=.94$, $p=.406$; T3: $F(2, 25)=1.05$, $p=.366$] and the advanced group [T1: $F(2, 36)=1.16$, $p=.325$; T2: $F(2, 36)=5.11$, $p=.011$; T3: $F(2, 36)=4.85$, $p=.014$].

The results of the separate *mixed ANOVAs*, summarized in Table 96 and Table 97 below, revealed a statistically significant time by group interaction at the upper-intermediate level $F(4, 50)=3.38$, $p=.016$, partial $\eta^2=.213$, suggesting that the three-upper-intermediate groups made different numbers of errors at different production stages. All other within- and between-subjects effects were statistically nonsignificant ($p>.025$).

Table 96

Errors (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	436.383	2	218.191	1.379	.261	.052
	Time*Group	2139.674	4	534.918	3.381*	.016	.213
	Error	7910.270	50	158.205			
Advanced	Time	272.859	2	136.430	1.298	.279	.035
	Time*Group	1237.649	4	309.412	2.943	.026	.141
	Error	7569.443	72	105.131			

*$p \leq .025$

Table 97

Errors (Task 2): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	140394.915	1	140394.915	206.002	.000	.892
	Groups	799.388	2	399.694	.586	.564	.045
	Error	17038.037	25	681.521			
Advanced	Intercept	102888.619	1	102888.619	240.437	.000	.870
	Groups	119.935	2	59.967	.140	.870	.008
	Error	15405.264	36	427.924			

The pairwise comparisons results confirmed that the statistically significant interaction at upper-intermediate level was due to the decrease in the mean scores of the upper-intermediate focused instruction group. The focused instruction group had a

statistically significantly higher mean score for their pretest (*M*=51.92, *SD*=26.15, *n*=13) than those for their posttest (*M*=34.92, *SD*=5.22, *n*=13, *MD*=17, *SE*=5.33, *p*=.011) and delayed posttest (*M*=35.38, *SD*=12.76, *n*=13, *MD*=16.54, *SE*=5.29, *p*=.013). None of the other mean differences within the upper-intermediate or advanced groups or between the three upper-intermediate or the three advanced groups at each time point approached statistical significance (*p*>.025).

These results suggested that while focused instruction of formulaic sequences helped the upper-intermediate students reduce the number of errors they made after the training period and retain the same level of performance at a later stage, it did not have any significant effects on the performance of the advanced students. Nor did making formulaic sequences salient in input or focusing on linguistic aspects help the upper-intermediate or advanced EAP students make fewer errors after the training period.

Errors% (task 2). *Descriptive statistics* were also computed on the percentage of errors the six groups had in their essay responses. According to Table 98, the upper-intermediate focused instruction group and the upper-intermediate control group had lower mean scores for their posttests (*M*=18.10, *SD*=5.42, *n*=13 and *M*=21.77, *SD*=10.36,

Table 98

Descriptive Statistics: Errors% (Upper-Intermediate)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	13	13	13	5	5	5	10	10	10
Mean	20.91	18.10	19.41	19.85	22.27	18.93	22.96	21.77	19.83
Median	23.57	16.66	17.70	17.67	19.90	20.63	20.30	20.99	19.41
Mode	11.06[a]	7.89[a]	8.59[a]	13.93[a]	16.16[a]	10.85[a]	14.00[a]	8.51[a]	7.45[a]
SD	6.48	5.42	7.85	5.45	6.30	5.88	9.23	10.36	6.72
Minimum	11.06	7.89	8.59	13.93	16.16	10.85	14.00	8.51	7.45
Maximum	28.30	25.56	32.70	26.05	32.75	26.14	43.26	46.25	31.90

a. Multiple modes exist. The smallest value is shown

n=10) and delayed posttests (*M*=19.41, *SD*=7.85, *n*=13 and *M*=19.83, *SD*=6.72, *n*=10) than those for their pretests (*M*=20.91, *SD*=6.48, *n*=13 and *M*=22.96, *SD*=9.23, *n*=10). The upper-intermediate saliency group, in turn, had the highest percentage of errors in their posttest (*M*=22.27, *SD*=6.30, *n*=5) and the lowest one in their delayed posttest (*M*=18.93, *SD*=5.88, *n*=5).

As for the three advanced groups, they all had lower mean scores for their posttests and delayed posttests than they did for their pretests (see Table 99 below). While the mean scores for the advanced focused instruction group dropped constantly from (*M*=16.11, *SD*=6.21, *n*=19) in the pretest to (*M*=15.34, *SD*=4.89, *n*=19) in the posttest and again to (*M*=10.91, *SD*=3.03, *n*=19) in the delayed posttest, the advanced saliency group had very similar mean scores at the posttest (*M*=14.24, *SD*=6.25, *n*=13) and the delayed posttest stages (*M*=14.89, *SD*=5.68, *n*=13), which were lower than the one they had at the pretest stage (*M*=18.17, *SD*=7.20, *n*=13). The advanced control group also had lower mean scores for their posttest (*M*=12.13, *SD*=4.62, *n*=7) and their delayed posttest (*M*=16, *SD*=6.07, *n*=7) than that for their pretest (*M*=16.61, *SD*=8.58, *n*=7).

Table 99

Descriptive Statistics: Errors% (Advanced)

	Focused Instruction			Saliency			Control		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
N	19	19	19	13	13	13	7	7	7
Mean	16.11	15.34	10.91	18.17	14.24	14.89	16.61	12.13	16.00
Median	15.69	14.56	10.45	17.25	14.60	14.28	13.88	11.36	17.04
Mode	8.23[a]	6.39[a]	6.56[a]	7.14[a]	3.10[a]	4.32[a]	6.55[a]	5.76[a]	6.76[a]
SD	6.21	4.89	3.03	7.20	6.25	5.68	8.58	4.62	6.07
Minimum	8.23	6.39	6.56	7.14	3.10	4.32	6.55	5.76	6.76
Maximum	33.11	26.76	17.51	33.75	26.79	27.09	32.92	17.48	25.92

a. Multiple modes exist. The smallest value is shown

The last *mixed factorial ANOVA* was carried out on the percentage of errors the six groups had in the second task to determine whether the differences shown in Table 98

and Table 99 above were statistically significant. Prior to running the *mixed factorial ANOVA*, several tests were compute in order to identify outliers, assess the distribution of scores, and check the assumptions of sphericity and homogeneity of variance. The *boxplots* and *z scores* generated for these variables revealed the presence of one outlier in the pretest of the advanced focused instruction group (see Table E35 and E36), which was replaced with the next highest score within that variable (i.e., 24.42). Moreover, the results of normality tests and Mauchly's tests indicated that the distribution of the summative scores was approximately normal (see Table F18) and the assumption of sphericity was met (W=.99, x^2=.54 df=2, p=.762). The assumption of homogeneity of variance, in turn, was violated in the delayed posttest $F(5, 61)=2.79$, p=.025 but not in the pretest $F(5, 61)=.85$, p=.518 or the posttest $F(5, 61)=.79$, p=.563. A more conservative alpha level ($p \leq .025$) was, thus, used in the interpretation of all the effects.

The *mixed factorial ANOVA* results, presented in Table 100 below, demonstrated that the main effect of time was statistically significant $F(2, 122)=6.13$, p=.003, partial η^2=.091, indicating that the six groups combined had different percentages of errors at different time points. The highest order three-way interaction was also statistically significant $F(4, 122)=4.70$, p=.001, partial η^2=.134, suggesting that the different percentages of errors at the three production stages were different for different groups at different proficiency levels, which can also be seen in Figure 40 and Figure 41 below. The results of the tests of between-subjects effects demonstrated that the only statistically significant effect was that of proficiency $F(1, 61)=13.45$, p=.001, partial η^2=.181, which implied that the overall percentage of errors made by the upper-intermediate and advanced EAP students was significantly different (see Table 101 below).

Table 100

Errors% (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Tm.	Sphericity As.	165.125	2	82.563	6.128*	.003	.091
Tm.*Gr.	Sphericity As.	42.833	4	10.708	.795	.531	.025
Tm.* Pro.	Sphericity As.	39.692	2	19.846	1.473	.233	.024
Tm.*Gr.*Pro.	Sphericity As.	253.334	4	63.333	4.701*	.001	.134
Error	Sphericity As.	1643.584	122	13.472			

*$p \leq .025$

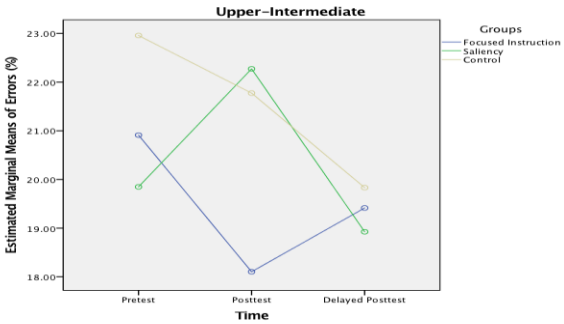


Figure 40. Error% task 2 inter.

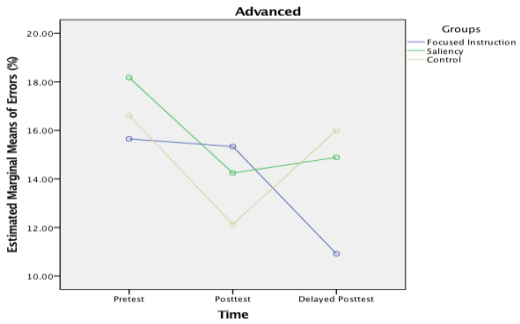


Figure 41. Error% task 2 adv.

Table 101

Errors% (Task 2): Tests of Between-Subjects Effects

	Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intercept	51905.775	1	51905.775	541.974	.000	.899
Groups	94.340	2	47.170	.493	.613	.016
Proficiency	1288.207	1	1288.207	13.451*	.001	.181
Groups*Proficiency	23.885	2	11.943	.125	.883	.004
Error	5842.071	61	95.772			

*$p \leq .025$

In light of the three-way significant interaction and the objectives of this study, the data were split based on proficiency and *a simple-effects analysis* was run. The alpha level was adjusted to $p \leq .05$ because the data, after being split, met the homogeneity of variance assumption at both the upper-intermediate [T1: $F(2, 25)=.72$, $p=.498$; T2: $F(2, 25)=.62$, $p=.547$; T3: $F(2, 25)=1.06$, $p=.362$] and advanced levels [T1: $F(2, 36)=.93$, $p=.405$; T2: $F(2, 36)=.47$, $p=.628$; T3: $F(2, 36)=2.41$, $p=.104$]. The sphericity assumption

was also met at both levels, which was evident in the nonsignificant Mauchly's test results for the three upper-intermediate groups (*W*=.98, x^2=.49, *df*=2, *p*=.782) and the three advanced groups (*W*=.87, x^2=4.81, *df*=2, *p*=.090).

The results of *the simple-effects analysis*, summarized in Table 102 and Table 103 below, demonstrated that both the effects of time *F*(2, 72)=8.58, *p*=.001, partial η^2=.193 and time×group interaction *F*(4 ,72)=5.07, *p*=.001, partial η^2=.220 were statistically significant within the three advanced groups; on the other hand, none of the within-subjects effects for the three upper-intermediate groups or the between-subjects effects for the upper-intermediate or the advanced groups was statistically significant (*p*>.05), as shown in Table 102 and Table 103 below.

Table 102

Errors% (Task 2): Tests of Within-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Time	43.331	2	21.665	1.243	.297	.047
	Time*Group	74.577	4	18.644	1.070	.381	.079
	Error	871.189	50	17.424			
Advanced	Time	184.152	2	92.076	8.583*	.000	.193
	Time*Group	217.563	4	54.391	5.070*	.001	.220
	Error	772.395	72	10.728			

*$p \leq .05$

Table 103

Errors% (Task 2): Tests of Between-Subjects Effects

Source		Type III SS	df	Mean Sq.	F	Sig.	Partial Eta Sq.
Intermediate	Intercept	29953.123	1	29953.123	228.288	.000	.901
	Groups	70.888	2	35.444	.270	.765	.021
	Error	3280.184	25	131.207			
Advanced	Intercept	21953.272	1	21953.272	308.490	.000	.895
	Groups	75.936	2	37.968	.534	.591	.029
	Error	2561.888	36	71.164			

The pairwise comparisons results indicated that the performance of the three advanced groups differed in dissimilar ways. That is, the delayed posttest mean score of

the advanced focused instruction group (*M*=10.91, *SD*=3.03, *n*=19) was statistically significantly lower than their pretest (*M*=15.65, *SD*=5.11, *n*=19, *MD*=-4.74, *SE*=1.04, *p*=.001) and posttest (*M*=15.34, *SD*=4.89, *n*=19, *MD*=-4.42, *SE*=.90, *p*=.001) mean scores. The pretest mean score of the advanced saliency group (*M*=18.17, *SD*=7.20, *n*=13) was statistically significantly higher than their posttest (*M*=14.24, *SD*=6.25, *n*=13, *MD*=3.93, *SE*=1.48, *p*=.036) and their delayed posttest (*M*=14.89, *SD*= 5.68, *n*=13, *MD*=3.28, *SE*=1.26, *p*=.039) mean scores. As for the advanced control group, the percentage of errors they had at the delayed posttest stage (*M*=16.00, *SD*=6.07, *n*=7) was statistically significantly higher than the one they had at the posttest stage (*M*=12.13, *SE*=4.62, *n*=7, *MD*=3.87, *SE*=1.48, *p*=.039). All other differences within the upper-intermediate and advanced groups or between the three upper-intermediate groups and the three advanced groups at each time point were statistically nonsignificant (*p*>.05).

The above-proposed results indicated that while focused instruction of formulaic sequences improved the advanced EAP students' overall accuracy one month after the training period, making formulaic sequences salient in input helped the advanced EAP students write more accurately at both production stages after the training period. A focus on linguistic aspects, on the other hand, did not affect the overall accuracy of the upper-intermediate or advanced EAP students, nor did focused instruction or saliency approaches affect the performance of the upper-intermediate EAP students.

Results summary. An explicit instructional approach to formulaic sequences seemed to help the upper-intermediate students make fewer errors in their posttest and delayed posttest essays, but it did not affect the number of errors they made in their summaries or the percentage of errors they had in both writing tasks. The same

instructional approach helped the advanced EAP students make fewer errors in their posttest and delayed posttest summaries, have a lower percentage of errors in their essays one month after the training period, and make fewer errors than the advanced control group in their delayed posttest summaries. However, it did not affect the number of errors the advanced EAP students made in their essay responses or the percentage of errors they had in their graphical summaries.

Making formulaic sequences salient in input, in turn, helped the upper-intermediate students make fewer errors in their graphical summaries at the end of the training period without any effects on perceived accuracy (i.e., error%) or retention. The same instructional approach helped the advanced EAP students respectively reduce the number and percentage of errors they made in their graphical summaries and essay responses. It also helped them make fewer errors than their peers in the control group when summarizing graphical information one month after the training period. Focusing on different linguistic aspects, on the other hand, did not have any effects on the accuracy of the advanced EAP students in their summary tasks or essay responses. However, it helped the upper-intermediate students make fewer errors in their delayed posttest graphical summaries than their pretest summaries without any effects on their accuracy when responding to essay prompts.

In sum, presenting formulaic sequences, whether by teaching them explicitly or by making them salient in input, seemed to have positive effects, at least to some extent, on the upper-intermediate and advanced EAP students' writing production in that they could make fewer errors at least in one writing task after the training period. These two presentation methods seemed to have positive effects on EAP students' accuracy (i.e.,

error%) at one proficiency level (i.e., advanced) and in one writing task (i.e., essay responses). The overall increased accuracy in the summary tasks cannot be attributed to a single target intervention because it was likely the result of the natural progression of the six groups. Focusing on linguistic aspects also seemed to have some positive effects on the upper-intermediate students, whose errors in the delayed posttest summaries were significantly fewer than their posttest summaries. A focus on linguistic aspects at the advanced level, on the other hand, did not affect the performance of the advanced EAP students in that their performance remained relatively stable over time and they were outperformed by their peers in both experimental groups in the number of errors they made in their delayed posttest summaries.

Research Questions Four and Five

To answer the fourth and fifth research questions that guided this research study, *Pearson correlation coefficients* were computed on pairs of average independent (i.e., the average of frequency and occurrence) and average dependent (the average of words, evaluation, and errors) for both writing tasks in order to determine whether there were any statistically significant positive and/or negative associations between these variables. Prior to computing correlational analyses, nine outlying scores were identified and replaced (see Table E37, Table E38, Table E39, and Table E40), the approximately normal distribution of the average scores within each dependent variable was confirmed (see Table F19), and the linear relationship between each pair of variables was visually assessed (see Figures G1 through G12).

The *Pearson correlation coefficients* results, summarized in Table 104 and Table 105 below, revealed several statistically significant positive associations between

different pairs of variables in both writing tasks. There were statistically significant positive associations between the number of words generated in graphical summaries and the frequency (r=.53, p=.001, n=67) and occurrence (r=.36, p=.003, n=67) of formulaic sequences, suggesting that 27% and 12% of the increase in the number of words generated by the 67 students were respectively accounted for by an increase in the frequency and occurrence of the target formulaic sequences.

Table 104

Correlation Results: Graphical Summaries

		Words		Evaluation		Errors	
Frequency	Pearson Correlation	1	.529*	1	.620*	1	-.202
	Sig. (2-tailed)		.000		.000		.102
	N	67		67		67	
	Association Strength (r^2)	27%		38%		∅	
Occurrence	Pearson Correlation	1	.356*	1	.566*	1	.-.210
	Sig. (2-tailed)		.003		.000		.088
	N	67		67		67	
	Association Strength (r^2)	12%		32%		∅	

*$p \leq .05$

Table 105

Correlation Results: Essays

		Words		Evaluation		Errors	
Frequency	Pearson Correlation	1	.548*	1	.593*	1	.090
	Sig. (2-tailed)		.000		.000		.467
	N	67		67		67	
	Association Strength (r^2)	30%		35%		∅	
Occurrence	Pearson Correlation	1	.462*	1	.567*	1	.027
	Sig. (2-tailed)		.000		.000		.828
	N	67		67		67	
	Association Strength (r^2)	21%		32%		∅	

*$p \leq .05$

Similarly, the number of words generated in the essay responses positively correlated with the frequency (r=.55, p=.001, n=67) and occurrence (r=.46, p=.001, n=67) of the target formulaic sequences, indicating that 30% and 21% of the increase in

the number of words generated in the essay responses could be respectively accounted for by an increase in the frequency and occurrence of the target formulaic sequences.

In addition, there were several statistically significant associations between the average independent frequency and occurrence variables and the average evaluation variable in both writing tasks. In other words, the evaluation assigned by the EAP judge to the graphical summaries was associated with the frequency (r=.62, p=.001, n=67) and occurrence (r=.57, p=.001, n=67) variables, suggesting that 38% of the increase in the use of formulaic sequences in graphical summaries was associated with better evaluation and 32% of the increase in the evaluation the EAP students received for their graphical summaries could be accounted for by an increase in the use of distinct types of formulaic sequences. The EAP judge's evaluation of the essay responses also positively correlated with the frequency (r=.59, p=.001, n=67) and occurrence (r=.57, p=.001, n=67) of the target formulaic sequences, indicating that 35% and 32% of the increase in the frequency and occurrence of the target formulaic sequences were associated with higher evaluation.

There were, however, no statistically significant associations between the frequency and occurrence of the target formulaic sequences and the overall number of errors that the students made whether in their summaries or their essays (p>.05), as shown in Table 104 and Table 105 above. These results suggested that an overall increase in the use of formulaic sequences (whether frequency or occurrence) in both writing tasks was associated with generating higher numbers of words and receiving better evaluation but not with making fewer errors.

Chapter Six: Interpretation and Discussion

The sixth chapter is divided into six sections, each of which aims to put flesh on the statistical results reported in the Findings Chapter and summarized in Appendix H (see Tables H1, H2, and H3). With this goal as a guide, each section recapitulates one major objective of this research study, summarizes the key statistical findings that relate to the objective under investigation, includes some excerpts, raw scores, and/or mean scores for exemplification and clarification purposes, and reveals the links between the key findings of this research study and the existing body of literature and experimental research on formulaic sequences and academic writing.

The Acquisition and Utilization of Formulaic Sequences

The first major objective of this research study was to assess the impact of an explicit instructional approach and a saliency approach to presenting formulaic sequences—two approaches whose pedagogical utility has been repeatedly advocated in the literature (e.g. Byrd & Coxhead, 2010; Wood, 2009, 2015)—on EAP students' ability (a) to acquire and utilize them when completing different writing tasks and (b) to outperform other students who were instructed pedagogical materials other than formulaic sequences. The findings of this study suggested that (a) teaching formulaic sequences explicitly and making them salient in input promoted, at least to some extent, the EAP students' acquisition and utilization of formulaic sequences in their writing production; (b) a focused instructional approach seemed to have stronger effects than a saliency approach on retaining formulaic sequences in the long-term memory, especially at the advanced proficiency level; and (c) the overall use of formulaic sequences in

writing production was affected by both the writing task and the EAP students' proficiency level.

Acquisition. The study results revealed the relative effectiveness of both an explicit instructional approach and a saliency approach in boosting the EAP students' ability to acquire and utilize this language phenomenon in their writing production and in helping them outperform the EAP students who were exposed to other pedagogical materials in at least one aspect.

Focused instruction groups. Teaching formulaic sequences explicitly seemed to have fostered the upper-intermediate students' ability, at least to some extent, to use this language phenomenon in their writing production. The raw scores presented in Table D2 illustrate that of the 13 upper-intermediate students, eight students could utilize greater numbers and more distinct types in their posttest and/or delayed posttest summaries, which can also be seen in the mean scores presented in Table 6 and Table 12. As for the second writing task, while nine upper-intermediate students used more formulaic sequences at least once after the training period (see the raw scores in Table D5 and the mean scores in Table 36), only six students implemented more various types in the essays they produced after the training period, which can be seen in the raw in scores Table D5 and the slightly decreasing posttest and delayed posttest mean scores in Table 42.

Of the above-listed differences, the upper-intermediate students' use of distinct types in their graphical summaries seemed to be the most conspicuous one. The statistical analysis showed that the upper-intermediate focused instruction group could implement statistically significantly more distinct types in their posttest graphical summaries and outperform their counterparts in the control group in two aspects: their use of distinct

types in their posttest summaries and their overall use of distinct types in the summary task, which were both higher than that by the control group.

An explicit instructional approach at an advanced level, in turn, substantially affected the performance of the 19 advanced students. The raw scores in Table D2 indicate that of the 19 advanced EAP students, 17 students utilized more formulaic sequences (both repeated and distinct ones) in their posttest and/or delayed posttest summaries, which is also reflected in the higher mean scores for the advanced focused instruction groups' frequency and occurrence at both production stages after the training period (see Table 7 and Table 13). Similar to the first writing task, the raw scores presented in Table D5 indicate that the majority of the advanced students in the focused instruction group could implement more formulaic sequences (i.e., 15 out of 19 students) and more distinct types (i.e., 16 out of 19 students) after the training period. Their successful utilization of more formulaic sequences in their essay responses after the training period can also be seen in the mean scores listed in Table 37 and Table 43.

The *mixed factorial ANOVA* results confirmed the substantial pedagogical effectiveness of an explicit instructional approach to presenting formulaic sequences at the advanced proficiency level. This was inferred from four statistically significant differences: (1) the statistically significantly higher numbers of formulaic sequences (both repeated and distinct ones) that the advanced focused instruction group used in their posttest and delayed posttest graphical summaries than their pretest summaries, (2) the statistically significantly higher numbers of formulaic sequences they implemented in their posttest essays, (3) the wider range of formulaic sequences they utilized in their posttest and delayed posttest essays, and (4) the statistically significantly greater numbers

of formulaic sequences they produced in their delayed posttest summaries than their peers in the saliency group.

Saliency groups. Making formulaic sequences salient in input also affected the upper-intermediate and advanced EAP students' tendency to use formulaic sequences in both writing tasks. An inspection of the scores presented in Table D3 and Table D6 shows that the five upper-intermediate students used more formulaic sequences and more distinct types in their posttest and delayed posttest summaries. The same was true for the formulaic sequences (both repeated and distinct ones) the upper-intermediate students used in their essay responses, except that only four of the five students used more formulaic sequences after the training period. The increasing tendency to use greater numbers and more distinct types of the target formulaic sequences in both writing tasks was also reflected in the mean scores for the upper-intermediate saliency group (see Tables 6, 12, 36 and 42).

The statistical results revealed several significant differences within the upper-intermediate group and between the former and the other two upper-intermediate groups: (1) the statistically significant increase in the distinct types the upper-intermediate EAP students implemented in their posttest and delayed posttest graphical summaries, (2) the significantly greater number of distinct types the upper-intermediate saliency group utilized in their delayed posttest essays than the one used by the upper-intermediate focused-instruction group at the same production stage, and (3) the greater overall numbers of formulaic sequences (both repeated and distinct ones) that the upper-intermediate students utilized in their essay responses than those used by the upper-

intermediate focused instruction and control groups. All other differences within the upper-intermediate group were not large enough to achieve statistical significance.

As for the advanced group, making formulaic sequences salient in input seemed to have encouraged most of the advanced students to use greater numbers (i.e., 9 out of 13 students) and more distinct types (8 out of 13 students) of the target formulaic sequences in their graphical summaries (see the raw scores in Table D3 and the mean scores in Tables 7 and 13). The same approach, however, had limited effects on the advanced students' use of formulaic sequences in their essay responses. While the raw scores presented in Table D6 shows that seven and 10 students respectively used greater numbers and more distinct types of formulaic sequences in their posttest and delayed posttest essays, these differences had very minimal, if any, effects on the mean scores obtained for the advanced group's essay responses (see Tables 37 and 43). The statistical analysis confirmed the statistically significant increase in the number of formulaic sequences (both repeated and distinct ones) used by the advanced EAP students in their posttest graphical summaries but, at the same time, showed that all other differences were very minor and, hence, did not approach statistical significance.

Control groups. The possible pedagogical utility of an explicit instructional and a saliency approach to addressing formulaic sequences was also deduced from the performance of the upper-intermediate and advanced control groups in both writing tasks. Table D4 and Table D7 show that out of the 10 upper-intermediate students in the control group, nine students could use more formulaic sequences in their summaries and essays after the training period; many upper-intermediate and advanced students could also augment their posttest and/or delayed posttest summaries and essays with more versatile

types (see Tables D4 and D7). However, their increasing tendency to use more formulaic sequences in both writing tasks was quite inconsistent, which was reflected in the relatively similar (see Tables12 and 42), increasing (see Tables 36, 37, 43), and fluctuating (see Tables 6, 7 and 13) mean scores for the upper-intermediate and advanced control groups. It was also evident in the absence of any statistically significant differences within the upper-intermediate and advanced control groups and the statistically significant differences between the upper-intermediate control group and their counterparts in the focused instruction group in the first task as well as the upper-intermediate control group and the saliency group at the same proficiency level in the second task.

Several scholars (e.g. Conklin & Schmitt, 2012; Cortes, 2004, 2006; Ellis et al., 2008; Fan, 2009; Willis, 2003) have cast doubts on the pedagogical efficacy of leaving the acquisition of formulaic sequences to chance encounters. The limited, if any, effectiveness of such an approach is mainly attributed to the underrepresentation of formulaic sequences in EAP textbooks and L2 learners' failure to notice, intuit, and acquire formulaic sequences because of their tendency to view language as composed of individual words, their propensity to disregard complex items in input, and/or their proclivity to utilize familiar ones (Boers & Lindstromberg, 2012; Byrd & Coxhead, 2010; Cortes, 2004; Jones & Haywood, 2004; Murray, 2016; Schmitt & Carter, 2004; Willis, 2003; Wood, 2010a; Wood & Appel, 2014).

The alleged limited pedagogical efficacy of incidental learning has led several scholars to argue in favor of teaching formulaic sequences explicitly or making them salient in input to facilitate their acquisition and internalization into L2 learners'

linguistic repertoire. For example, Boers et al. (2006), Byrd and Coxhead (2010), Nesselhauf (2003), Staples et al. (2013), and Wood (2009) contend that focused instruction of formulaic sequences is likely to increase the chances of acquiring and committing such a complex language phenomenon as formulaic sequences to L2 learners' long-term memory and to ultimately result in promoting L2 learners' tendency to utilize them in their language production. Other scholars (e.g. Bishop, 2004a, 2004b; Cortes, 2004, 2006; Ward, 2007; Wood, 2010b, 2015) stress the pedagogical efficacy of making formulaic sequences salient in input. On the pedagogical utility of making formulaic sequences salient, Bishop (2004a, 2004b), Erman and Warren (2000), Schmitt and Carter (2004), and Wood (2010a) suggest that such an approach may foster L2 learners' noticing of the form and function of formulaic sequences, raise L2 learners' awareness of the importance of utilizing them in their language production, and develop L2 learners' independent learning skills.

The study results are partially congruent with the above-summarized arguments on the pedagogical utility of an explicit instructional and/or a saliency approach to presenting formulaic sequences. Despite the differential effects that the different approaches had on the participants at different proficiency levels, the participants' performance in each of the four experimental groups differed at least in one aspect after the training period, and some groups could outperform their peers in the control or the other experimental group in one or several aspects. The theoretical account of the effectiveness of focused instruction of formulaic sequences in increasing the chances of their acquisition, retention, and utilization (e.g. Byrd & Coxhead, 2010; Wood, 2009) was most apparent in the advanced focused instruction group whose use of formulaic

sequences in both writing tasks substantially increased after the training period and occasionally surpassed the performance of the advanced saliency group. It was also noticeable in the performance of the upper-intermediate group who did not only augment their graphical summaries with more distinct types at the end of the training period but also outperformed their peers in the control group in two different aspects.

Similar to the explicit instructional approach, the possible pedagogical efficacy of making formulaic sequences salient in input (e.g. Bishop, 2004b; Cortes, 2004; Wood, 2015) was obvious at both proficiency levels. In other words, the upper-intermediate group could utilize a wider range of formulaic sequences in their graphical summaries after the training period and outperform the upper-intermediate focused instruction and control groups in their use of formulaic sequences (i.e., both repeated and distinct ones) in the second task. The same approach had moderate effects on the performance of the advanced students in that it affected their use of formulaic sequences (repeated and distinct ones) in graphical summaries at the end of the training period. However, it did not affect their use of this language phenomenon in their essay responses; nor did it help them outperform the advanced control or the other experimental group.

The pedagogical utility of focused instruction and saliency was also deduced from the fairly stable performance of both control groups. Since the two control groups mainly functioned as a baseline in this study, their unaltered performance may be perceived as another piece of evidence that the differences detected in the experimental groups' performance were likely the result of a target pedagogical intervention rather than their natural progression because of being registered in an EAP program. While it is hard to provide a definitive explanation, the control group's relatively stable performance may be

due to the possible underrepresentation of formulaic sequences in the input to which they were exposed during the study, the limited pedagogical efficacy of mere exposure to formulaic sequences in input, or the students' tendency to avoid them in the writing samples they produced during the data collection stages.

Retention. Another interesting finding that is worth reporting in this section relates to the advanced focused instruction group's greater ability to retain the target formulaic sequences for later use than the other three experimental groups. As mentioned in the previous section, the advanced focused instruction group's use of formulaic sequences in the delayed posttest remained very similar to (i.e., occurrence in both tasks) or was even higher than the posttest (i.e., frequency in task 1) or than the delayed posttest of the other experimental group (i.e. the advanced saliency group). When compared to the performance of the other experimental groups and the two control groups, the upper-intermediate saliency group was the only one that could retain the same level of performance (i.e., occurrence in task 1).

The process through which formulaic sequences are successfully acquired is believed to be a protracted and an arduous one and requires multiple exposures to salient input (Conklin & Schmitt, 2012; Durrant & Schmitt, 2010; Ellis, 2012; Laufer & Waldman, 2011; Schmitt, 2008; Verstraten, 1992). Unless salient formulaic sequences are repeatedly encountered in input, L2 learners may fail to acquire them or they may integrate them into their linguistic repertoire but fail to use them productively (Boers & Lindstromberg, 2012; Cortes, 2004; Coxhead, 2008; Verstraten, 1992). One way to accelerate L2 learners' uptake of formulaic sequences is to teach them explicitly (Boers et al., 2006; Byrd & Coxhead, 2010; Hinkel, 2004; Nesselhauf, 2003; Wood, 2009).

In view of the pertinent literature, one potential explanation for the very minimal retention on the part of the two saliency groups is that the limited numbers of exposures to salient formulaic sequences might have negatively affected their ability to retain formulaic sequences: "taking notice of a given word sequence just once or twice is hardly enough to leave durable memory traces" (Boers & Lindstromberg, 2012, p. 89). An alternative explanation is that focused instruction may have stimulated the relatively rapid acquisition and internalization of formulaic sequences into the 19 advanced EAP students' long-term memory because explicitly teaching formulaic sequences is believed to expedite their successful acquisition (Boers et al. 2006; O'Donnell et al., 2013; Staples et al., 2013; Wood, 2009).

While these explanations seem plausible, it is hard to demonstrate or even claim that focused instruction of formulaic sequences is more likely to lead to higher learning gains, for three reasons. First, it is not clear whether the advanced saliency group failed to retain the target formulaic sequences, or they successfully learned the ones they noticed but avoided them in their delayed posttest summaries. Second, the upper-intermediate saliency group's ability to retain the same level of performance cannot be overlooked. Although there were only five participants in the upper-intermediate saliency group and only one statistically significant indication of retention, their ability to use more distinct types in their delayed posttest summaries and to outperform their counterparts in the focused instruction group in their use of distinct types in their delayed posttest essays may be suggestive of some learning gains based on noticing. Last but not least, focused instruction of formulaic sequences did not have any significant effects on the upper-intermediate EAP students' ability to retain the target formulaic sequences, which again

raises questions about the utility of this teaching method in promoting retention at other proficiency levels.

Proficiency level and writing tasks. The statistical findings also revealed some effects of both the proficiency level and the type of the writing task on the participants' use of formulaic sequences. The effect of proficiency level was, first and foremost, evident in the statistically significant effect of proficiency on the 67 EAP students' use of formulaic sequences (i.e., frequency and occurrence) in their graphical summaries and was attributed to higher mean scores for the advanced EAP students than those for the upper-intermediate students. However, different results were obtained for the essay responses; that is, the performance of the upper-intermediate and advanced EAP students, as far as the use of repeated and distinct formulaic sequences is concerned, did not differ significantly.

Research on L2 learners' use of formulaic sequences in academic writing has provided mixed findings. Some researchers (e.g. Chen & Baker, 2010; Paquot & Granger, 2012) found that the use of formulaic sequences is positively associated with language proficiency; that is, students with higher proficiency levels tend to implement more formulaic sequences in their writing production. Other researchers (e.g. Appel & Wood, 2016; Staples et al., 2013) noted that the use of formulaic sequences decreases as L2 learners become more proficient in writing. In other words, L2 learners with limited language proficiency tend to cling onto phrases they have memorized or copied from other sources (e.g. in summary tasks) as production techniques to compensate for their limited lexis and underdeveloped language proficiency (Appel & Wood, 2016; Ohlrogge, 2009; Staples et al., 2013). Since the advanced EAP students used more formulaic

sequences than the upper-intermediate students in their graphical summaries, the results of this study seem to concur, at least in part, with the former findings.

The writing prompt also seemed to have some effects on the participants' tendency to utilize formulaic sequences in their written texts. A glance at the summarized results above reveals an obvious difference in the number of statistically significant findings for both writing tasks. Most of the within-groups differences were detected in the first task (i.e., graphical summaries) of the four experimental groups (i.e., six out of eight), and only two differences were identified in the second writing task (i.e., essay responses). This observation is suggestive of the effects of some complex writing tasks (e.g. reporting graphical information) on L2 learners' tendency to use formulaic sequences and, hence, requires a focus on the characteristics of the writing task and the utility of formulaic sequences as a production tool.

The literature on formulaic sequences suggests that many L2 learners view them as '*zones of safety*,' '*islands of reliability*,' '*islands*,' and/or '*fixed anchorage points*,' which become particularly handy when L2 learners have a relatively underdeveloped lexical repertoire to address a complex subject matter in writing and/or in speech (Boers et al., 2006; Ellis, 2012; Dechert, 1984, as cited in Granger, 1998; Millar, 2011). One such complex task is graph interpretation (Brasell & Rowe, 1993; Glazer, 2011; Mevarech & Kramarsky, 1997). Despite their widespread use in academic materials to visually display complex and dense scientific data in a clear and concise manner, graph interpretation is a daunting process for many students (Brasell & Rowe, 1993; Friel, Curcio, & Bright, 2001; Glazer, 2011; Mevarech & Kramarsky, 1997; Salanti, Ades, & Ioannidis, 2011). The complexity of this process is mainly ascribed to the perplexity of

the information presented in a graph, the differences between science language and everyday language, and the intricacy of transforming one mode of representing information (i.e., *graphical*) into another, that is, *verbal* (Brasell & Rowe, 1993; Glazer, 2011; Mevarech & Kramarsky, 1997).

The complexity of the first task, graph interpretation, may have led the participants in the experimental groups to view the target formulaic sequences as trustworthy resources and utilize them in order to deal with the complexity of summarizing graphical information. While this explanation needs further investigation (e.g. through a follow-up interview or questionnaire), it can be taken as a possible one for the differences between the number of formulaic sequences used in the first writing task and the ones utilized in the second writing tasks. It may also be one of the main reasons for the different results reported in the research studies that attempted to investigate the effects of a particular presentation method on EAP students' writing (e.g. AlHassan & Wood, 2015; Čolović-Marković, 2012; Jones & Haywood, 2004).

The Accurate Utilization of Formulaic Sequences

The second major objective of this research study was to measure the effectiveness of different instructional methods in promoting L2 learners' accurate utilization of formulaic sequences in different writing tasks, which is associated in the literature with increased writing proficiency (e.g. Byrd & Coxhead, 2010; Hyland, 2008a; Jones & Haywood, 2004; Laufer & Waldman, 2011; Lewis, 1997; Paquot, 2008). The results of the *mixed factorial ANOVAs* computed on the numbers and percentages of correct and incorrect formulaic sequences indicated that (1) an explicit instructional approach to formulaic sequences boosted the advanced students' ability to utilize more

accurate formulaic sequences and helped them outperform their counterparts in the saliency group, (2) making formulaic sequences salient in input had minimal effects on the advanced group's performance and no significant effects on the upper-intermediate students' production, (3) focusing on structural aspects did not affect the control groups' accurate production of formulaic sequences, and (4) the participants' proficiency influenced their production of accurate formulaic sequences in at least one aspect in both writing tasks.

Focused instruction groups. An explicit instructional approach to formulaic sequences seemed to have different effects on the upper-intermediate and advanced EAP students. After explicitly teaching them the target formulaic sequences, most of the upper-intermediate students could produce, especially in their posttest summaries, more accurate and fewer erroneous formulaic sequences, which can be seen in the raw scores in Table D2 and the mean scores in Tables 18, 24, and 30. Bonapeite, for example, produced inaccurate stretches of discourse to report changing trends in the pretest (e.g. *was going down*, *between 40 to 20*, *were stables*, etc.), but she utilized several relatively more accurate collocations in her posttest and delayed posttest summaries (e.g. *a tremendous decrease*, *remained stable*, *was fluctuating*, *was slightly decreasing*, etc.).

Other upper-intermediate students, however, could produce more accurate formulaic sequences only at the end of the training period. Zoey, for instance, produced more accurate stretches of discourse in her posttest summary (e.g. *there was a significant decline*, *was stable*, etc.) than her pretest (*Sugar is dramticly declining and then rising*, *sugar is signifacantly increasing*, etc.) and delayed posttest (e.g. *A is dramatic increasing*, *the university B students rise and down significantly*, etc.). Despite their

ability to produce more correct and fewer incorrect formulaic sequences at the end of the training period, none of the differences, except for the decreased accuracy in the number of correct formulaic sequences they produced in their delayed posttest summaries when compared to their posttest, reached statistical significance.

Similar to the graphical summaries, most of the upper-intermediate students could produce more accurate and fewer inaccurate formulaic sequences in their essay responses at both production stages after the training period, as shown in the raw scores presented in Table D5 and the mean scores listed in Tables 48, 52, and 58. In the pretest, Bonapeite, for instance, produced noun phrases based on *the (A) of (B)*, which were, for the most part, erroneous (e.g. *One of the most reason of death*, *two important technology reasons of people inactivities*, etc.). Most of the noun phrases Bonapeite composed in her posttest and delayed posttest essays based on the same formulaic sequence were structurally accurate (e.g. *her domain of study*, *The percentage of people born intelligent*, *the number of intelligent people*, etc.). Victoria, in turn, accurately utilized a summarizer (*All in all*) in her posttest conclusion after using *in a conclusion* in her pretest (see Appendix D, Analyzed Texts). However, the differences in the upper-intermediate students' performance in the second writing task were statistically negligible.

Unlike the upper-intermediate students, an explicit instructional approach to formulaic sequences improved the advanced EAP students' performance in graphical summaries. The raw scores presented in Table D2 show that most of the advanced students had higher numbers and percentages of correct formulaic sequences in their graphical summaries after the training period. The *ANOVA* results, in turn, demonstrated that the advanced EAP students could produce more accurate formulaic sequences at both

production stages after the training period, could have higher percentages of correct formulaic sequences in their delayed posttest than that in their pretest, and could outperform the saliency group in the number of correct formulaic sequences at the third production stage. To illustrate, the stretches of discourse which Leo produced in his posttest (*the percentage of international students remained the same*, *it decreased slightly from 15% to 13%*, etc.) and delayed posttest (*there was a sharp increase*, *the production of rice remained the same*, etc.) were, for the most part, syntactically and/or semantically more accurate than the ones he produced in his pretest (e.g. *is increasing a little bit, the increase is steadly uping*, etc.). Yin, in turn, produced a wider range of semantically more accurate collocations to report graphical summaries at both production stages after the training period (*increased flatly*, *decreased sharply*, *decreased slightly*, etc.) after repeatedly using the same collocate to report different types of trends in his pretest (e.g. *increased a lot*, *did not increase a lot or decrease a lot*, etc.).

The mean scores listed in Table 19, in turn, indicate that the total numbers of correct formulaic sequences produced by the advanced focused instruction group after the training period were obviously higher than the ones produced by the advanced saliency group, with the highest difference at the third stage of data collection, and, hence, they could outperform their peers in the saliency group in the delayed posttest. It is worth noting here that although most of the advanced students could successfully reduce the number of incorrect formulaic sequences after the training period, as can be seen in the raw scores listed in Table D2 and the mean scores in Table 25, the statistical findings indicated that the advanced focused instruction group's production of inaccurate formulaic sequences was relatively stable over time.

Similar to the first writing task, the raw scores presented in Table D5 reflect the ability of most of the participants to produce more accurate formulaic sequences (both numbers and percentages) and fewer inaccurate ones in their essay responses after the training period. Tom, for example, dropped the head noun in the phrases he produced using *the (A) of (B)* at the first production stage (i.e., *the difficulty of Chinese* and *the timetable of Canadian*), an error that was not detected in the noun phrases he produced in his posttest or delayed posttest essays (e.g. *the overuse of technology*, *the negative effects of the overuse of technology*, and *the grades of high school*). Similarly, the noun phrases that Fares produced based on the same formulaic sequence in the posttest and the delayed posttest (e.g. *the regime of the government*, *The isolation of men and women*, *the laws of movement*, *the number of creators*, etc.) did not include usage or word order errors, two issues that were identified in the two noun phrases that he produced in the pretest (i.e., *the side negative of using it* and *the addiction of technology*).

Leo, in turn, used *there is* before both singular and plural nouns in the pretest (*There is an argument* and *There is no cases*), but he accurately utilized *there is* and *there are* at both production stages after the training period (i.e., *there is a country-wide exam called Gao Kao*, *There is no special treatment*, *There are two negative effects,* and *There are two apparent differences*). Dana's utilization of formulaic sequences also became more accurate in that she corrected the punctuation that follows a summarizer; that is, Dana added a period after *in conclusion* in her pretest but a comma after *in sum* in the posttest and after *in conclusion* in the delayed posttest. While these examples, along with the raw and mean scores for the advanced focused instruction group (see Tables D5, 49, 53, and 59), are suggestive of some increased accuracy, the statistical results

demonstrated that none of the above-mentioned differences achieved statistical significance, and, thus, focused instruction of formulaic sequences did not seem to have any significant effects on the advanced EAP students' ability to produce more correct formulaic sequences or to reduce the numbers of incorrect ones in their essay responses.

Saliency groups. Making formulaic sequences salient in input did not have any significant effects on the upper-intermediate students' utilization of accurate formulaic sequences, but it slightly affected the number of correct formulaic sequences that the advanced EAP students produced at the end of the training period. The raw scores (see Table D3) and mean scores (see Tables 18, 24, and 30) for the total numbers and percentages of the correct/incorrect formulaic sequences the upper-intermediate students produced in their graphical summaries revealed very minimal improvement within the saliency group. For instance, the formulaic-based stretches of discourse that Arthur produced in his pretest (e.g. *the parcentage of the University A Students is increased steadly*, *the number of students at University B seems fluctuating*, etc.) were very similar to the ones he produced at both production stages after the training period (e.g. *the nombor of homless people started to encrease signeficntly*, *the production of suger in Country X was droped dramatically… the suger production was jumped rapidally*, etc.) in that they all contained spelling and/or grammatical errors. Very similar issues (i.e., persistent grammatical and spelling errors) were identified in most of the formulaic sequences that Mike produced at the three production stages (see Appendix D, Analyzed Texts).

The raw scores for the upper-intermediate students' production of correct formulaic sequences in their essays (see Table D6), in turn, indicated that most of the

students could produce greater numbers of accurate formulaic-based stretches of discourse at the posttest and/or delayed posttest stages, which was also reflected in the mean scores listed in Tables 48 and 58 for the saliency. An examination of the essays produced by the upper-intermediate students, however, revealed very minimal improvement in their use of formulaic sequences. Mike, for example, correctly utilized *(A) as well as (B)* to combine two noun phrases in the posttest after placing part of this formulaic sequence at the end of his sentence in the pretest (see Appendix D, Analyzed Texts). Moe, in turn, used *there is* before a noun phrase with a plural head noun in the pretest (i.e., *there is no stupid people*), but he correctly used *there are* before plural nouns in the delayed posttest (e.g. *There are many people*, *there are positive effects*, etc.). The production of other formulaic-based stretches of discourse by the other upper-intermediate students remained relatively consistent. The statistical results confirmed the absence of any improvement within the upper-intermediate saliency group in both writing tasks.

The raw and mean scores for the advanced students suggested that most of the advanced EAP students could produce more accurate and fewer inaccurate formulaic sequences in their posttest and/or their delayed posttest summaries (see Table D3 and Tables 19, 25, and 31) and essays (see Table D6 and Tables 49, 53, and 59). The statistical findings, however, demonstrated that the only statistically significant difference within the advanced saliency group was the number of correct formulaic sequences they produced to report graphical summaries in their posttest when compared to the ones they utilized in their pretest. Their production of formulaic sequences in the delayed posttest summaries was relatively inconsistent, which explains the absence of any statistically

significant difference between their pretest and delayed posttest and the advanced focused instruction group's ability to outperform them at the delayed posttest stage.

The advanced saliency group's writing production clearly explicates these results. The formulaic sequences that Franklin utilized in his posttest summary to introduce the line graph (i.e., *The line chart demonstrates*) was more accurate than the ones he produced in the pretest (*In the line chart illustrate above*) and the delayed posttest (*Refer to the line chart shown above*). Similarly, Andrea produced syntactically and/or semantically more accurate stretches of discourse to report graphical changes in her posttest (e.g. *a remarkable flactuation*, *increased dramatically*, *remained the same*, *there is an increase in the percentage of homeless people*, etc.) than her pretest (i.e., *it's trend of consumption decrised*, *we have maximum using of sugar*, *the needs of rice have not changed*, etc.) and delayed posttest (e.g. *remain stable*, *a sligh decrease*, *inflectuated dramatically, there was an inflactuation*, etc.).

The essay responses, in turn, revealed the inconsistency in the advanced saliency group's accurate utilization of formulaic sequences. To illustrate, Franklin correctly used *there is* before singular nouns and *there are* before plural nouns in his pretest and delayed posttest (e.g. *there is a survey*, *there are several traits*, etc.), but he used there are in the posttest before a singular pronoun (i.e., *there are something*). On the other hand, most of the noun phrases that Jinner produced in the pretest and the delayed posttest based on *the (A) of (B)* were inaccurate (e.g. *development of technologies*, *the most scary part of overuse technologies*, *perspecties of teaching*, etc.), and some of the noun phrases he constructed at the three production stages were atypical of English (e.g. *their shields of eyes*, *the following of his life*, and *the abilities of them*).

Control groups. The raw and mean scores for the number and percentage of correct and wrong formulaic sequences which the upper intermediate students produced in their graphical summaries (see Table D4 and Tables 18, 24, and 30) and their essay responses (see Table D7 and Tables 48, 52, and 58) were suggestive of increased accuracy. However, the results of the *mixed ANOVAs* showed that the upper-intermediate students' increased accuracy in both writing tasks was very minimal and, hence, did not reach statistical significance, which can be illustrated with reference to their writing production.

Most of the formulaic sequences that Dead Pool produced in his pretest summary, for example, were syntactically and/or semantically erroneous (e.g. *The diagram indicate*, *had a rapidly increase*, *the rest of two university*, *at 2000 to 2002*, etc.). While Dead Pool could accurately produce some formulaic sequences at the posttest and the delayed posttest stages (*The chart indicates*, *from 2005 to 2010*, *the production of rice and whea*t, etc.), some of the formulaic sequences he produced at both stages after the training period had similar errors to those detected in his pretest (e.g. *were a flucuated*, *was fluctuated*, *there was a dramatic down trend*, *these paragraph shows*, etc.). Aiden, in turn, produced more accurate formulaic sequences in his posttest summary (e.g. *had a fluctuation*, *decreased significantly*, *The production of sugar*, etc.) than his pretest (e.g. *the decrease of that*, *with the 40 percent of international students in 1999*, *were fluctuated*, etc.). However, Aiden's accurate production of formulaic sequences in the delayed posttest significantly decreased (e.g. *a trend of number of people*, *rised dramatically*, *the most amount of people*, *a significat decline*, etc.).

An inspection of the formulaic-based stretches of discourse the upper-intermediate group produced in their essay responses also revealed a relatively minimal and inconsistent improvement. Hazem, for example, utilized *(A) and (B)* to construct noun phrases in his three essay responses. While all the phrases he produced in the pretest contained structural errors (*a different standards and Regulations*, *a first priority and vital*, and *postive and negative in difference*), he could accurately produce two noun phrases based on the same formulaic sequence after the training period (*certain gifts and talents* and *very creative and smart people*), but the third one was inaccurate (i.e., *inventors and scientist*). Another example of the inconsistent performance can be seen in Jonathan's essay prompts. After misspelling an exemplifier in the pretest (i.e., *For exsmple,*), Jonathan correctly spelled *for example* in the posttest and the delayed posttest but dropped the comma after *for example* at the delayed posttest stage.

As for the advanced students, the raw scores presented in Tables D4 and D7 revealed the highly inconsistent scores for the number and percentage of correct and incorrect formulaic sequences within the advanced control group, which is also reflected in the mean scores for the advanced control group's summaries (see Tables 19, 25, and 31) and essays (see Tables 49, 53, and 59) as well as the absence of any statistically significant differences within that group. The three graphical summaries that Charlie produced (see Appendix D, Analyzed Texts) may serve as an example of the inconsistent performance. In other words, Charlie alternatively used *from (A) to (B)* and *during/in (A) to (B)* to refer to price and time ranges at the three production stages. He also incorrectly utilized the passive voice in his pretest and posttest, an error that he seemed to be able to correct at the delayed posttest stage (see Appendix D, Analyzed Texts).

Another example of the inconsistent performance is the writing production of Madara. Most of the formulaic sequences that Madara produced in the three graphical summaries were suggestive of a relatively advanced control over different types of formulaic sequences (e.g. pretest: *The line graph represents*, *declined gradually, there were several fluctuations*, *remained stable between 2000 and 2005*, *the production of wheat and corn*; posttest: *the percentage of homelessness in country B*, *increased steadily*, *fluctuated dramatically during the same period of time*; delayed posttest: *from 40% in 1999 to 60% in 2005, decreased slightly*, *during the same period of time*, *the highest proportion of students*). Madara could also correctly use *it can be seen* in the delayed posttest after inaccurately constructing it in the pretest (i.e. *It can clearly seen*) and the posttest (i.e., *As it can been seen*). Madara, however, made errors in some of the formulaic sequences that he could accurately utilize at the first stage of data collection in both the posttest (e.g. *The line graph reprents, has remained stable*, *period of recession*, etc.) and the delayed posttest (e.g. *there was dramatic fluctuations*).

The essay responses of the advanced control group also include some inconsistency in their performance. For example, all the noun phrases that Crizik composed using *the (A) of (B)* had grammatical and/or spelling errors (e.g. *an seprate part of the society*, *The definations of universities*, etc.). Most of the noun phrases he produced using the same formulaic sequence in both the posttest and the delayed posttest were accurate (e.g. *the development of technology*, *the invention of 3D-printers*, *the invention of IA and VR*, *the results of millions of IQ tests*, etc.). Good's use of *the (A) of (B)* to construct noun phrases at the three production stages, on the other hand, was inconsistent; he produced correct noun phrases at the three production stages (e.g. *the*

health of our next generation, the status of intelligence, *the spirit of a country*, etc.) as well as inaccurate ones (e.g. *every psychological needs of human*, *the source of siginal*, *the reflection of intelligent difference*, etc.). These examples may explain the absence of significant differences within the two control groups.

In addition to the above-listed results and examples, two other points that relate to the participants' performance, irrespective of their groups, are worth highlighting. While none of the differences in the performance of different groups in the essay responses reached statistical significance, the six groups' production of correct formulaic sequences (both numbers and percentages) was significantly more accurate than that before the training period. The results also revealed the effects of the participants' proficiency level on their accurate production of formulaic sequences. The advanced students, irrespective of their groups, could produce statistically significantly more accurate formulaic sequences (both numbers and percentages) than the upper-intermediate students in their graphical summaries. They also had significantly higher percentages of correct formulaic sequences than the upper-intermediate students in their essay responses. However, the number of incorrectly constructed formulaic sequences by the upper-intermediate and advanced EAP students in both writing tasks was relatively similar.

Writing proficiency is not merely contingent on integrating formulaic sequences into written texts; it is rather a by-product of their accurate and proper utilization in writing production because (a) linguistic accuracy is perceived as a major marker of writing proficiency (Ferris, 2002; Hinkel, 2004; Llach, 2011) and (b) erroneous formulaic sequences signal written texts as lacking proficiency (Chen & Baker, 2010; Cortes, 2004). While it stands to reason that L2 writers need to have control over the linguistic

features of the formulaic sequences they utilize in their writing production, such control is not easily attainable in view of the transparent, opaque, and/or idiosyncratic nature of formulaic sequences; the prolonged process of their acquisition; and the effects of L1 interference, L2 overgeneralization, and L2 proficiency on their accurate production (Granger, 1998; Hasselgren, 1994; Kennedy, 2008; Paquot & Granger, 2012; Schmitt & Carter, 2004; Siepmann, 2008; Staples et al., 2013). The importance and complexity of accurately producing formulaic sequences, together with the body of research that empirically revealed L2 writers' failure to do so, have led several researchers to advocate making formulaic sequences salient in input (e.g. Cortes, 2004; Lewis, 2002), providing L2 learners with focused practice and feedback (e.g. Laufer & Waldman 2011; Wood, 2015), or teaching formulaic sequences explicitly (e.g. Byrd & Coxhead, 2010; Nesselhauf, 2003; Wood, 2009) as pedagogical techniques that are likely to help L2 learners gain greater control over this language phenomenon.

The above-mentioned pedagogical techniques are plausible and may indeed result in increased accuracy, yet the study results seem to be partly in agreement with one proposed method (i.e., focused instruction), especially at the advanced level. To elaborate, the advanced focused instruction group, when compared to the other five groups, was the only group whose performance in the first task substantially differed after the training period and who could also outperform the advanced saliency group at the delayed posttest stage. The detected differences in the other five groups' summaries, on the other hand, were either very minimal (i.e., the advanced saliency group) or negligible (i.e., the other four groups). One possible reason why the only group that had substantial improvement was the advanced focused instruction group is that gaining mastery over the

productive use of formulaic sequences is a slow process (Laufer & Waldman, 2011; O'Donnell et al., 2013; Paquot & Granger, 2012) and requires advanced linguistic knowledge, time, and/or explicit instruction (Byrd & Coxhead, 2010; O'Donnell et al., 2013; Paquot & Granger, 2012; Staples et al. 2013).

Another possible, and probably more plausible, explanation for the superior effects of focused instruction at an advanced level is the combined effect of the writing task, the proficiency level, and the method of instruction. As previously mentioned, graph interpretation is a challenging task for many students because of the complex nature of the information presented and the language used in such tasks (Brasell & Rowe, 1993; Friel et al., 2001; Glazer, 2011). The accurate production of such a complex language phenomenon as formulaic sequences, in turn, is believed to be influenced by L2 learners' proficiency level (Staples et al., 2013) and explicit instruction and learning (Nesselhauf, 2003). Taken together, it seems justifiable to assume that both the method of instruction (i.e., focused instruction) and the participants' proficiency level (i.e., advanced) promoted the advanced focused instruction group's accurate production of formulaic sequences in such a complex task as graph interpretation.

The latter interpretation also seems plausible when comparing the performance of the advanced focused instruction group in the first task with that of the upper-intermediate students who were assigned the same method of instruction but did not show any improvement or with the advanced saliency group that showed very minimal improvement. The fact that the performance of the upper-intermediate saliency group and the two control groups remained unaffected also provides further evidence for the latter interpretation; that is, the accurate production of formulaic sequences in such a complex

task as graph interpretation is affected, at least to some extent, by the method of instruction and the students' proficiency level. Another piece of evidence that stresses the crucial effects of proficiency level on the accurate production of formulaic sequences was the overall performance of the advanced EAP students in the first writing task, which was significantly more accurate than that of the upper-intermediate one.

These interpretations are not meant to invalidate the pedagogical efficacy of other techniques (i.e., saliency, corrective feedback, word-focused practice, etc.). They are rather intended to highlight the positive effects a focused instructional approach may have on advanced EAP students' accurate production of formulaic sequences in complex tasks within a limited time period. This is particularly crucial to note, given the performance of the participants in the second writing task. The study results showed that neither focusing on formulaic sequences nor teaching structural aspects could effect, on its own, differences in the performance of a single group over time. Rather, the six groups combined could produce more accurate formulaic sequences (both numbers and percentages) in their posttest and/or delayed posttest essays.

One possible reason why all the groups improved is, again, the writing task. Since the formulaic sequences that were used in the essay responses were, for the most part, typical of general academic writing (i.e., discourse organizers, exemplifiers, etc.), the participants' accurate production may have improved as a result of one or a combination of the following factors: (a) having received the three target pedagogical interventions, (b) having been repeatedly exposed to these formulaic sequences in different materials, (c) having received corrective feedback on their use from different instructors (i.e., their instructors and the researcher), and/or (d) having achieved greater proficiency over time.

All these factors seem possible when taking into account the theoretically expected positive effects of being multiply exposed to rich input over time, achieving higher proficiency, and receiving explicit instruction and/or corrective feedback on L2 learners' productive use of formulaic sequences (Byrd & Coxhead, 2010; Lewis, 2002; Paquot & Granger, 2012; Wood, 2015). Last but not least, the statistical analysis revealed, once again, the effects of proficiency level on the participants' accurate production of formulaic sequences in the second writing, an unsurprising result because of the expected positive effects of advanced proficiency on L2 writers' accurate production of formulaic sequences (O'Donnell et al., 2013; Staples et al., 2013).

Formulaic Sequences and Text Length

The third major objective of this study was to examine the possible positive effects of an explicit instructional and a saliency approach on the number of words the EAP students generate when completing different types of writing tasks—which is associated in the literature with writing proficiency (e.g. González-Bueno & Pérez, 2000; Jones & Haywood, 2004; Lembke et al., 2003; Perkins, 1980; Schoonen et al., 2011)—and compare their performance to other EAP students who were not introduced to the target formulaic sequences. Another objective was to explore any possible association between the overall number of formulaic sequences that the participants, irrespective of their group and proficiency level, utilize and the overall number of words they generate in different writing tasks. The statistical results demonstrated that (a) both focused instruction and saliency had greater impacts on the number of words the EAP students produced in their graphical summaries than those they generated in their essay responses

and (b) an overall increase in the number of formulaic sequences was associated with generating greater numbers of words.

The effectiveness of an explicit instructional approach and a saliency approach to teaching formulaic sequences was, first and foremost, deduced from the upper-intermediate experimental groups' ability to produce fewer numbers of words in their graphical summaries after the training period. The raw scores presented in Tables D2 and D3 reveal the ability of almost all the upper-intermediate participants in both the focused instruction and saliency groups to produce fewer words in their posttest and/or their delayed posttest summaries, which can also be seen in the mean scores for these two groups (see Table 62). The results of the *mixed factorial ANOVA*, in turn, demonstrated that the focused instruction group could produce more concise summaries at both production stages after the training period, and the saliency group could successfully generate fewer words in their posttest summaries. The efficacy of the two presentation methods of formulaic sequences was also inferred from the performance of the upper-intermediate control group. Although several EAP students in the control group generated fewer words in their delayed posttest summaries than they did at the pretest and posttest stages (see Table D4 and Table 62), the decrease was very minimal and, hence, was statistically negligible.

Unlike the results obtained for the upper-intermediate students, the fewer numbers of words generated by several advanced students in the focused instruction and saliency groups (see Tables D2 and D3), together with the lower mean scores for the focused instruction group's posttest summaries and the saliency group's posttest and delayed posttest summaries (see Table 63), were statistically nonsignificant; nor were the higher

numbers of words that several advanced EAP students in the control groups generated at both production stages after the training period (see Table D4 and Table 63), except for the significantly higher numbers of words that the advanced control group generated in their delayed posttest summary when compared to their pretest and posttest. Despite the absence of any statistically significant differences within the two advanced experimental groups, they could outperform their peers in the control group in one or multiple aspects. The overall number of words the advanced saliency group produced in their graphical summaries was significantly lower than the one produced by the control group. The advanced saliency group could also write more concise summaries at the posttest stage than the advanced control group and at delayed posttest stage than both the advanced focused instruction and saliency groups. The advanced focused instruction group, in turn, produced more concise summaries than their peers in the control group at the third production stage.

The minimal pedagogical efficacy of explicitly teaching formulaic sequences and making them salient in input, in turn, was deduced from the very limited statistically significant differences that were identified in the performance of the four experimental groups and the absence of any statistically significant differences between the four experimental groups and the two control groups in the second writing task. The raw scores for the total numbers of words that the six groups generated in their essay responses revealed opposing trends. In other words, several participants in the upper-intermediate focused instruction and control groups as well as the advanced saliency group generated fewer words in their posttest and delayed posttest essays (see Tables D5, D6, and D7), which was also reflected in the lower mean scores for these groups (see

Tables 72 and 73). Many participants in the other three groups, on the other hand, produced more words after the training period (see Tables D5, D6, and D7), which was also associated with higher mean scores for their posttest and delayed posttest essays (see Tables 72 and 73).

The statistical results demonstrated that the only two significant differences within the six groups were in the numbers of words that the upper-intermediate focused instruction group and the advanced focused instruction group generated in their delayed posttest essays when compared to their pretest essays. However, these two differences went in opposite directions, that is, a decrease for the former and an increase for the latter. The *ANOVA* results also revealed the advanced focused instruction group's ability to produce significantly higher number of words in their delayed posttest essays than the advanced saliency group. The results of the Pearson correlation coefficient, in turn, revealed strong positive associations between word counts and formulaic sequence counts. Irrespective of group, proficiency level, and data collection stage, the EAP students who utilized more formulaic sequences, both repeated and distinct ones, generated longer timed summaries and essays.

The above-summarized results corroborate, to some extent, the efficacy of formulaic sequences in promoting writing proficiency. Among the determiners of academic writing proficiency is the writer's ability to abide by the conventions of academic writing and the expectations of the target reader (Hinkel, 2004; Hyland, 2007). A major objective of writing a summary is to concisely communicate a plethora of information, which entails that the process of summary writing is equally based on the writer's ability to select the most important information and to report it concisely in a

written text (Celce-Murcia & Olshtain, 2000; Hidi & Anderson, 1986). While the former is partly influenced by the writer's comprehension of the target material, that latter is partly contingent on knowledge of formulaic sequences (e.g. Hill, 2000; Jones & Haywood, 2004; Laufer & Waldman, 2011; Schoonen et al., 2011). Knowledge of formulaic sequences enables writers, especially L2 writers, to communicate the intended message concisely and efficiently (Hill, 2000; Laufer & Waldman, 2011).

As mentioned above, the experimental groups could produce more concise summaries at one or both production stages after the training period or outperform their peers in the control groups. These results are suggestive that the experimental groups could better abide by the conventions of summary writing than could the control groups. Two major factors have led to this inference. The fact that the six groups were registered in the same program, completed the same summary tasks, and received the same type of written corrective feedback but only the experimental groups could produce more concise summaries after the training period may suggest that integrating formulaic sequences into classroom materials affected the number of words they generated. Moreover, multiple significant differences were detected in the use of formulaic sequences (frequency and/or occurrence) in the graphical summaries of the experimental groups but not the control groups, as explicated in the first section of this chapter. When combined together, these two sets of findings may imply that the groups whose use of formulaic sequences increased after the training period produced more concise summaries.

In addition to writing efficiency, formulaic sequences, as Gledhill (2000), Hyland (2008a, 2008b), and Laufer and Waldman (2011) suggest, may promote writing fluency—which is another perceived indicator of writing proficiency (González-Bueno &

Pérez, 2000; Lembke et al., 2003; Schoonen et al., 2011; Silva, 1993). The findings partly attest to the probable effects of focused instruction of formulaic sequences at an advanced proficiency level on generating longer timed essays and, therefore, writing fluency. In other words, the only group that could produce longer essays after the training period (i.e., in the delayed posttest) was the advanced focused instruction group who also outperformed their peers in the saliency group. When these results are viewed in combination with the advanced focused instruction group's ability to use more formulaic sequences (both repeated and distinct ones) in their essays after the training period and the advanced saliency and control groups relatively stable performance in the same task, one may conclude that using more formulaic sequences in writing production is likely to result in generating longer essays.

A word of caution is in order here. Not all the groups who used more formulaic sequences in their essay responses could generate longer timed essays. In other words, the upper-intermediate saliency group outperformed their peers in the focused instruction and control groups in the overall numbers of repeated and distinct types of formulaic sequences, as explicated in the first section of this chapter. The overall number of words they generated, however, did not differ significantly over time or from the numbers of words generated by the other two upper-intermediate groups. While focused instruction of formulaic sequences may be the reason for generating longer timed texts by the advanced EAP students, the decrease in the performance of the upper-intermediate focused instruction group is likely to suggest that the focused instruction group's ability to produce longer texts and outperform the advanced saliency group at the third stage of data collection may be attributed to other factors that were not explored in this study (e.g.

motivation) or intervening variables that were beyond the researcher's control (e.g. the topics and the nature of writing tasks they discussed/practiced with their instructors).

The statistical findings also corroborated the possible association between the overall use of formulaic sequences and writing fluency (Gledhill, 2000; Hyland, 2008a, 2008b; Laufer & Waldman, 2011), which was inferred from the statistically significant positive association between the average use of formulaic sequences (i.e., frequency and occurrence) in both writing tasks and the average number of generated words. It is worth noting here that based on the above-discussed *mixed factorial ANOVA* results for the graphical summaries, a statistically significant negative correlation between formulaic sequences and words would be intuitively expected. However, a restatement of the major objective for running correlation coefficients and the nature of correlation tests may help explain these results.

The correlation coefficients were computed to reveal any statistically significant associations between formulaic sequence counts and word counts. Such an objective necessitated, as recommended by Bland and Altman (1995a, 1995b), redistributing the target scores to eliminate the effects of time (i.e., multiple stages of data collection) by generating average scores for each EAP student's scores within the target variables. It also entailed collectively analyzing the performance of the EAP students in different groups and at different proficiency levels, that is, without splitting the data based on groups and proficiency levels. Moreover, correlation statistical tests reveal association rather than causative relations between variables (Dörnyei, 2007; Vogt, 2007). Accordingly, the significant positive associations between formulaic sequence counts and word counts suggest that the increasing use of formulaic sequences was associated with

generating more words rather than implying that using more formulaic sequences resulted in generating longer texts. The latter interpretation is not only methodologically invalid but also results in inferences that are contradictory in nature when viewed in relation to the *mixed factorial ANOVA* results.

Formulaic Sequences and Overall Linguistic Accuracy

The fourth major objective of this research study was to identify any positive effects of teaching formulaic sequences explicitly and/or making them salient in input on EAP students' ability to make fewer errors, which is another major marker of writing proficiency (Ellis, 2015; Perkins, 1980; Polio, 1997), compare their performance to other EAP students who were instructed with structural materials, and reveal any possible associations between error counts and formulaic sequence counts. The statistical results indicated that (a) an explicit instructional and a saliency approach had differential effects on the numbers and percentages of errors that upper-intermediate and advanced EAP students made in both writing tasks; (b) focusing on structural aspects had almost no positive impact on the upper-intermediate and advanced students' ability to produce texts with fewer errors; (c) the writing task and the proficiency level affected the participants' performance; and (d) the target formulaic sequences did not correlate with errors.

As for the focused instruction groups, the raw scores for the total number and percentage of errors presented in Tables D2 and D5 indicated that most of the upper-intermediate and advanced EAP students made fewer errors in their posttest and/or delayed posttest graphical summaries and essay responses. The lower numbers and percentages of errors after the training period were also reflected in the mean scores for the focused instruction groups, which were, for the most part, lower after the training

period in both their graphical summaries (see Tables 82, 83, 88, and 89) and essay responses (see Tables 92, 93, 98, and 99). The *mixed factorial ANOVA* results, in turn, confirmed that the upper-intermediate students made fewer errors in their essay responses at both production stages after the training period and the advanced students made fewer errors in their posttest and delayed posttest summaries and had a lower percentage of errors in their delayed posttest essays. The statistical analysis also revealed the advanced focused instruction's group ability to make fewer errors in their delayed posttest summaries than their counterparts in the control group. All the other differences that were detected within the two focused instruction groups were found to be statistically nonsignificant.

Unlike the focused instruction groups, the performance of the saliency groups differed in the two writing tasks. The raw scores (see Table D3) and mean scores (see Tables 82, 83, 88, and 89) for the number of errors that the upper-intermediate and advanced students made in their graphical summaries revealed most of the participants' ability to make fewer errors at least at one stage after the training period. On the other hand, the raw scores in Table D6, together with the mean scores in Tables 92, 93, 98, and 99, showed that while most of the advanced EAP students had lower scores for the errors they made, both numbers and percentages, in the essays they produced after the training period, the upper-intermediate students generally had higher numbers but fewer percentages of errors in the essays they produced after the training period.

The *mixed factorial ANOVA* results revealed three statistically significant differences in the performance of the two saliency groups: (1) the statistically significantly fewer errors that the upper-intermediate students made in their posttest

summaries than their pretest summaries, (2) the statistically significantly fewer errors the advanced saliency group made in their posttest and delayed posttest summaries (i.e., error counts) and posttest and delayed posttest essays (percentages), and (3) the advanced saliency group's ability to make fewer errors in their summaries than the advanced control group at the delayed posttest stage.

The scores for the numbers and percentages of errors that the control groups made in both writing tasks, in turn, were highly inconsistent. In other words, while the raw scores in Tables D4 and D7 indicated that several upper-intermediate students and most advanced students made fewer errors in both writing tasks after the training period, the mean scores for the numbers and percentages of errors that both groups made in both writing tasks either decreased at a single data collection stage after the training period, declined at both production stages after the training period, or increased (see Tables 82, 83, 88, 89, 92, 93, 98, and 99). The *mixed factorial ANOVA* results, however, showed that the only group that had higher proficiency after the training period was the upper-intermediate one, which was inferred from the statistically significantly fewer errors they made in their delayed posttest summaries than their posttest. A second statistically significant difference was detected within the advanced control group. Unlike the former, however, it revealed decreased proficiency because of the statistically significantly higher number of errors that the advanced students made in their delayed posttest essays than their posttest essays.

The *mixed ANOVA* results also revealed the partial effects of proficiency level on the errors the participants made; that is, when compared to the upper-intermediate students, the advanced EAP students, irrespective of their groups, had significantly lower

scores for the errors they made in their graphical summaries (percentages) and their essay responses (both numbers and percentages). The correlation coefficient results, in turn, revealed that the number of errors did not correlate significantly with the frequency or occurrence of the target formulaic sequences. In other words, an increase in using formulaic sequences was not associated with a decrease in making errors.

Linguistic accuracy, which is frequently investigated based on such indicators as counts of errors and/or error-free stretches of discourse in written texts, is perceived as a major and an objective determiner of writing proficiency (Ellis, 2005a, 2015; Ferris, 2002; Ishikawa, 1995; Perkins, 1980). Given its central role in defining L2 learners as proficient writers, SLA researchers have suggested different strategies to promote L2 learners' linguistic accuracy. Among the suggested remedial strategies is explicitly teaching the linguistic aspects that L2 learners repeatedly violate despite being immersed in rich input and/or involved in communicative activities (e.g. Butler, 2002; Ellis, 2005c; Harper & de Jeong, 2004; Willis, 2003). Providing L2 learners with corrective feedback and/or engaging them in editing practice are also among the recommended instructional strategies (e.g. Ellis, 2015; Ferris, 2002; Ferris & Roberts, 2001; Gass & Mackey, 2007; Hyland, 2003a). Other effective strategies include focusing on formulaic sequences, whether by making them salient or teaching them explicitly, and providing L2 learners with extensive practice and/or corrective feedback on their use (e.g. Coxhead & Byrd, 2007; Ellis, 2005a; Fan, 2009; Granger, 1998; Hinkel, 2004; Wood, 2015).

While the above-listed instructional strategies, as their proponents suggest, are likely to have profound impacts on L2 writers' linguistic accuracy, the results of this study were somewhat congruent with the pedagogical effectiveness of the latter. The

effectiveness of focusing on formulaic sequences in promoting the participants' linguistic accuracy was generally extrapolated from the four experimental groups' ability to reduce, at least to some extent, the number of errors they made in both writing tasks after the training period. While the experimental groups' increased accuracy was demonstrated statistically, its source is relatively equivocal. The four experimental groups may have made fewer errors as a result of developing a rich repertoire of formulaic sequences after the training period, which is believed to help L2 learners write more accurately because they retrieve and utilize these items as single units (Ellis, 2005a; Hill, 2000; Schoonen et al., 2011; Willis, 2003). Their increased linguistic accuracy might also have been the result of combining written corrective feedback with focusing on formulaic sequences, which may promote accurate language production (Coxhead & Byrd, 2007; Wood, 2015). It might be attributed to learning formulaic sequences during the training period and receiving grammar instruction from their instructors, a combination that may promote linguistic accuracy (Ellis, 2005a; Lewis, 2008), or the higher proficiency they achieved over time.

Another factor that contributed to inferring the efficacy of formulaic sequences in promoting linguistic accuracy was the performance of the two control groups. The two control groups were explicitly instructed in several linguistic aspects during the training period and were provided with written corrective feedback before they completed their delayed posttests. Such an approach to teaching academic writing is presumed to boost, at least partially, accurate linguistic production (Ellis, 2005a, 2015; Hyland, 2003a; Willis, 2003). However, the statistical results showed that the upper-intermediate control group exhibited a very limited increased accuracy and the advanced control group's writing

production remained unaffected. The results also revealed the significantly greater number of errors the advanced control group made in their delayed posttest essays than they did in their posttest and than those made by their peers in the advanced focused instruction and saliency groups in their delayed posttest essays.

Not only do these findings reveal the very limited effects that focusing on structural aspects and providing corrective feedback had on the upper-intermediate and advanced students' linguistic accuracy, but they also corroborated the efficacy of different approaches to presenting formulaic sequences in promoting the EAP students' accurate language production. These interpretations, however, remain tentative in light of the overall increased accuracy that the six groups combined exhibited in their posttest summaries when compared to their pretest summaries and the absence of any statistically significant differences in the percentage of errors for each of the six groups in the summary task.

The results of the study also revealed the profound effects of language proficiency on the EAP students' linguistic accuracy, with higher linguistic proficiency at an advanced proficiency level. Such inference was drawn based on the performance of the six groups in both writing tasks, in that most of the increased linguistic accuracy was detected at the advanced level. Another indicator of the significant effect of proficiency on EAP students' accurate language production was the overall greater error counts the upper-intermediate students had than the advanced EAP students in the graphical summaries (i.e., the percentage of errors) and essay responses (i.e. both numbers and percentages of errors). These results are, once again, reminiscent of the association between higher proficiency and increased accuracy (e.g. Ellis, 2005c). The study results

also indicated that although different presentation methods of formulaic sequences helped the participants, at least partially, to produce more accurate texts after the training period, the fewer errors they made did not correlate with their use of formulaic sequences.

The Greater Control over Other Markers of Writing Proficiency

The summaries and essays that were collected from the six groups at the three production stages also revealed the increased awareness of some upper-intermediate and advanced EAP students of textual elements (i.e., topic and concluding sentences) after the training period and their tendency to creatively utilize formulaic sequences as production strategies in both writing tasks, which are perceived as other determiners of academic writing proficiency (Cowie, 1992; Erman & Warren, 2000; Mitchell, 2004; Silva, 1990; Willis, 2003).

Textual constituents. An inspection of the textual data revealed a greater awareness of the basic paragraph elements on the part of several upper-intermediate and advanced EAP students after the training period, especially in the first writing task. The textual data also reflected the very limited and sometimes no effects of the training period on the performance of some upper-intermediate and advanced students.

Focused instruction groups. In the pretest, several upper-intermediate and advanced EAP students in the focused instruction groups started and/or concluded their graphical summaries with supporting sentences, an issue that some students could overcome after the training period. To elaborate, several upper-intermediate EAP students produced a concluding sentence that collectively reiterated the graphical changes after the training period. For example, after focusing on a single trend in the last sentence of his pretest (*The highest rate is in 2000 which is 11%*), Brian summarized the graphical

information at the end of his posttest (*In conclusion, the percentage of international students in university A and B had a increased trend while C and D had a declined trend*) and delayed posttest (*The line paragraph shows the four production in country X. The production wheat and corn kept increasing while the production of sugar rose and fell. Rice remained the same*). However, other participants did not add concluding sentences to any of the paragraphs they produced (e.g. Victoria, John, Sara, etc.) or inconsistently produced concluding sentences (e.g., Zoey, Amy, Curry, etc.).

Several advanced students had some control over the basic elements of a paragraph after the training period, that is, topic sentences (i.e., Yin and Mary 2) and concluding sentences (i.e., Mary, Kareem, Leo, and Fares). For instance, after starting his pretest summary with a supporting sentence (*Form this chart we can know the country production of wheat and corn was lower than the sugar*), Yin produced a topic sentence at the beginning of his posttest (*This line chart shows that the rate of international students in four university*) and delayed posttest (*The line chart shows the number of homeless people in four different countries*). Fares added a concluding sentence at the end of his posttest (*In conclusion, There are four exports produced by Country X between 1995 and 2010*) and delayed posttest (*In conclusion, it is clear that the percentage of homeless people in Country B and Country D increased while it in Country A and Country C decreased from 1940 to 2000*) after focusing on two trends in the last sentence of his pretest (*When it comes to the other universities, University C and University D show decreasing over the six years*). Similar to the upper-intermediate students, the performance of other advanced students was relatively inconsistent in that they produced

topic sentences (e.g. Joker and Fill) or concluding sentences (e.g. Amanda, Fill, Mary 2, etc.) in some paragraphs but not others.

The essay responses that the upper-intermediate and advanced EAP students produced illustrated the growing awareness of students at both proficiency levels of topic sentences and the advanced EAP students greater control over concluding and/or transition sentences. Several upper-intermediate students started their posttest and delayed posttest body paragraphs with a topic sentence after dropping it in their pretest (e.g. John, Amy, and Bonapeite). John, for example, wrote two sentences in response to the pretest essay prompt (*Education system improvement is the trend that every country spends a lot of mony to reach satisfing level. Sadi Arabia government has spent over than billions to improve the education system*). In the posttest and the delayed posttest, however, he attempted to write multiple paragraphs with topic sentences for each body paragraph:

T2 *Many intlectual people who invented several revoloutionary technologies that, changed the world, did not complet their studying.…To be a successful in any field, it is better to live it and work hard to achive the goals that you have made previousely that helps you in your dedication.*

T3 *Lonlieness is one of the biggest negative effects that cause a person to be seperated from his/her society then distroys his/her life entierly…. Addiction is very huge dilemma that always facing societies.*

The excerpts above represent the first sentence of each body paragraph that John produced after the training period. These sentences were deemed to be topic sentences because they were semantically related to the subsequent sentences within each paragraph. The upper-intermediate students' use of concluding or transition sentences was inconsistent and/or unaffected. In other words, several students did not conclude any of the body paragraphs they produced at the three production stages (e.g. Victoria, Paul, Bonapeite, etc.); the performance of other students was relatively inconsistent in that they

added concluding and/or transition sentences to some of the body paragraphs they produced but not others (e.g. John, George, Sara, and Curry).

An examination of the essay responses that the advanced students produced revealed the overall tendency of several students to start and/or conclude all the body paragraphs they produced after the training period with topic and concluding sentences (e.g. Maya, Dana, Amanda, Ana, Mary, etc.). After starting her body paragraphs with supporting sentences, Amanda, for example, added topic sentences to the body paragraphs she produced after the training period, which can be seen in the first sentences of the paragraphs she produced at the pretest stage and the ones she constructed at the beginning of each body paragraph she generated after the training period:

T1 *Firstly, on the scientific bases, people only develop one of millions their brain in their whole lives, so if a person strives to think and learn for his or her brain, he or she also can be intelligent…. Other than the angle of Science, it is easy for us to find that people are intelligent when they learn more.*

T2 *One of the negative effect of technology addiction is overweight…. Another neigative effect of technology addiction is that people can be isolated.*

T3 *The first difference in the education system between Canada and China is that students have different levels of studying pressure…. The second different element in the education system between Canada and China is the way of evaluating students.*

Maya, in turn, attempted to conclude the body paragraphs she produced after the training period, an attempt that was not present in her pretest. This can be seen in the last sentence of the single body paragraph Maya produced at the pretest stage, which introduced a supporting detail, and the last sentences of the body paragraphs she produced after the training period, which briefly restated the main points of discussion:

T1 *Many factors may interven with the type of courses that had been given over the semstar.*

T2 *Overall technology is main enemy for socail activity…. Technology plays important role twoard people behivour.*

T3 *This example shows that people can have different level of Intelligance…. Smartness*

is important in daily life, and it is a tool to for easy and successfullife.

Other participants, however, inconsistently dropped the topic sentences or the concluding sentences (e.g. Mia and Silva) from the body paragraphs they produced after the training period.

Saliency groups. Similar to the focused instruction groups, some upper-intermediate and advanced students in the saliency groups showed a slightly increased awareness of paragraph constituents, especially in the first writing task. For instance, Eric, an upper-intermediate student, started his pretest graphical summary with a sentence that focused on a single trend (*The statistics shows us that in University A, the population of international students has increased in the last six years*...). After the training period, Eric seemed to be more aware of the importance of beginning a paragraph with a topic sentence, which can be clearly seen in the first sentence he produced in his posttest (*The line graph shows it, that production in country X of some basic foods has been changing in a period of fifteen years*) and delayed posttest (*The graphic line shows a percentage of Homelessness between 60 years since 1940 to 2000 in four different countries*). Arthur, in turn, concluded his posttest summary (*To sum up, the graph illustrated that the over all idea is that only two countries has reached the highest percentage of homlessness and the other two countries has the lowest percentage of homelessness*) after ending his pretest with a supporting sentence (*On the other hand, the parcentage of international students in university C and D was 10 percent…*). Arthur's delayed posttest, however, was very similar to his pretest (*The production of which were remaining steady from 2005 till to 2010*). Other students did not include a topic sentence (i.e., Aleed) or a concluding sentence (i.e., Aleed, Eric, and Moe) at any of the production stages.

The advanced EAP students' graphical summaries, in turn, revealed the increasing tendency of several students to produce the basic elements of a paragraph after the training period. In other words, some participants who did not conclude their pretest summaries produced concluding sentences in their posttest and/or their delayed posttest (e.g. Nart, Franklin, Andrea, Zia, etc.). Such a growing awareness can be seen in Nart's graphical summaries; that is, he concluded his pretest summary with reference to a single trend but added a concluding sentence to his posttest and delayed posttest summaries (see Appendix D, Analyzed Texts). Other advanced (i.e., Kristen and Hank) did not conclude any of their graphical summaries.

The essay responses, in turn, revealed the growing tendency of the upper-intermediate students to produce topic sentences after the training period, a tendency that did not seem to be present at the advanced level. The essays produced by Mike at the three production stages show his increased awareness of the characteristics of a topic sentence. The first sentences of the two body paragraphs that Mike produced at the pretest stage were respectively too general and too specific and, hence, did not reflect the content the subsequent sentences; the sentences he produced at the beginning of each body paragraph after the training period, on the other hand, had the characteristics of topic sentences (see Appendix D, Analyzed Texts). Mike, however, did not conclude any of the body paragraphs he produced at the three production stages, nor did Arthur or Moe. Eric and Aleed, in turn, produced concluding sentences for some body paragraphs but not others. The advanced students' performance in the second writing task was relatively consistent. Several students started the body paragraphs they produced with a topic sentence (e.g. Lulu, Lee, Jinner, Riki, etc.); other students did not add a concluding

or a transition sentence to any of their body paragraphs (e.g. Nart and Zia). Other students, however, inconsistently added topic and/or concluding sentences to some body paragraphs but not others (e.g. Lucian, Hank, etc.).

Control groups. The graphical summaries produced by the upper-intermediate and advanced control groups also indicated that some participants became more aware of the importance of including topic and/or concluding sentences after the training period. Several upper-intermediate students could successfully produce topic and/or concluding sentences in their posttest and/or delayed posttest summaries (i.e., Alex, Laila, Jessy, and Todd). The following excerpts from Todd's graphical summaries, which represent the first and last sentences he produced at each production stage, may serve as examples to illustrate his increasing awareness of paragraph elements and the task itself over time:

T1 *According to the graphic chart, there was a huge significant change from 1995 to 2000 shows the decreasing of sugar for country X production…. However the amount of sugar is very unstable and it tells that might some earthquake or other deasters happened and it totally destroyed sugar factories or the environment for the plants' growth which provide sugar.*

T2 *The line chart present the number of homelessness in 4 different countries between 1940 and 2000…. Therefore, country A must B North Korea. For America, people had competition every time and most of people could not find jobs, or they enjoyed being homeless, but for north Koreans, the plicis supported everyone to have home and jobs.*

T3 *The graph shows the differences about the percentage of international students between 4 different university from 1999 to 2005…. Personaly, I would like to suggest international students to go to university A, because it is more friendly and more skilled to solve different issues.*

It is interesting to note that none of the upper-intermediate students consistently dropped topic or concluding sentences at the three production stages.

Similar to the upper-intermediate control group, some participants in the advanced control group were more aware of paragraph elements after the training period (i.e., Madara and Good). To illustrate, Madara successfully summarized the graphical

trends in the last sentence of his posttest summary (*Overall, the proportion of homeless people was high in countries B and D, whereas the lowest figures were noted in countries C and A*) and delayed posttest summary (*Overall, it is clear that among these four universities, the highest proportion of students is noted at university A, while the number of undergraduates at other universities was changeable or remained stable with further gradual decrease*) after ending his pretest summary with a supporting sentence (*In the contrast, there were not discrepancies in the production of wheat and corn during given period of time*). Other participants (i.e., Anita and Crizik) did not conclude any of the graphical summaries they produced.

As for the second writing task, the essay responses that Laila and Aiden produced reveal their tendency to include topic sentences to the body paragraph they produced at a single production stage after the training period. The excerpts below, which represent the first sentence of each body paragraph Laila produced at the three production stages, provide an example:

T1 *In my country Saudi Arabia, the government is responsible for education…. Another thing, is that in Canada the education system has rules that are strict and cannot be overrules, for example the attendance policy or the grading system*

T2 *Sitting too long to watch a movie or play a video game can have serious effects on someone's vision and body…. Technology addiction also causes mental illnesses.*

T3 *Many kids are not given the opportunity to explore what their mind can do.*

The first sentence of each body paragraph Laila produced in the pretest focused on a single country, while all the subsequent sentences compared the education system in her country to that in Canada. At the posttest stage, Laila introduced the content of both body paragraphs with topic sentences, but she dropped the topic sentence again in the delayed posttest. The other participants either consistently produced topic sentences but not concluding sentences (e.g. Wukong, Hazem, etc.) or inconsistently dropped topic and/or

concluding sentences (e.g. Alex, Dead Pool, Jessy, Jonathan, Christine, etc.) in their essays.

Madara's essay responses also exemplify the increasing awareness of paragraph elements at the advanced level, especially topic sentences. Unlike the sentences Madara produced at the first data collection stage, which focused on a single country but were followed by comparisons, the sentences he produced at the posttest and the delayed posttest stages introduced the subsequent sentences within each body paragraph:

T1 *First of all, it is important to note that Canadian educational system has many benefets…. Secondly, the education process in Canada includes extraordinary methods of absorbing knowledge by going through brain storms, team-based learning and recreational activities*

T2 *First of all, I would like to say that some scientists claim that sometimes rare human beings can be born with hereditary talents…. Secondly, on the other hand, it should be noted that our intelligence is determined by our social environment: family, friends and colleagues*

T3 *First of all, if people overuse their gadgets, their spine and muscles can be suffered from it because physical activity is replaced by electronic devices…. Secondly, technology addiction may be associated with cognitive disorders in humans particularly among children*

The performance of all the other participants in the advanced control group was relatively inconsistent in that they dropped either topic or concluding sentences from at least one body paragraph.

In the 19th century, several traditional approaches to L2 writing instruction (e.g. *the traditional approach* and *the current-traditional rhetoric*) stressed the importance of focusing on topic and concluding sentences in the language classroom (Hyland, 2007; Mitchell, 2004; Raimes, 1991; Seow, 2002; Shih, 1986; Silva, 1990). This emphasis stemmed from viewing topic and concluding sentences as the basic elements that constitute an academic English paragraph and as *rhetorical strategies* that promote coherence as well as smooth transition and, hence, are suggestive of increased writing

proficiency (Bamberg, 1983; D'Angelo, 1986; Mitchell, 2004; Siepmann, 2006; Silva, 1990; Welch & Link, 1992). However, this orthodoxy has faced serious challenges since the early 20th century based on two major arguments: many expert writers do not abide by these elements and the view of a paragraph as being sequentially composed of a topic sentence, supporting sentences, and a concluding sentence is not only outdated but also misguided (D'Angelo, 1986; Welch & Link, 1992).

Despite these challenges, several researchers argue in favor of the pedagogical utility of focusing on topic and concluding sentences in the language classroom, especially when L2 learners' writing skills are relatively underdeveloped (Bamberg, 1983; D'Angelo, 1986; Welch & Link, 1992). Such an approach is particularly helpful because L2 learners can use these elements as strategic devices to make their writing more organized, coherent, and comprehensible; that is, they help L2 writers divide their texts into meaningful and coherent discursive units, prepare the reader for the subsequent discourse elements, and reiterate the major discussion points (Bamberg, 1983; D'Angelo, 1986; Welch & Link, 1992). On the effectiveness of teaching topic sentences, D'Angelo (1986) argues that "[s]ince composition characteristically deals with the discovery of the available means of exposition and persuasion, it makes little sense not to teach the available means simply because some writers do not choose to avail themselves of those means" (p. 439).

The textual data revealed that the upper-intermediate and advanced EAP students' performance was partially in accord with the literature on the efficacy of focusing on topic and concluding sentences in promoting their utilization in L2 writing production (e.g. D'Angelo, 1986; Mitchell, 2004; Siepmann, 2006; Silva, 1990). The examples

provided above revealed the tendency of some EAP students in each of the six groups to utilize topic and/or concluding sentences as *rhetorical strategies* in their writing production after the training period. In other words, they produced topic and concluding sentences to respectively introduce and reiterate the major points of discussion within a paragraph. While the inconsistent use or absence of topic and concluding sentences in the second writing task might be perceived as a marker of underdeveloped writing skills, caution should be taken when interpreting these observations. Since some expert writers produce topic and/or concluding sentences to some paragraphs rather than others or drop these elements altogether (D'Angelo, 1986; Mitchell, 2004; Siepmann, 2006; Welch & Link, 1992), it is neither accurate nor logical to describe some of the EAP students' inconsistent use of topic and concluding sentences at the three production stages as a marker of underdeveloped writing skills. They can be rather seen as other variations of academic writing to which the students were exposed and by which they were influenced.

The creative use of formulaic sequences. During the data quantification stage, a number of creatively used formulaic sequences in both writing tasks was identified within each of the six groups. Although inspecting the creative use of formulaic sequences was beyond the scope of this research study, the recurrence of creatively utilized formulaic sequences in the textual data collected from the six groups and the proposed concepts in the literature regarding the creative use of formulaic sequences as a production technique and a marker of language proficiency (e.g. Erman & Warren, 2000; Willis, 2003) made those formulaic sequences worth consideration. The creative use of formulaic sequences was mainly based on combining two or more formulaic sequences to form one, extending

or changing part of a formulaic sequence, adding a modifier, and/or changing the grammatical class of a formulaic sequence.

When reporting graphical summaries and responding to essay prompts, the participants in the upper-intermediate and advanced focused instruction groups constructed a single formulaic sequence by combining two distinct ones (e.g. *the (A) of (B)+(A) and (B)→ the production of wheat and corn*) and sometimes three different ones (e.g. *in the (A) of (B)+(A) and (B)+between (A) and (B)→ in terms of the number of students in schools and the relationship between students and teachers*). They also adapted formulaic sequences (e.g. *both (A) and (B)→Country A and Country B both*; *both of these→both of them*; *it is possible→it is impossible*), added modifiers (e.g. *it can be* ***clearly*** *seen*, ***dramatic and sharp*** *fluctuation*, *the* ***main*** *reason*), and extended some formulaic sequences (e.g. *(A) and (B)→(A), (B) and (C)* as in *work, family and social communication* or *(A), (B), (C), and (D)* as in *rice, wheat, sugar and corn*). Changing the part of speech of a formulaic sequence, in turn, was mainly utilized when reporting graphical information (e.g. *a sharp increase/decrease→ increased/decreased sharply*).

Similarly, the upper-intermediate and advanced saliency groups' creative use of formulaic sequences was consistent and much the same as that of the focused instruction groups. In both writing tasks, the saliency groups combined two formulaic sequences to construct one (e.g. *the (A) of (B)+(A) and (B) →the consumption of wheat and corn*; *two different types of economy and power*), partly utilized formulaic sequences (e.g. *decreased dramatically→ to drop dramatically*; *both of these→both of them*; *it is possible→it is impossible*), and inserted a modifier (e.g. *remained* ***relatively*** *stable*; *it is* ***more*** *difficult*). They also created longer formulaic sequences based on shorter ones (e.g.

(A) and (B)→(A), (B) and, (C) as in *social activities, video games, and visual pollution* or *(A), (B), (C), and, (D)* as in *rice, wheat, sugar, and corn*) and changed noun phrases into verb phrases, particularly in their summaries (e.g. *sharp/slight increase→increased sharply/slightly*).

The same patterns were also identified in the control groups' summaries and essays, that is, combining two formulaic sequences (e.g. *the (A) of (B)+(A) and (B)→the production of wheat and corn* and *According to (A)+(A) and (B)→According to some psychologists and managers*), adapting formulaic sequences (e.g. *both of these→both of them* and *the reason for→the reason why*), adding a modifier (e.g. *an **intense** fluctuation*; *it is **sometimes** possible*), extending formulaic sequences (e.g. *(A) and (B)→(A), (B) and (C)* as in *brainstorms, team-based learning and recreational activities* or *(A), (B), (C), and (D)* as in *wheat, corn, sugar and rice*), and altering the part of speech in task 1 (e.g. *gradual decline→declined gradually*).

As can be seen from the examples provided above, the six groups' creative utilization of formulaic sequences was relatively similar in that they all resorted to the same techniques to manipulate formulaic sequences in both writing tasks. The manual coding phase also revealed a clearly higher numbers of creatively used formulaic sequences in the graphical summaries than essay responses. To check this observation, all creatively utilized formulaic sequences were counted and then divided by the total

Table 106

The Creative Use of Formulaic Sequences

	Focused Instruction		Saliency		Control	
	Summaries	Essays	Summaries	Essays	Summaries	Essays
Inter.	17.07%	5.60%	12.17%	7.03%	11.49%	7.01%
Adv.	15.63%	8.73%	13.87%	7.80%	15.76%	5.76%
Overall	16.14%	7.83%	13.43%	7.54%	13.65%	6.35%

number of the ones used by each group. The percentages derived based on this manual quantification revealed the greater tendency on the part of the participants in all six groups to creatively utilize formulaic sequences in their graphical summaries than in their essay responses (see Table 106 above).

The participants' ability to creatively use formulaic sequences in this study is, first and foremost, partly in accord with what Erman and Warren's (2000) call '*combinability*' in writing production; that is, formulaic sequences "are conjoined with each other and with other words to form larger prefabs,… [and] they can be embedded within one another" (p. 51). It is also in accordance with the assumption that L2 learners initially acquire and utilize formulaic sequences as fixed chunks and, at later stages of language development, start to analyze and creatively utilize them in their writing production (e.g. Cowie, 1992; Staples et al., 2013; Willis, 2003). It as well entails that L2 learners do not necessarily rely heavily on fixed formulaic sequences in their writing production; they rather utilize flexible ones as production techniques to produce innovative, error-free stretches of discourse (Cowie, 1992; Erman & Warren, 2000; Staples et al., 2013; Willis, 2003).

The participants' higher tendency to creatively utilize more formulaic sequences in their graphical summaries than in their essay responses is, once again, suggestive of the possible effects of the writing task. As mentioned earlier, graph interpretation is a complex task (Brasell & Rowe, 1993; Glazer, 2011), and formulaic sequences function as "building blocks for managing the highly informational writing found in academic texts" (Staples et al., 2013, p. 224) and as production tools that help L2 writers "make individual choices, with no serious risk of error" (Willis, 2003, p. 45). Taken together, the

complex nature of graph interpretation might have pressured the participants to not only rely on the formulaic sequences they had already mastered but also manipulate them in order to reduce the complexity of the writing task. While this explanation seems plausible, further investigation in future research is required before suggesting a relationship or an association between the complexity of a writing task and the creative use of formulaic sequences.

One point worth mentioning before concluding this section is that it is not clear whether the participants' greater awareness of paragraph constituents and their ability to creatively use formulaic sequence were the result of the target pedagogical interventions and/or corrective feedback they received and/or because they were introduced to this aspects during regular class hours.

Formulaic Sequences and Academic Achievement

The last objective of this research study was to identify any possible impact of and/or association between formulaic sequences and academic achievement. The statistical analysis suggested that (a) none of the pedagogical interventions had substantial effects on the evaluation that the six groups received for both writing tasks; (b) the advanced EAP students, irrespective of their groups, received better evaluation than the upper-intermediate students; and (c) there were strong positive associations between formulaic sequences and academic evaluation.

The raw and mean scores for the evaluation of the six groups revealed numerous inconsistencies within and/or between different groups. The raw scores presented in Table D2 and Table D5 show that some upper-intermediate students and many advanced students in the focused instruction groups received higher grades on at least one of the

summaries and/or essays they completed after the training period. However, Tables 68, 69, 78 and 79 indicate that the mean scores for the upper-intermediate students' summaries and essays were different from those for the advanced students, that is, lower mean scores after the training period for the former but higher mean scores at one or both production stages for the later. Similar to the focused instruction groups, the raw scores for the saliency groups' evaluation indicated that some upper-intermediate students and many advanced students received better evaluation for at least one of the summaries and/or essays they produced after the training period (see Tables D3 and D6). The mean scores for the posttest and/or the delayed posttest of the upper-intermediate students' essays and the advanced students' summaries were higher than those for their pretests (see Tables 69 and 78), unlike the mean scores for the upper-intermediate saliency group's graphical summaries and the advanced saliency group's essays, which were lower after the training period (see Table 70 and Table 81).

The raw scores for the two control groups, on the other hand, revealed that most of the upper-intermediate students and some advanced students received higher grades on the graphical summaries and essay responses they produced after the training period (see Table D4 and D7). The mean scores presented in Tables 68, 69, 78 and 79 indicated that the evaluation that the upper-intermediate and the advanced students received was inconsistent, that is, sometimes higher mean scores and other times lower mean scores after the training period.

The results of the *mixed factorial ANOVAs*, in turn, showed that none of the above-summarized differences approached statistical significance, and the only statistically significant difference was between the evaluation of the upper-intermediate

students and that of the advanced students in both writing tasks; that is, the advanced students in all groups earned significantly higher grades than the upper-intermediate students. As for the association between formulaic sequences and evaluation, the *Pearson correlation coefficient* results uncovered strong positive associations between the frequency and occurrence of the target formulaic sequences and the evaluation of the EAP judge in both writing tasks. In other words, an increase in the use of formulaic sequences, whether repeated or distinct ones, was positively associated with better grades.

Several SLA researchers have theoretically established links between formulaic sequences and academic achievement in that they partly attributed L2 learners' poor academic achievement to their failure to use formulaic sequences in their writing production (e.g. Hill, 2000; Laufer & Waldman, 2011; Lewis, 1997). Other researchers (e.g. Ab Manan & Pandian, 2014; AlHassan, 2016; AlHassan & Wood, 2015; Ohlrogge, 2009) have attempted to investigate the links between formulaic sequences and academic achievement in different types of written texts (e.g. timed-writing EFL exams, proficiency tests, graphical tasks, and academic essays). These studies have either revealed higher achievement after formulaic sequences instruction (i.e., AlHassan, 2016; AlHassan & Wood, 2015) or strong positive correlations between formulaic sequences and academic achievement (e.g. Ab Manan & Pandian, 2014; AlHassan, 2016; AlHassan & Wood, 2015; Ohlrogge, 2009).

The results of the current study are in agreement with the latter. The statistical analysis revealed the statistically significant positive correlation between using more repeated and distinct formulaic sequences in writing production and gaining better marks.

These results indicate that the increase in the average evaluation that the EAP judge assigned for the graphical summaries and essay responses could be partly accounted for by an increase in the average use of formulaic sequences. The results of the *mixed factorial ANOVA*, in turn, showed that none of the pedagogical interventions on its own resulted in receiving better evaluation by any of the six groups. The results also indicated that the advanced EAP students could outperform the upper-intermediate ones in that they obtained significantly higher grades in both writing tasks than the upper-intermediate students. Although these findings were not unexpected, they can be seen as an indicator of the validity of the EAP judge's evaluation. In other words, the EAP judge assigned higher grades to the texts that the advanced EAP students produced despite the nature of the evaluation process (i.e., blind evaluation).

The absence of any statistically significant differences in the evaluation the six groups received before and after the pedagogical intervention was somewhat surprising because of three major factors. First, a similar writing task (i.e., line graphs) and one of the target intervention methods (i.e., focused instruction) were investigated in AlHassan's pilot study (2016) and AlHassan and Wood's (2015) empirical study. Moreover, no major revisions were made to the focused instruction group's worksheets after being piloted in AlHassan (2016). Furthermore, the participants in this study were registered in an EAP program and the study took place over time, which may lead to the expectation that they might receive better evaluation after some time. While it is not possible to provide a clear explanation for such discrepancies, the above-listed factors may indicate that the differences between these studies may be the result of the level of motivation that the students had or the way in which the EAP judge interpreted and used the analytical

rubrics. These possible explanations may seem plausible, but they cannot be taken for granted because other factors that the researcher might not have been aware of may have resulted in this discrepancy. Without follow-up interviews or questionnaires and unless the same results are obtained from multiple EAP judges and/or in other studies, these results remain open to several possible interpretations.

Chapter Seven: Conclusion

The Conclusion Chapter is divided into five sections with five distinct objectives. The first section is devoted to recapitulating the major objectives and the key findings of the present study. The subsequent three sections are respectively intended to discuss the study implications, address its limitations, and suggest some directions for future research, and the fifth section restates the importance of the present research study and concludes the entire dissertation.

Major Objectives and Key Findings

The theoretical accounts of the effectiveness of formulaic sequences in improving L2 academic writing skills and of the pedagogical efficacy of the different approaches to addressing formulaic sequences in promoting their successful acquisition and proper utilization in writing production have been associated with a limited number of research studies that attempted to investigate the effects of a single presentation method, focused on different types of writing tasks, and yielded inconsistent results. To address these gaps in the literature, the present research study explored the effects of two different presentation methods of formulaic sequences on the performance of EAP students at two proficiency levels when completing two different timed writing tasks and compared their performance to other EAP students who were at the same proficiency levels but were instructed pedagogical materials other than formulaic sequences.

The results of the present study were broadly consistent with the literature on the possible pedagogical efficacy of both focused instructional and saliency approaches in enhancing the EAP students' acquisition and utilization of formulaic sequences and in promoting their overall writing proficiency. Such effectiveness was evident in the four

experimental groups' ability to (a) utilize more formulaic sequences (i.e., repeated and/or distinct ones), (b) produce greater numbers of correct formulaic sequences, (c) generate different numbers of words, and (d) make fewer errors in the graphical summaries and/or the essay responses they produced after the training period. The effectiveness of teaching formulaic sequences explicitly and making them salient in input was also deduced from the experimental groups' ability to outperform their counterparts in the control groups in several aspects as well as the performance of both control groups, which remained relatively the same in both writing tasks at the three production stages.

The statistical results also revealed the greater effects that focused instruction of formulaic sequences had on the upper-intermediate and advanced EAP students' performance than making formulaic sequences salient in input, which was inferred from the greater numbers of differences that were detected within the focused instruction groups when compared to their counterparts in the saliency groups and the advanced students' ability to retain formulaic sequences for later use. The statistical results were also reminiscent of the body of literature on the possible links between the proper use of formulaic sequences and L2 writers' proficiency level and revealed some possible effects of the writing task on the EAP students' performance. The study results were as well in accord with the association between formulaic sequences and academic achievement and writing fluency, but they were at odds with the association between formulaic sequences and linguistic accuracy as well as the improved academic achievement as a result of focusing on formulaic sequences. Last but not least, the textual data revealed the six groups' increasing awareness of textual elements and their tendency to creatively utilize formulaic sequences in both writing tasks, two aspects that are perceived as markers of

writing proficiency and might be attributed to the three different types of the delivered pedagogical intervention and/or the participants' natural progression as a result of being registered in an EAP program.

Implications: Pedagogy, Curriculum, and Training

Given the pedagogical nature and the academic context of this research study, the study provides several implications that are of direct relevance to classroom practices, curriculum design, and teacher training.

Classroom practices. The study reinforces the recommendations for addressing formulaic sequences, whether by making them salient or teaching them explicitly, in academic writing instruction (e.g. Boers & Lindstromberg, 2009; Cortes, 2004; Coxhead & Byrd, 2007; Hyland, 2008a; Lewis, 2008; Wood, 2015). The study results corroborate the suggestions proposed by Boers and Lindstromberg (2009), Byrd and Coxhead (2010), Hyland (2008b), Lewis (2008), and Wood (2010b, 2015) among others, that a deliberate focus on and practice of formulaic sequences in EAP or other academic writing courses are likely to promote the chances of acquiring this language phenomenon, raise L2 learners' awareness of the importance of utilizing it in their writing production, improve their overall lexical and syntactic accuracy, and help them abide by academic writing conventions. Such an approach to teaching academic writing can also help writing instructors enrich the impoverished input to which students are exposed with formulaic sequences and reduce the chances of disregarding this language phenomenon in rich input on the part of L2 learners (Laufer, 2005; Wood, 2009, 2010a).

A major implication of this research study is the need to take into account the time constraints and the students' proficiency level prior to selecting a presentation

method. While several differences were detected within the four experimental groups and between the experimental and the control groups, the study results seem to echo the theoretical accounts of the better learning gains of an explicit instructional approach than other possibly effective presentation methods (i.e., saliency). It seems warranted, based on the performance of the four experimental groups, to assume that an explicit instructional approach seems to transcend instantaneous learning gains and result in prolonged learning grains, especially with the time constraints under which writing instructors need to work and the moderating effects that EAP students' proficiency level may have on their ability to acquire salient formulaic sequences.

Advocating the pedagogical efficacy of an explicit instructional approach is reminiscent of scholars' claims that focused instruction of formulaic sequences may promote the chances of committing them to the long-term memory for later use (Boers et al., 2006; Byrd & Coxhead, 2010; Nesselhauf, 2003; Staples et al., 2013; Wood, 2009). Such a recommendation, however, is not meant to devalue other pedagogical practices (i.e., saliency and multiple exposures). Rather, it is to suggest that focused instruction of formulaic sequences may lead to higher chances of their acquisition and retention under the limitations of time and proficiency level. It is also meant to highlight the potentially longer time that EAP students may need to fully acquire a complex language phenomenon as formulaic sequences based on salient input or repeated exposure and to reiterate the limited pedagogical effects the latter may have on L2 learners' productive skills (Byrd & Coxhead, 2010; Cortes, 2004; Durrant & Schmitt, 2010; Laufer & Waldman, 2011).

Another pedagogical implication of this research study concerns the potential benefits of providing EAP students with lists of formulaic sequences they can utilize in complex writing tasks. English academic writing is an inherently complex skill that university students need to master in order to succeed in postsecondary education (Biber, 2006; Murray, 1985; Swales & Feak, 2012). This inherent complexity may substantially increase when university students produce academic texts in their L2 (Ferris, 2012; Hinkel, 2004; Hüttner, 2010) and/or report complex information (e.g. graphical information) in written texts (Glazer, 2011). The literature on formulaic sequences reveals not only their frequent occurrence in academic prose but also their high utility in producing accurate stretches of discourse (e.g. Ellis, 2012; Hasselgren, 1994; Millar, 2011; Simpson-Vlach & Ellis, 2010).

Providing EAP students with thematic lists of formulaic sequences as a prewriting activity—a practice proposed by Conzett (2000) and Howarth (1996)—may help bridge the lexical gap they have in a particular subject matter and consequently mitigate, at least in part, the complexity of the writing task they need to complete. Such a practice may be of high utility in the case of complex writing tasks as graph interpretation as they equip EAP students with stretches of discourse that are typical of the subject matter. It can also provide EAP students with production tools that they can use as '*islands of reliability*' (Dechert, 1984, as cited in Granger, 1998) to complete complex timed written tasks without the risk of making lexical errors or the need to check lexical use.

Curriculum design. The study findings also provide implications for curriculum development. The findings revealed the experimental groups' ability to integrate greater numbers of formulaic sequences, to alleviate the number and/or percentage of syntactic

errors they make in different types of writing tasks, and to produce more concise summaries after the training period. The study results also revealed the relatively stable performance of the control groups as far as the aforementioned aspects are concerned, except for the fewer number of errors the upper-intermediate control group made in their delayed posttest summaries, and the unaffected numbers of erroneous formulaic sequences that the six groups produced in both writing tasks. While these findings remain dubious unless investigated further in a longitudinal study with a larger sample size, they contribute some suggestions for EAP curriculum, mainly curriculum content and assessment.

The major objective of EAP programs is to prepare L2 learners from different disciplines for tasks they may encounter in English medium universities, a complex objective that necessitates developing EAP curricula based on students' needs and contemporary research (Hyland & Hamp-Lyon, 2002; Reid, 2001). The body of literature on phraseology and L2 academic writing has yielded four major findings: (1) different types of formulaic sequences are recurrent in academic prose; (2) L2 learners' use of formulaic sequences differs significantly from that of proficient writers; (3) a major obstacle facing L2 learners in writing production is eliminating linguistic errors, an issue that is often exacerbated by their underdeveloped sense of L2 phraseology; and (4) developing L2 learners' phraseological competence requires deliberate attention that transcends incidental learning (Bishop, 2004a, 2004b; Chen & Baker, 2010; Hasselgren, 1994; Hyland, 2008a; Simpson-Vlach & Ellis, 2010).

The above-listed concepts, together with the study findings, are suggestive of the importance and the potential pedagogical utility of integrating different types of

formulaic sequences into EAP curricula. When designing a curriculum for an EAP program, curriculum developers should integrate different types of formulaic sequences (i.e., collocations, lexical bundles, etc.) that are recurrent in and/or peculiar to different types of written texts (e.g. general academic writing, graph interpretations, business reports, etc.) into curricular content materials and learning objectives. Incorporating formulaic sequences into curricular content provides EAP instructors with pedagogical materials that focus on a recurrent language phenomenon in academic writing and, thus, helps them account for their students' needs. This is particularly important because, as Byrd and Coxhead (2010) explicate, academic writing is marked by the recurrence of different types of formulaic sequences, and language instructors should focus on those different types in the classroom. It may also oblige EAP instructors who assign formulaic a weak position in the language classroom—whether because of time constraints or teaching beliefs—to integrate them into their teaching practices. It can also draw EAP students' attention to the importance of this language phenomenon and encourage them to practice and master its use.

Curriculum developers should also partly base strategies and measures of writing assessment on EAP students' ability to properly utilize different types of formulaic sequences in different types of writing tasks. While assessment is a major indicator of whether a learning objective is attained or not, it is believed to substantially affect students' learning gains and habits (Black & Wiliam, 1998; Earl 2013; Gibbs & Simpson, 2004; Harden, 2001; Meyers & Nulty, 2009). The general consensus regarding the positive effects of assessment on learning gains has emerged from the widespread belief that, in higher education, most students exclusively focus on assessed aspects and

materials (Earl, 2013; Gibbs & Simpson, 2004; Meyers & Nulty, 2009). Accordingly, integrating different types of formulaic sequences into writing assessment may pressure EAP students to consciously attend to and master the productive use of formulaic sequences in order to meet one of the assessment requirements. This, in turn, might promote EAP students' accurate and proper use of formulaic sequences in their writing production, raise their awareness of the restricted use of formulaic sequences based on the target genre and/or register, and help them, later on, meet the expectations of their academic communities.

Assessment—especially formative assessment that is repeatedly administered during the learning process and focuses on learning gains—also has a significant potential to inform and reform classroom instruction; this potential resides in its use as an effective tool to highlight the learning objectives that have been met and the ones that have not been attained yet and, hence, require more focus and/or instruction (Black & Wiliam, 1998; Earl, 2013; Gibbs & Simpson, 2004; Stiggins, 2005). Incorporating formulaic sequences into formative assessment may help EAP instructors identify the issues that their students have in using formulaic sequences (i.e., overuse, underuse, or misuse). In other words, formative assessment may help EAP instructors evaluate their students' productive knowledge of formulaic sequences in writing, which is not only different but also lags behind receptive knowledge (Laufer & Waldman, 2011; O'Donnell et al., 2013; Paquot & Granger, 2012). It may also help EAP instructors provide written corrective feedback based on individual students' weaknesses and to pinpoint recurrent issues that their students, as a group, need instruction on.

Teacher training. The study findings draw some implications for teacher training in EAP contexts. The literature on formulaic sequences and language pedagogy has not only stressed the importance of focusing on formulaic sequences that account for L2 writers' needs, but it has also assigned this task to language instructors because they are the vehicles that deliver language content and, more importantly, the best determiners of their students' needs (Byrd & Coxhead, 2010; Fan, 2009; Hyland, 2008a). The statistical analyses revealed the experimental groups' tendency to increasingly rely on more versatile and accurate formulaic sequences and make fewer errors in their graphical summaries after the training period.

The textual data, in turn, showed the six groups' proclivity to creatively utilize formulaic sequences in their graphical summaries. Irrespective of the reasons for the participants' increased and creative use of formulaic sequences, these findings—together with the literature on the ubiquity of graphs, the under- or misrepresentation of formulaic sequences in EAP textbooks, the scarcity of corpus linguistics research (to the researcher's knowledge) on similar writing tasks, and the diverse needs of EAP students (e.g. Glazer, 2011; Murray, 2016; Reid, 2001; Wood & Appel, 2014)—suggest the possible need for providing EAP instructors with formal training to enable them to extract formulaic sequences from reliable sources (i.e., online corpora) and integrate them into classroom materials.

Providing EAP instructors with formal training on using mega corpora, such as the COCA or the BNC, may equip them with the skills needed to extract formulaic sequences that account for their students' needs. Such a skill may be necessary in EAP contexts because a major objective of EAP programs is to prepare students from diverse

disciplines for different types of writing tasks peculiar to academic context (Hyland & Hamp-Lyon, 2002; Reid, 2001), an objective that may entail focusing on different types of formulaic sequences (Ellis & Simpson-Vlach, 2009; Hyland, 2008a). EAP instructors may need to focus on such writing tasks as graph interpretation. However, they may not have access to lists of formulaic sequences that they can use as a springboard to design materials because of three main reasons: the possible limited access they may have to published research, the scarce and sometimes unreliable formulaic sequences that are present in EAP textbooks, and/or the possible limitations of published lists that mainly focus on formulaic sequences in published research or provide lists for certain genres rather than others (Byrd & Coxhead, 2010; Coxhead, 2008; Flowerdew, 2015; Wood & Appel, 2014).

Developing the necessary skills for using online corpora may therefore enable EAP instructors to extract formulaic sequences that are typical of the writing tasks they address in the language classroom. It may also help EAP instructors overcome the shortcomings of textbooks and classroom materials that are used in EAP contexts, for most EAP textbooks and other materials incorporate formulaic sequences with very limited, if any, information about the source from which they were extracted (i.e., intuition, research, etc.) and the register in which they are used, that is, spoken or written (Byrd & Coxhead, 2010; Coxhead, 2008; Flowerdew, 2015; Murray, 2016; Wood, 2010a). It might as well reduce EAP instructors' tendency to identify formulaic sequences based on intuition, whose reliability as an identification procedure has been repeatedly questioned in literature even in the case of expert researchers (e.g. Coxhead, 2008; O'Keefe et al., 2007; Read & Nation, 2004). In addition, it is likely to encourage

EAP instructors who tend to resist formulaic-based instruction due to their beliefs (e.g. formulaic sequences impede creativity) to integrate formulaic sequences into classroom materials, for they will get a hands-on opportunity to test the frequency and creative use of formulaic sequences in a reliable source as an online mega corpus.

EAP instructors may also need some training on integrating formulaic sequences into classroom materials because the sheer presentation of lists of formulaic sequences to L2 learners is believed to be of very limited pedagogical utility, and EAP instructors may not be familiar with effective pedagogical techniques for integrating formulaic sequences into classroom materials (Byrd & Coxhead, 2010; Jones & Haywood, 2004). During training sessions, EAP instructors may be familiarized with the effectiveness of presenting formulaic sequences with explanatory information and illustrative examples and the pedagogical efficacy of some strategies such as contextual cues (e.g. Bishop, 2004a) and discussion questions that they can adapt in order to have their students attend to salient formulaic sequences in input. Teacher training may also raise EAP instructors' awareness of the possibility of presenting a considerable number of formulaic sequences in decontextualized worksheets that can be communicatively completed within a short period of time.

Limitations

This research study suffers from a number of limitations that restrict the possibility of drawing firm conclusions based on its findings. The first limitation relates to the number of participants in this research study. One of the merits of repeated-measures designs is that, when compared to completely randomized designs, it can provide robust results based on a relatively small sample size (Myers et al., 2010; Pituch & Stevens,

2016; Stevens, 2002; Tabachnick & Fidell, 2007a). Despite this methodological merit, the study results are far from being generalized to its context or any other similar contexts. The total number of EAP students who participated in this study was 67, which may seem adequate to obtain generalizable results. However, the 67 participants constituted six different groups; two groups (i.e., the upper-intermediate saliency group: 5 participants and the advanced control group: 7 participants) were very small. This, in turn, raises questions about the robustness and generalizability of the study findings because of issues that relate to representativeness and inflated *Type II error* (Dörnyei, 2007; Pallant, 2016; Stevens, 2002; Vogt, 2007).

While there is no clear-cut rule for what constitutes a reliable and representative sample except for '*the larger the sample size, the better'* (Dörnyei, 2007; Pallant, 2016; Vogt, 2007), some researchers suggest that a minimum of 15 subjects within each group and 30 subjects in total are respectively required for (quasi-)experimental and correlational designs in order to obtain credible results (Dörnyei, 2007; Rovai, Baker, & Ponton, 2014). Based on these "rough estimates" which Dörnyei (2007) lists as "*rules of thumb*" (p. 99), the sample size used for the correlational analysis (i.e., 67 subjects) seems to be adequate, so are the numbers of participants within the upper-intermediate and advanced focused instruction groups, the advanced saliency group, and the upper-intermediate control group, for they are slightly higher or lower than 15. The numbers of participants within the upper-intermediate saliency and advanced control groups, on the other hand, are clearly lower than the recommended sample size for comparative studies.

The fact that only one EAP judge evaluated the textual data is another limitation. Recruiting more than one EAP judge would also have increased the robustness and

reliability of the study findings. Such a practice, as Dörnyei (2007) and Vogt (2007) explicate, allow researchers who utilize elicitation instruments (i.e., different raters) to statistically explore the internal consistency and reliability of the results (e.g. by computing Cronbach's alpha) or, in Vogt's (2007) words, "the extent to which two or more raters or judges agree" (p. 114). Accordingly, the reliability of the evaluation obtained in this study would have been increased had there been more than one EAP judge. While among the major limitations in this study, the small sample size does not invalidate the study results; rather, it substantially restricts their generalizability to its context or other contexts.

Respondent self-selection, also know as *self-selection effects*, is another limitation in this research study. *Respondent self-selection*, defined as the subjects' ability to make the choice to participate in a research study or not, is perceived as a potential threat to research validity because the subjects who are willing to participate in a research study may be incidentally but inherently different from the ones who don't (Dörnyei, 2007; Vogt, 2007). While this limitation is almost inevitable in most research studies because of ethical considerations, it raises several questions about the degree to which the participants who volunteered in a research study are representative of the target population and, hence, restricts the generalizability of the results (Dörnyei, 2007; Pallant, 2016; Vogt, 2007). Given the fact that the 67 EAP students were self-selected, it is not clear whether they were representative of the students to whom the intervention was delivered and consequently the target population or whether they differed systematically from the ones who were unwilling to participate in such major aspects as aptitude and writing proficiency.

Other limitations in this research study concern the nature and quality of the textual data. The textual data that were analyzed in this study comprised 382 graphical summaries and essay responses, which may seem sufficient to draw firm conclusions. The textual data, however, represent the performance of 67 EAP students, that is, 58.26% of the total number of students to whom the target pedagogical interventions were delivered (i.e., 115). The textual data are not inclusive of all the EAP students, and, consequently, it is hard to determine whether the same results would have been obtained had all 115 students accepted to participate in this study. Moreover, the delayed posttest was collected one month after the posttest, which can be seen as another limitation. While the date for the third round of data collection was determined based on the session length (one term), collecting the delayed posttest at a later point in time may reveal more liable results regarding learning gains.

Another issue relates to absenteeism; that is, some participants were absent on a day during which the pedagogical intervention was delivered or textual data were collected. Although such an issue is quite common in studies that take place over time or involve repeated measurements (Dörnyei, 2007; Meyers et al. 2006;), it is possibly among the major limitations in this study for two reasons. First, it resulted in missing data that had to be manually imputed to avoid sample size reduction. Moreover, a major objective of this research study was to investigate the effects of different intervention methods on the participants' writing skills. Therefore, being absent on an intervention day possibly affected the performance of some participants for they were not exposed to some of the materials and the researcher could not check whether or not they reviewed the missed materials.

The other possible effects on the quality of textual data pertain to the nature of the collected tasks. Several scholars (e.g. Black & Wiliam, 1998; Gibbs & Simpson, 2004; Stiggins, 2005) believe that assessment, especially in post secondary education, bears a relation to students' performance and learning gains. To them, assessment can substantially promote learning because students tend to focus on assessed themes and topics and rarely complete or give careful attention to unassessed work. Gibbs and Simpson's (2004) argue that "students are unlikely to engage seriously with… practice unless it is assessed, or at least required, by the assessment regulations" (p. 15). The fact that students usually concentrate "on topics associated with assessment and nothing else" (Gibbs & Simpson, 2004, p. 14) makes one wonder about the extent to which the collected textual data reflected the participants' learning gains and writing skills. In other words, it is not possible to know whether the participants took the instructed materials and/or the writing tasks seriously, especially that the collected textual data were only practice tasks (i.e., not part of their assessed work) and the practice materials were not delivered by their instructors.

The other limitations that might have affected the participants within some groups relate to some intervening variables that were beyond the researcher's control. While every effort was made to control for intervening variables that might affect the participants' performance, some factors (e.g. students' motivation, aptitude, learning beliefs, etc.) might have affected the participants' performance and learning gains, whether positively or negatively. Other factors that might have potentially impacted the performance of some groups include the type of instruction and the amount of practice and feedback they received from their instructors. The last limitation relates to the

possible sharing of instructional materials between groups. In other words, it was not possible for the researcher to know whether the participants in a particular group (e.g. experimental) shared the materials that they received from the researcher with their colleagues who were registered in a different group and were assigned a different type of pedagogical intervention (e.g. control).

Because of the above-listed limitations, the results obtained in this research study should be interpreted with caution and should be taken as tentative and exploratory until further studies with larger sample sizes draw more firm and robust conclusions that go beyond the study context and its participants.

Directions for Future Research

The study provides several directions for future research. The study results revealed the positive effects of both salient and explicit instructional approaches and the possible supremacy of the latter on the participants' academic writing skills. While these results seem promising, it is not possible to take the study results for granted or generalize them unless similar results are obtained in future research on a larger sample size. Moreover, since the study context (i.e., EAP programs) comprises multiple levels and teachers with different instructional beliefs, future quasi-experimental research on lower proficiency levels (e.g. intermediate, lower-intermediate, etc.) and on the effects of the third proposed presentation method (i.e., repeated exposure) would also be of interest. Not only do future larger research studies yield generalizable results, but they may also reveal the effects of different presentation methods of formulaic sequences (i.e., repeated exposure, salient input, and explicit instruction) on EAP students' writing skills at different proficiency levels.

Future research could also compare the possible effects of solely focusing on formulaic sequences with integrating form-focused instruction with formulaic sequences on EAP students' linguistic accuracy. While such research requires full control over the instruction and the materials to which the participants are exposed, which is not easily attainable, it may solve a major disagreement in the literature regarding the extent to which formulaic sequences alone are likely to promote EAP students' linguistic accuracy (e.g. Granger, 1998; Hinkel, 2004; Willis, 2003). Another possible comparative research study may focus on the effects of a single or multiple presentation methods with different types of writing practice (e.g. independent and/or group practice) of different types of practice materials (e.g. contextualized and decontextualized) to highlight the effects of practice on the acquisition and accurate utilization of formulaic sequences in writing production. Such research is required to better understand the effects of different presentation methods and whether follow-up practice plays a role in the acquisition of this complex language phenomenon.

Since assessment is believed to have significant effects on students' learning habits and performance (Black & Wiliam, 1998; Earl, 2013; Gibbs & Simpson, 2004; Harden, 2001; Meyers & Nulty, 2009), future research can also attempt to explore the effects of presenting formulaic sequences using different approaches as part of assessment criteria on students' assessed writing production. Future research studies may also identify all recurrent formulaic sequences, whether instructed or not, during the data quantification stage. Integrating formulaic sequences as part of assessment criteria and analyzing students' assessed work may minimize EAP students' inclination to disregard formulaic sequences in input because of their complexity (Granger & Meunier, 2008)

and/or irrelevance to assessment (Gibbs & Simpson, 2004) and may yield results based on data that may be a better reflection of the participants' writing proficiency. Counting both instructed and uninstructed formulaic sequences, in turn, reveals the extent to which recognizing formulaic sequences as important elements would pressure EAP students to become *independent learners* (Boers & Lindstromberg, 2012); that is, whether they would look for and master the formulaic sequences they may encounter in other academic materials.

While the study results revealed strong positive correlations between formulaic sequences and the number of words generated in timed writing tasks, the question of whether formulaic sequences per se promote writing fluency remains unanswered. A considerable number of researchers argue that timed writing tasks and speech production share in common several characteristics: they both involve '*lexical retrieval*,' '*syntactic processing*,' and '*content selection*' and are subject to the limitations of the short-term memory (e.g. McCutchen, 2000; Schoonen et al., 2011; Snellings, van Geldren, & De Glopper, 2004; van Gelderen, Oostdam, & Schooten, 2011). Those researchers also hypothesize that, in timed writing tasks, the fluent retrieval of lexical and grammatical constructions from the long-term memory help writers counterbalance the limitations of the working memory and have more time to focus on content.

Since the effects of formulaic sequences on bypassing the limitations of the short-term memory and in producing fluent speech has been empirically demonstrated (e.g. Conklin & Schmitt, 2012; Wood, 2010b) and lexical retrieval from the long-term memory is theoretically considered a major facilitator of writing fluency (Snellings et al., 2004; van Gelderen et al., 2011), future research may explore the effectiveness of

formulaic sequences in promoting writing fluency and proficiency. In so doing, future research will first and foremost be able to address a yet-to-be-answered question (Wood, 2015): *Does the retrieval of formulaic sequences in timed writing tasks promote writing fluency?* It may also explore other uncharted territories: *Does formulaic sequences fast retrieval give L2 writers more time to produce better content and/or higher numbers/ percentages of accurate stretches of discourse*?

Conclusion

This quasi-experimental research study is novel in that it has attempted to empirically investigate the effectiveness of a focused instructional and a saliency approach to presenting formulaic sequences in upgrading upper-intermediate and advanced EAP students' writing skills when approaching two different writing tasks. The study results have highlighted (1) the possible pedagogical effectiveness of both presentation methods in promoting EAP students' successful acquisition and proper utilization of formulaic sequences in their writing production and in improving, at least to some extent, their writing proficiency; (2) the higher prolonged learning gains that focused instruction may lead to; (3) and the effects that the level of proficiency and the type of writing task may have on EAP students' tendency to utilize formulaic sequences in their writing production. The importance of this study resides in revealing the different learning outcomes that may be achieved from different approaches to presenting formulaic sequences at different proficiency levels in EAP contexts. It also lies in highlighting two effective ways in which formulaic sequences can be integrated in academic writing courses to help EAP students treat them as frames that can be utilized in different timed writing tasks in order to help them produce more accurate and proper

stretches of discourse while expressing the intended message.

By highlighting the greater learning gains that both approaches may lead to than solely relying on incidental exposure in EAP contexts, the study results suggest the need for integrating formulaic sequences into teaching academic writing in EAP contexts and highlight the better learning outcomes that may be achieved through an explicit instructional approach in such contexts. While the study results are not definitive or generalizable because of the small sample size, they formed the bases for suggesting some pedagogical implications for EAP programs in Canadian universities and/or colleges and some paths in this field for future research to address.

References

Ab Manan, N. A., & Pandian, A. (2014). The effects of formula instruction on knowledge and use of the target formulas. *Journal of English Language and Literature*, *1*(3), 71-78.

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, *31*(2), 81-92.

Alali, F. A., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal*, *3*(2), 153-180.

AlHassan, L. (2016). Learning all the parts of the puzzle: Focused instruction of formulaic sequences through the lens of activity theory. In H. M. McGarrell & D. Wood (Eds.). *Contact-Special Research Symposium Issue 42*(2), 44-65. Toronto: TESL Ontario.

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes, 17*, 51-62.

Amiri, M., & Fatemi, A. H. (2014). The impact of content-based instruction on students' achievement in ESP courses and their language learning orientation. *Theory and Practice in Language Studies*, *4*(10), 2157-2167.

Appel, R., & Wood, D. (2016). Recurrent word combinations in EAP test-taker writing: Differences between high-and low-proficiency levels. *Language Assessment Quarterly*, *13*(1), 55-71.

Arnaud, P. J. L., & Savignon, S. J. (1997). Rare word, complex lexical units and the

advanced learner. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 157-173). Cambridge: Cambridge University Press.

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

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, *47*(1), 56-63.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Bamberg, B. (1983). What makes a text coherent? *College Composition and Communication*, *34*(4), 417-429.

Barlow, M. (2000). Usage, blends and grammar. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (pp. 315-346). Stanford: CSLI Publications.

Beglar, D., & Hunt, A. (2002). Implementing task-based language teaching. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 96-106). Cambridge: Cambridge University Press.

Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J.,... & Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology*, *20*(4), 517-529.

Benedict, R. H., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 339-352.

Biber, D. (2000). Investigating language use through corpus-based analysis of association

patterns. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (pp. 287-313). Stanford: CSLI Publications.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275-311.

Biber, D. (2010). Corpus-based and corpus-driven analysis of language variation and use. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 159-191). Oxford: Oxford University Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oskserjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181-190). Amsterdam: Rodopi.

Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly, 35*(2), 331-336.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Essex: Pearson Education Limited.

Birch, B. M. (2014). *English grammar pedagogy: A global perspective*. New York, NY: Routledge.

Bishop, H. (2004a). Noticing formulaic sequences—A problem of measuring the subjective. *LSO Working Papers in Linguistics, 4*, 15-19.

Bishop, H. (2004b). The effects of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 227-244). Amsterdam: John Benjamins.

Björkman, B. (2013). *English as an academic lingua franca: An investigation of form and communicative effectiveness*. Boston: De Gruyter Mouton.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Bland, J. M. (2015). *An introduction to medical statistics* (4th ed.). Oxford: Oxford University Press.

Bland, J. M., & Altman, D. G. (1995a). Calculating correlation coefficients with repeated observations: Part 1-Correlation within subjects. *BMJ, 310*, 446.

Bland, J. M., & Altman, D. G. (1995b). Calculating correlation coefficients with repeated observations: Part 2-Correlation between subjects. *BMJ, 310*, 633.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245-261.

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke: Palgrave Macmillan.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*,

32, 83-110.

Brasell, H. M., & Rowe, M. B. (1993). Graphing skills among high school physics students. *School Science and Mathematics*, *93*(2), 63-70.

Brick. J. (2012). Teaching English for academic purposes. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide for pedagogy and practice in second language teaching* (pp. 170-178). Cambridge: Cambridge University Press.

Brown, C. L. (2004). Content based ESL curriculum and academic language proficiency. *The Internet TESL Journal*, *10*(2), 1-7.

Butler, Y. G. (2002). Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system. *Studies in Second Language Acquisition*, *24*(3), 451-480.

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, *5*, 31-64.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English*. Cambridge: Cambridge University Press.

Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and context in language teaching: A guide for language teachers*. Cambridge: Cambridge University Press.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, *14*(2), 30-49.

Chomsky, N. (1982). A note on the creative aspects of language use. *The philosophical*

Review, *91*(3), 423-434.

Chomsky, N. (2006). *Language and mind* (3rd ed.). Cambridge: Cambridge University Press. Retrieved from http://lib.myilibrary.com.proxy.library.carleton.ca/Open.aspx?id=43197

Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools*, *44*(5), 451-459.

Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 225-237). Cambridge: Cambridge University Press.

Cogo, A., & Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation*. London: Continuum.

Collins, L. (2012). Form-focused instruction. *The Encyclopedia of Applied Linguistics*. Retrieved from http://onlinelibrary.wiley.com.proxy.library.carleton.ca/ doi/10.1002/9781405198431.wbeal0432/full

Čolović-Marković, J. (2012). *The effects of explicit instruction of formulaic sequences on second-language writers*. (Unpublished doctoral dissertation). The University of Utah, Salt Lake City.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32,* 45-61.

Connors, R. J., & Lunsford, A. A. (1988). Frequency of formal errors in current college writing, or Ma and Pa Kettle do research. *College Composition and Communication*, *39*(4), 395-409.

Conrad, S. M. (1999). The importance of corpus-based research for language teachers. *System*, *27*(1), 1-18.

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, *34*(3), 548-560.

Conzett, J. (2000). Integrating collocation into a reading and writing course. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 70-87). Boston, MI: Heinle.

Cook, V. 1997. *Inside language.* London: Arnold.

Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131-145). Amsterdam: John Benjamins.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, *17*(4), 391-406.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, *3*(1), 43-57.

Cowie, A. P. (1992). Multiword lexical units and communicative language teaching. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 1-12).

London: Macmillan Academic and Professional.

Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 149-161). Amsterdam: John Benjamins.

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147.

Creswell, J. W., & Plan Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Curry, M. J., & Lillis, T. M. (2003). Issues in academic writing in higher education. In C. Coffin, M. J. Curry, S. Goodman, A. Hewings, T. M., Lillis & Swann, J. (Eds.), *Teaching academic writing: A toolkit for higher education* (pp. 1-18). London: Routledge.

D'Angelo, F. J. (1986). The topic sentence revisited. *College Composition and Communication, 37*(4), 431-441.

Davies, M. (2008). *The corpus of contemporary American English (COCA): 520 million words, 1990-present*. Retrieved from http://corpus.byu.edu/coca/

Davies, M. (2009). The 385+ million word corpus of contemporary American English (1990-2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*(2), 159-190.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing, 25*(4), 447-464.

Dias, P., & Paré, P. (Eds.). (2000). *Transitions: Writing in academic and workplace*

settings. Cresskill, NJ: Hampton Press.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Dörnyei, Z. (2013). Communicative language teaching in the twenty-first century: The 'principled communicative approach.' In J. Arnold & T. Murphey (Eds.), *Meaningful action: Earl Stevick's influence on language teaching* (pp. 161-171). Cambridge: Cambridge University Press.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, *28*(3), 157-169.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, *38*(2), 165-193.

Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, *30*(1), 58-72.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, *47*(2), 157-177.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163-188.

Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning* (2nd ed.). Thousand Oaks, CA: Corwin.

Elbow, P. (2012). *Vernacular eloquence: What speech can bring to writing*. Oxford: Oxford University Press.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*(1), 91-126.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17-44.

Ellis, N. C., & Simpson-Vlach, R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory*, *5*(1), 61-78.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 375-396.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1998). Teaching and research: Options in grammar teaching. *TESOL Quarterly 32*(1) 39-60.

Ellis, R. (2002). Grammar teaching-Practice or consciousness raising? In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 167-174). Cambridge: Cambridge University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227-275.

Ellis, R. (2005a). Principles of instructed language learning. *System*, *33*(2), 209-224.

Ellis, R. (2005b). *Instructed second language acquisition: A literature review*.

Wellington: Ministry of Education.

Ellis, R. (2005c). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141-172.

Ellis, R. (2015). *Understanding second language acquisitio*n (2nd ed.). Oxford: Oxford University Press.

Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus-on-form. *System*, *30*(4), 419-432.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.

Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, *12*(1), 25-53.

Erman, B., Lewis, M., & Fant, L. (2013). Multiword structures in different materials, and with different goals and methodologies. In J. Romero Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2013: New domains and methodologies* (pp. 77-103). London: Springer.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62.

Fan, M. (2009). An exploratory study of collocational use by ESL students–A task based approach. *System*, *37*(1), 110-123.

Feild, L., Pruchno, R. A., Bewley, J., Lemay, E. P., & Levinsky, N. G. (2006). Using probability vs. nonprobability sampling to identify hard-to-access participants for health related research: Costs and contrasts. *Journal of Aging and Health*, *18*(4), 565-583.

Ferris, D. (2002). Teaching students to self-edit. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 328-334). Cambridge: Cambridge University Press.

Ferris, D. (2012). Writing instruction. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide for pedagogy and practice in second language teaching* (pp. 226-235). Cambridge: Cambridge University Press.

Ferris, D. R., & Hedgcock, J. S. (2014). *Teaching L2 composition: Purpose, process, and practice* (3rd ed.). New York, NY: Routledge.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*(3), 161-184.

Fidell, L. S., & Tabachnick. B. G. (2003). Preparatory data analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 115–141). New York, NY: John Wiley.

Flowerdew, L. (2010). Using corpora for writing instruction. In A. O'keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 444- 457). London: Routledge.

Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, *48*(1), 99-116.

Fox-Wasylyshyn, S. M., & El-Masri, M. M. (2005). Focus on research methods: Handling missing data in self-report measures. *Research in Nursing & Health*, *28*(6), 488-495.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for*

Research in Mathematics Education, 32(2), 124-158.

Fromkin, V., Rodman, R., & Hyams, N. (2003). *An introduction to language* (9th ed.). Boston, MA: Wadsworth.

Gass, S. (1999). Incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*(2), 319-333.

Gass, S., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 175-199). Mahwah, NJ: Lawrence Erlbaum.

George, A. (2009). Quantitative and qualitative approaches to content analysis. In K. Krippendorff & M. A. Bock (Eds.), *The content analysis reader* (pp. 144-155). Thousand Oaks, CA: SAGE.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486-489.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3-31.

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, *6*(4), 319-335.

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, *47*(2), 183-210.

Gledhill, C. J. (2000). *Collocations in science writing*. Tübingen: Gunter Narr Verlag.

González-Bueno, M., & Pérez, L. C. (2000). Electronic mail in foreign language writing: A study of grammatical and lexical accuracy, and quantity of language. *Foreign*

Language Annals, *33*(2), 189-198.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford: Clarendon Press.

Granger, S., & Meunier, F. (2008). Phraseology in language learning and teaching: Where to from here? In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247-252). Amsterdam: John Benjamins.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27-49). Amsterdam: John Benjamins.

Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 3-25). Amsterdam: John Benjamins.

Hairston, M. (1982). The winds of change: Thomas Kuhn and the revolution in the teaching of writing. *College Composition and Communication*, *33*(1), 76-88.

Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*(4), 597-618.

Handl, S. (2008). Essential collocations for learners of English: The role of collocational direction and weight. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 43-66). Amsterdam: John Benjamins.

Harden R. M. (2001). AMEE Guide No.21: Curriculum mapping: A tool for transparent and authentic teaching and learning. *Medical Teacher, 23*(2), 123-137.

Harper, C., & de Jong, E. (2004). Misconceptions about teaching English-language

learners. *Journal of Adolescent & Adult Literacy*, *48*(2), 152-162.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, *4*(2), 237-258.

Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas, TX: Southern Methodist University Press.

Henry, A., & Roseberry, R. L. (1998). An evaluation of a genre-based approach to the teaching of EAP/ESP writing. *TESOL Quarterly, 32*(1), 147-156.

Hewings, M., & Hewings, A. (2002). "It is interesting to note that…": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, *21*(4), 367-383.

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*(4), 473-493.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 47-69). Boston, MI: Heinle.

Hill, J., Lewis, M., & Lewis, M. (2000). Classroom strategies, activities and exercises. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 88-117). Boston, MI: Heinle.

Hilton, A., & Armstrong, R. (2006). Stat note 6: Post-hoc ANOVA tests. *Microbiologist, 7*, 34-36.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary*

and grammar. Mahwah, NJ: Lawrence Erlbaum Associates.

Hinton, P. R., McMurray, I., & Brownlow, C. (2014). *SPSS Explained* (2nd ed.). Hove: Routledge.

Howarth, P. A. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making.* Tübingen: Max Niemeyer Verlag.

Howarth, P. (1998a). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford: Oxford University Press.

Howarth, P. (1998b). Phraseology and second language proficiency. *Applied Linguistics*, *19*(1), 24-44.

Hunt, A., & Beglar, D. (2002). Current research and practice in teaching vocabulary. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 258-266). Cambridge: Cambridge University Press.

Hüttner, J. (2010). The potential of purpose-built corpora in the analysis of student academic writing in English. *Journal of Writing Research*, *2*(2), 197-218.

Hyland, K. (2003a). *Second language writing*. Cambridge: Cambridge University Press.

Hyland, K. (2003b). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, *12*(1), 17-29.

Hyland, K. (2004). *Genre and second language writing*. Ann Arbor, MI: The University of Michigan Press.

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, *16*(3), 148-164.

Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, *18*(1), 41-62.

Hyland, K. (2016). *Teaching and researching writing: Applied linguistics in action* (3rd ed.). New York, NY: Routledge.

Hyland, K., & Hamp-Lyons, L. (2002). EAP: Issues and directions. *Journal of English for Academic Purposes*, *1*(1), 1-12.

IBM Support. (n.d.). IBM. (accessed in January, 2018). Retrieved from http://www-01.ibm.com/support/docview.wss?uid=swg21475404

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, *4*(1), 51-69.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, *24*(4), 541-577.

Jacobs, G., & Small, J. (2003). Combining dictogloss and cooperative learning to promote language learning. *The Reading Matrix*, *3*(1), 1-15.

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching*, *44*(3), 281-315.

Jeong, H. (2013). Verbal data analysis for understanding interactions. In C. E. Hmelo-Silver, C. A. Chinn, C. K. Chan & A. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 168-183). New York, NY: Routledge.

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: 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). London: Lawrence Erlbaum.

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 269-300). Amsterdam: John Benjamins.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*(2), 130-144.

Kaplan, R. B. (1967). Contrastive rhetoric and the teaching of composition. *TESOL Quarterly*, *1*(4), 10-16.

Kasper, L. F. (2000). Content-based college ESL instruction: Theoretical foundations and pedagogical applications. In L. F. Kasper, M. Babbitt, R. W. Mlynarczyk, D. M. Briton, J. W. Rosenthal, P. Master,… L. S. Wood (Eds.), *Content-based college ESL instruction* (pp. 3-23). Mahwah, N.J: Lawrence Erlbaum.

Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology, 49*(2), 183-192.

Kazemi, M., Katiraei, S., & Rasekh, A. E. (2014). The impact of teaching lexical bundles on improving Iranian EFL students' writing skill. *Procedia-Social and Behavioral Sciences*, *98*, 864-869.

Kennedy, G. (2008). Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 21-

41). Amsterdam: John Benjamins.

Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, *38*(1), 52-54.

Knapp, P., & Watkins, M. (1994). *Context, text, grammar: Teaching the genres and grammar of school writing in infants and primary classrooms*. Sydney: Text Productions.

Kress, G. (1994). *Learning to write.* London: Routledge.

Laufer, B. (2005). Instructed second language vocabulary learning: The fault in the 'default hypothesis'. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 311–329). Berlin: Mouton de Gruyter.

Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review*, *63*(1), 149-166.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647-672.

Lee, C. (2013). Using lexical bundle analysis as discovery tool for corpus-based translation research. *Perspectives*, *21*(3), 378-395.

Leedy, P. D. (1997). *Practical research: Planning and design* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Leki, I. (1991). Twenty-five years of contrastive rhetoric: Text analysis and writing

pedagogies. *TESOL Quarterly*, *25*(1), 123-143.

Lembke, E., Deno, S. L., & Hall, K. (2003). Identifying an indicator of growth in early writing proficiency for elementary school students. *Assessment for Effective Intervention*, *28*(3-4), 23-35.

Lewis, M. (1996). Implications of a lexical view of language. In D. Willis & J. Willis (Eds.), *Challenge and change in language teaching* (pp. 10-16). Oxford: Heinemann.

Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255-270). Cambridge: Cambridge University Press.

Lewis, M. (2000a). Materials and resources for teaching collocation. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 186-204). Boston, MI: Heinle.

Lewis, M. (2000b). Language in the lexical approach. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 126-154). Boston, MI: Heinle.

Lewis, M. (2000c). Learning in the lexical approach. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 155-185). Boston, MI: Heinle.

Lewis, M. (2002). *The lexical approach: The state of ELT and a way forward*. Boston, MA: Heinle.

Lewis, M. (2008). *Implementing the lexical approach: Putting theory into practice*. London: Heinle.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85-102.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.

Liu, D. (2008). Linking adverbials: An across-register corpus study and its implications. *International Journal of Corpus Linguistics*, *13*(4), 491-518.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, *31*(1), 25-35.

Liu, W., & Huo, Y. (2011). On the role of formulaic sequences in second language acquisition. *US-China Foreign Language, 9*(11), 31-36.

Llach, A. (2011). *Lexical errors and accuracy in foreign language writing*. Bristol: Multilingual Matters.

Luke, A. (1996). Genres of power? Literacy education and the production of capital. In R. Hasan & A. G. Williams (Eds.), *Literacy in society* (pp. 308–338). London: Longman.

Marco, M. J. L. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes*, *19*(1), 63-86.

Mark, M. M., & Reichardt, C. S. (2009). Quasi-experimentation. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research* (2nd ed.) (pp. 181-213). Thousand Oaks, CA: SAGE.

Martin, W. (2008). A unified approach to semantic frames and collocational patterns. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*

(pp. 51-65). Amsterdam: John Benjamins.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299-320.

Matsuda, P. K. (2003a). Process and post-process: A discursive history. *Journal of Second Language Writing*, *12*(1), 65-83.

Matsuda, P. K. (2003b). Second language writing in the twentieth century: A situated historical perspective. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 15–34). Cambridge: Cambridge University Press.

Mayher, J. S., Lester, N. B., & Pradl, G. M. (1983). *Learning to write/writing to learn.* Portsmouth, NH: Boynton/Cook.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*(1), 13-23.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, *32*, 111-129.

Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics*, *32*(3), 229-263.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research*. Thousand Oaks, CA: Sage.

Meyers, N. M., & Nulty, D. D. (2009). How to use (five) curriculum design principles to align authentic learning environments, assessment, students' approaches to thinking and learning outcomes. *Assessment & Evaluation in Higher Education*, *34*(5), 565-577.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129-148.

Mitchell, C. (2004). *Writing and power: A critical introduction to composition studies*. London: Paradigm.

Model Course 3.17: Maritime English (2nd ed.). (2009). London: International Maritime Organization.

Moon, R. (1992). Textual aspects of fixed expressions in learners' dictionaries. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 13-27). London: Macmillan Academic and Professional.

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Oxford University Press.

Mora-Flores, E. (2009). *Writing instruction for English learners: A focus on genre*. Thousand Oaks, CA: Corwin Press.

Murray, D. M. (1985). *A writer teaches writing* (2nd Ed.). Boston, MI: Houghton Mifflin.

Murray, L. (2016). An analysis of the usefulness of formulaic sequences and related acquisition exercises in an EFL academic writing textbook. *Studies of Language and Culture*, *20*, 29-45.

Myers, J. L., Well, A. D., & Lorch, R. F. (2010). *Research design and statistical analysis* (3rd ed.). New York, NY: Routledge.

Nabei, T. (1996). Dictogloss: Is it an effective language learning task? *Working Papers in Educational Linguistics*, *12*(1), 59-74.

Nation, I. S. P. (1991). Dictation, dicto-comp, and related techniques. *English Teaching Forum*, *29*(4), 12-14.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2002). Best practice in vocabulary teaching and learning. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 267-272). Cambridge: Cambridge University Press.

Nation, I. S. P., & Newton, J. (1997). Teaching vocabulary. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 238-254). Cambridge: Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Research and analyzing vocabulary*. Boston, MA: Heinle.

Nattinger, J. R. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly, 14*(3), 337-344.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223-242.

Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). Thousand Oaks, CA: SAGE.

Oakey, D. (2002). Formulaic language in English academic writing: A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore Linguistic Variation* (pp. 111-129). Amsterdam: John

Benjamins.

O'Donnell, M. B., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics*, *18*(1), 83-108.

Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E. A. Moravcsik & H. Ouali (Eds.), *Formulaic language: Acquisition, loss, psychology, reality, and functional explanations* (pp. 375-386). Amsterdam: John Benjamins.

O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Omazic, M. (2008). Processing of idioms and idiom modifications: A view from cognitive linguistics. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 67-79). Amsterdam: John Benjamins.

Oshima, A., & Hogue, A. (2006). *Writing academic English*. While Plains, NY: Pearson.

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (6th ed.). Sydney: Allen & Unwin.

Paltridge, B. (2004). Academic writing. *Language Teaching*, *37*(2), 87-105.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 101-119). Amsterdam: John Benjamins.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32*, 2-19.

Paré, A. (2009). What we know about writing, and why it matters. *Compendium 2, 2*(1), 1-12.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). New York, NY: Longman.

Pecorari, D. (2009). Formulaic language in biology: A topic-specific investigation. In M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 91-104). London: Continuum International Publishing Group.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, *14*, 84-94.

Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly, 14*(1), 61-69.

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for Academic Purposes*, *20*, 28-39.

Philip, G. (2008). Reassessing the canon: 'Fixed' phrases in general reference corpora. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 95-108). Amsterdam: John Benjamins.

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS* (6th ed.). New York, NY: Routledge. Available from https://ebookcentral-proquest-com.proxy.library.carleton.ca/lib/oculcarleton-ebooks/detail.action?docID=4218570

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, *47*(1), 101-143.

Pratt, D. (2011). *Modelling written communication: A new systems approach to modeling in the social sciences.* London: Springer.

Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language Teaching Research*, *12*(1), 61-82.

Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. *TESOL Quarterly*, *25*(3), 407-430.

Raimes, A. (2002). The steps in planning a writing course and training teachers of writing. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 306-314). Cambridge: Cambridge University Press.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21-33.

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 23-35). Amsterdam: John Benjamins.

Reichardt, C. S. (2009). Quasi-experimental design. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 46-71). Los Angeles, CA: SAGE.

Reid, J. (2001). Advanced EAP writing and curriculum design: What do we need to know? In T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 143–

160). Mahwah, NJ: Erlbaum.

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: Cambridge University Press.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, *15*(1), 18-39.

Richards, J. C. (2002). Addressing the grammar gap in task work. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 153-166). Cambridge: Cambridge University Press.

Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press.

Römer, U. (2009). English in academia: Does nativeness matter. *International Journal of English Studies*, *20*(2), 89-100.

Rovai, A. P., Baker, J. D., & Ponton, M. K. (2014). *Social science research design and statistics: A practitioner's guide to research methods and IBM SPSS*. Chesapeake, VA: Watertree Press LLC.

Rutherford, W. E., & Smith, M. S. (1985). Consciousness-raising and universal grammar. *Applied Linguistics*, *6*(3), 274-282.

Salanti, G., Ades, A. E., & Ioannidis, J. P. (2011). Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: An overview and tutorial. *Journal of Clinical Epidemiology*, *64*(2), 163-171.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data

management in counseling psychology. *Journal of Counseling psychology*, *57*(1), 1-10.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329-363.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1-22). Amsterdam: John Benjamins.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psychologically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 127-148). Amsterdam: John Benjamins.

Schoonen, R., van Gelderen, A., Stoel, R. D., Hulstijn, J., & de Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, *61*(1), 31-79.

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: Cambridge University Press.

Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programmes. *Language Teaching Research*, *19*(1), 89-106.

Shah, M. I. A. (2003). Language learning content-based English as a second language (ESL) Malaysian classrooms. *Journal of Language and Learning*, *1*(2), 73-97.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591-611.

Shaw, P. (2009). Linking adverbials in student and professional writing in literary studies: What makes writing mature. In M. Charles, S. Hunston & D. Pecorari (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 215-235). London: Continuum.

Shih, M. (1986). Content-based approaches to teaching academic writing. *TESOL Quarterly*, *20*(4), 617-648.

Shin, Y. K., & Kim, Y. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, *69*, 79-91.

Siepmann, D. (2006). Academic writing and culture: An overview of differences between English, French and German. *Meta*, *51*(1), 131-150.

Siepmann, D. (2008). Phraseology in learners' dictionaries: What, where, how? In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 185-202). Amsterdam: John Benjamins.

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: Cambridge University Press.

Simard, D. (2009). Differential effects of textual enhancement formats on intake. *System*, *37*(1), 124-135.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly, 37*(3) 419-441.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in

phraseology research. *Applied Linguistics, 31*(4), 487-512.

Sinclair, J. (1991) *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429-458.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*(1), 38-62.

Smit, U. (2010). *English as a lingua franca in higher education: A longitudinal study of classroom discourse*. Berlin: Walter de Gruyter.

Smith, M. S. (1981). Consciousness-raising and the second language learner. *Applied Linguistics, 11*(2), 159-168.

Smith, M. S. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, *7*(2), 118-132.

Smith, M. S. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*(2), 165-179.

Snellings, P., Van Gelderen, A., & De Glopper, K. (2004a). The effect of enhanced lexical retrieval on L2 writing: A classroom experiment. *Applied Psycholinguistics, 25*(2), 175-200.

Spada, N. (2005). Conditions and challenges in developing school-based SLA research programs. *The Modern Language Journal*, *89*(3), 328-338.

Spada, N. (2007). Communicative language teaching: Current status and future prospects. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 271-288). New York, NY: Springer.

Spence, L. K. (2010). Discerning writing assessment: Insights into an analytical rubric. *Language Arts*, *87*(5), 337.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, *87*(4), 324-328.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language*, *2*(1), 23-55.

Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, *7*(2), 215-244.

Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type: A quantitative method and some findings. *Functions of Language*, *10*(1), 61-104.

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills*. Ann Anbor, MI: University of Michigan Press.

Swan, M. (2002). Seven bad reasons for teaching grammar- And two good ones. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 148-152). Cambridge: Cambridge University Press.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Tabachnick, B. G., & Fidell, L. S. (2007a). *Experimental design using ANOVA*. Belmont, CA: Duxbury.

Tabachnick, B. G., & Fidell, L. S. (2007b). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Essex: Pearson Education Limited.

Tansey, O. (2007). Process tracing and elite interviewing: A case for non-probability sampling. *PS: Political Science and Politics*, *40*(4), 765-772.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Van den Branden, K. (2016). Task-based language teaching. In G. Hall (Ed.), *The Routledge handbook of English language teaching* (pp. 238-251). Abingdon: Routledge.

van Gelderen, A., Oostdam, R., & van Schooten, E. (2011). Does foreign language writing benefit from increased lexical fluency? Evidence from a classroom experiment. *Language Learning*, *61*(1), 281-321.

Van Ness, P. H., Murphy, T. E., Araujo, K. L., Pisani, M. A., & Allore, H. G. (2007). The use of missingness screens in clinical epidemiologic research has implications for regression modeling. *Journal of Clinical Epidemiology*, *60*(12), 1239-1245.

VanPatten, B., & Williams, J. (2007). Early theories in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 17-35). Mahwah, NJ: Lawrence Erlbaum.

Van Roey, J. (1990). *French-English contrastive lexicology: An Introduction*. Leuven: Peeters.

Verstraten, L. (1992). Fixed phrases in monolingual learners' dictionaries. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 28-40). London: Macmillan Academic and Professional.

Vogt, W. P. (2007). *Quantitative research methods for professionals*. Boston, MA: Pearson.

Wajnryb, R. (1990). *Grammar dictation*. Oxford: Oxford University Press.

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, *6*(1), 18-35.

Warrington, S. D. (2008). Concerns with content-based instruction (CBI) In Asian EFL contexts. *The Buckingham Journal of Language and Linguistics*, *1*, 129-141.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, *16*(2), 180-205.

Welch, M., & Link, D. P. (1992). Informal assessment of paragraph composition: A seven-step informal assessment procedure for evaluating the form and function of paragraph composition. *Intervention in School and Clinic, 27*(3), 145–149.

Williams, J. (2005). *Teaching writing in second and foreign language classrooms.* Boston, MA: The McGraw-Hill Companies.

Williams, J. (2013). *LEAP advanced: Reading and writing*. Montreal: Pearson.

Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching.* Cambridge: Cambridge University Press.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review*, *57*(4), 573-589.

Wood, D. (2002). Formulaic language in acquisition and production: Implications for

teaching. *TESL Canada Journal, 20*(1), 1-15.

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review, 63*(1), 13-33.

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, *12*(1), 39-57.

Wood, D. (2010a). Lexical clusters in an EAP textbook corpus. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 88-106). London: Continuum International Publishing Group.

Wood, D. (2010b). *Formulaic language and second language speech fluency: Background, evidence and classroom applications.* London: Continuum International Publishing Group.

Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. London: Bloomsbury.

Wood, D. C., & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, *15*, 1-13.

Woolard, G. (2000). Collocation—Encouraging learner independence. In M. Lewis (Ed.), *Teaching collocations: Further developments in the lexical approach* (pp. 28-46). Boston, MI: Heinle.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 21*(4), 463-489.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Xu, X. (2010). The effects of Glosses on incidental vocabulary acquisition in reading. *Journal of Language Teaching and Research, (1)*2, 117-120.

Young, R. (1996). Form-function relations in articles in English interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 135-175). Amsterdam: John Benjamin.

Zhu, W. (2006). Understanding context for writing in university content classrooms. In P. K. Matsuda, C. Ortmeire-Hooper & X. You (Eds.), *The politics of second language writing: In search of the promised land* (pp. 129-146). West Lafayette, IN: Parlor Press.

Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 5-19). Cambridge: Cambridge University Press.

Appendix A: Selected Formulaic Sequences

Based on their structure and usage, the selected formulaic sequences for this study fall into three major categories: sentence builders, discourse organizers, and collocations

Sentence Builders

The first group of formulaic sequences for this study comprises 31 formulaic sequences whose function in academic writing pertains to what has been referred to in the literature as sentence builders. Sentence builders, sometimes referred to as sentence frames, are continuous or discontinuous strings of words that can be as short as a phrase or as long as a sentence; they provide a highly variable skeleton for constructing stretches of discourse (Birch, 2014; Granger, 1998; Nattinger, 1980; Nattinger & DeCarrico, 1992; Oakey, 2002). The sentence builders chosen for this study are:

- *The/A (A) of (B)* is one of the most common phrases in academic prose, which is constructed of a noun phrase and an of-phrase fragment, (Biber, 2009; Byrd & Coxhead, 2010; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). Not only is it used to frame the attributes of the following noun phrase, but it can also be used to formulate an entire clause or connect a clause with a preceding verb (Simpson-Vlach & Ellis, 2010). It may function as both a tangible and an intangible framing attribute; it conveys a range of different meanings that include, but are not limited to, highlighting quality, specifying quantity, marking existence, and specifying the physical or measurable attributes of the following noun or noun phrase (Biber et al., 2004; Hyland, 2008b; Liu, 2012; Simpson-Vlach & Ellis, 2010).
- *In the (A) of* is a tangible framing attribute and/or a framing signal which is composed of a prepositional phrase and an of-phrase fragment (Biber et al., 2004; Cortes, 2006;

Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010). It is mainly used to organize text elements and to "situate arguments by specifying limiting conditions" (Hyland, 2008a, p. 14).

- *Over a period of* is another tangible framing attribute which is, as indicated in Simpson-Vlach and Ellis's (2010) AFL, highly frequent in academic writing. It is typically used to specify the measurable attributes of the following noun (Simpson-Vlach & Ellis, 2010).
- *Between (A) and (B)* is extracted from the COCA to provide the participants with a formulaic sequence that can be used in order to present time or price range, two elements that are typical of reporting trends in graphic charts. This formulaic sequence occurs 449.09 per one million words and is introduced as an alternative for *over a period of*.
- *From (A) to (B)* is another formulaic sequence that is extracted from the COCA as an alternative for *over a period of* and *between (A) and (B)* due to its high frequency in academic discourse (i.e., 302.10 per one million words). It has a two-fold function in that it can be used to introduce time and/or price range.
- *At the beginning of/ At the end of* are multifunctional referential time expressions that are recurrent in academic writing (Biber, 2009; Chen & Baker, 2010; Cortes, 2004, 2008; Coxhead & Byrd, 2007; Liu, 2012; Stubbs, 2002). Both are composed of a preposition and a noun phrase fragment and are utilized to present temporal elements in a text, refer to the beginning and the end of an event, or set forth and evaluate content (Ädel & Erman, 2012; Byrd & Coxhead, 2010; Cortes, 2008).

- *Play$_{(s)}$ an important role in* is an evaluative formulaic sequence that is highly frequent in academic prose; its paramount function is to either provide an evaluation of the preceding noun or establish causative links between discourse elements (Hyland, 2008a, 2008b; Liu, 2012; Simpson-Vlach & Ellis, 2010).
- *(Due to) the fact that* is an intangible framing attribute—also referred to as an impersonal epistemic bundle—that can be alternatively used with such prepositions as *on*, *despite*, *from*, etc. and is followed by a postmodifier fragment (Biber et al., 2004; Coxhead & Byrd, 2007; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010). It is used to highlight resultative elements or to propose objective attitudes in academic writing (Chen & Baker, 2010; Hyland, 2008a; Simpson-Vlach & Ellis, 2010).
- *As a result of* is a cause and effect prepositional formulaic sequence, also referred to as a resultative signal; despite its fixed structure, *as a result of* is not perceptually salient due to its recurrence in dense academic prose (Byrd & Coxhead, 2010; Cortes, 2006; Coxhead & Byrd, 2007; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). It is composed of a prepositional phrase and an of-fragment and is one of the most frequent extended collocations in the academic register that is used establish causative relations among text elements and is considered a marker of writing proficiency (Chen & Baker, 2010; Coxhead & Byrd, 2007; Hyland, 2008a; Simpson-Vlach & Ellis, 2010).
- *The reason for* is another cause and effect formulaic sequence that is highly frequent in written academic English and is primarily used to elaborate on a topic and to reveal the causative relations between elements in a text (Simpson-Vlach & Ellis, 2010).

- *It can be seen/observed* and *As can be seen* are two formulaic sequences, also referred to as participant-oriented bundles, which are utilized as clause or sentence stems in order to provide an objective evaluation of or direct the reader's attention to previously introduced elements in a written text (Chen & Baker; 2010; Hyland, 2008a; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). They are highly frequent in academic prose (i.e., one of the top 10 formulaic sequences in the AFL) because identification is a primary function in academic writing (Simpson-Vlach & Ellis, 2010).
- *(A) and (B)* and *(A) as well as (B)* are two sentence builders that are also referred to as deictics and locative formulas and are highly frequent in academic prose; they are used as sentence builders or as constituents of a prepositional phrase in order to connect similar elements of or to refer to spatial or temporal elements in a text (Byrd & Coxhead, 2010; Chen & Baker, 2010; Liu, 2012; Nattinger & DeCarrico, 1992; Simpson-Vlach & Ellis, 2010).
- *Both (A) and (B)* is a formulaic sequence extracted from the COCA as an alternative for *(A) and (B)* and *(A) as well as (B)* due to its high frequency in the academic register (319.62 per one million words) and its function as a sequence that helps cluster and establish links between similar textual elements.
- *Both of these* is a quantity specification formulaic sequence that is recurrent in academic writing; its main function is anaphoric in that it refers to a noun phrase that has been previously introduced in a text (Simpson-Vlach & Ellis, 2010).
- *There is a/an/not* is an identification and focus formulaic sequence, also referred to as a hedging device or a three-word bundle (Chen & Baker, 2010; Hyland, 2008a;

Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). It is highly frequent in academic writing and functions as a sentence stem or an introducer to identify or refer to the existence of specific elements in a text (Byrd & Coxhead, 2010; Erman & Warren, 2000; Simpson-Vlach & Ellis, 2010).

- *There are a number (of)* is a quantity specification formulaic sequence, also referred to as a five-word lexical bundle; it is cataphoric in nature and has a quantifying function in that it specifies the number of the following noun phrase (Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010).
- *It has been (adverb) noted/believed/asserted* is a hedging verb phrase or a sentence head that is extensively used in academic prose (Chen & Baker, 2010; Hyland, 2008a; Nattinger & DeCarrico, 1992; Oakey, 2002). Among the complex discursive functions it executes in academic writing are foregrounding points of discussion objectively, priming topics, presenting arguments, and linking textual elements (Hyland, 2008a; Nattinger & DeCarrico, 1992; Oakey, 2002).
- *The (X) shows* is a sentence head that is multifunctional in nature and is highly recurrent in academic writing; it is used to execute such discursive functions as highlighting the existence of discourse elements, reporting information in a text, and describing textual elements (Liu, 2012; Nattinger & DeCarrico, 1992).
- *The (X) present(s)* is a sentence head that is extracted from the COCA as a synonym for *The (X) shows* due to its high frequency in the academic register, that is, 38.27 occurrences per one million words.
- *It is important to* is an evaluation formulaic sequence that is also referred to as a directive stance bundle and is composed of an anticipatory *it* and an adjective (Byrd

& Coxhead, 2010; Chen & Baker, 2010; Hyland, 2008a; Stubbs & Barth, 2003). It either foregrounds the writer's objective evaluation of previously presented information or influences the reader's perception about the importance of given information in a text; it can also direct the reader to execute certain actions (Biber et al., 2004; Chen & Baker, 2010; Hyland, 2008a).

- *It is necessary to* is another evaluation formulaic sequence that is also referred to as modality stance and obligatory directive stance bundle (Biber, 2009; Byrd & Coxhead, 2010; Chen & Baker, 2010; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). It is highly frequent in academic writing and is typically used to express an impersonal obligation, to instruct the reader to perform particular actions, or to influence the reader's evaluation of the presented information (Ädel & Erman, 2012; Biber et al., 2004; Hyland, 2008a; Simpson-Vlach & Ellis, 2010).
- *It is difficult to* is also an evaluation formulaic sequence that is recurrent in multiple disciplines (Byrd & Coxhead, 2010; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). This formulaic sequence, which is also referred to as an ability stance bundle, is composed of an anticipatory *it*, a copular, an adjective and a clause fragment and can be used with modals (e.g. *It would be difficult*); it is often utilized in academic writing to evaluate information presented in a text (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004; Simpson-Vlach & Ellis, 2010).
- *It is possible to* is an expression of ability and possibility formulaic sequence—also referred to as an impersonal ability bundle—that has a fixed structure and is widespread in academic writing (Ädel & Erman, 2012; Biber, 2009; Coxhead & Byrd, 2007; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs &

Barth, 2003). It is utilized as a hedging device to express predictions of future events or to present the writer's objective evaluation of textual elements (Biber et al., 2004; Chen & Baker, 2010; Hyland, 2008a).

- *Is/Are likely to* is one of the most frequent stance expressions in academic writing and is used in order to frame and hedge the writer's attitude towards textual elements (Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2004; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). It is composed of a copular and an adjective and can be modified by such adverbs as *more, less,* and *most* (Biber, 2009; Liu, 2012; Simpson-Vlach & Ellis, 2010).
- *On the basis of* is an intangible framing attribute that is widespread in academic prose (Biber, 2009; Chen & Baker, 2010; Cortes, 2004, 2008; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). This formulaic sequence is composed of a prepositional phrase and is typically followed by a noun or a noun phrase (Cortes, 2004; Simpson-Vlach & Ellis, 2010). It can occupy an initial or a medium position in a sentence and is utilized to present the characteristics of the following noun phrase, frame the basis of a decision, or emphasize meaning (Byrd & Coxhead, 2010; Simpson-Vlach & Ellis, 2010).
- *Based on the* is another intangible framing attribute that is composed of a passive verb and a prepositional phrase fragment (Chen & Baker, 2010; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). It is highly frequent in academic discourse and is used to frame the attributes of the following noun phrase (Biber, 2009; Durrant, 2009; Simpson-Vlach & Ellis, 2010).

- *Not only… but also* is a correlative conjunction that is composed of two parts which, according to the COCA, are highly frequent in the academic register (91.83 and 33.22 respectively). It is mainly utilized in academic writing to show that the head noun has two characteristics or is the result of two things.

Discourse Organizers

The second set includes 15 formulaic sequences that function as discourse markers. Discourse markers, also referred to as discourse organizers or discourse organizing bundles, are used to mark relationships between prior and upcoming discourse elements, connect different discursive elements, signal the introduction of a new point of discussion, and introduce a restatement or an elaboration of a previously proposed discussion point; therefore, they promote coherence in writing (Biber & Barbieri, 2007; Biber et al., 2004; Chen & Baker, 2010; Cortes, 2004; Simpson-Vlach & Ellis, 2010). The 15 discourse organizers selected for this study are:

- *In other words* is a discourse marker that is highly frequent in academic writing; it functions as a connective in that it semantically connects the constituents of clauses or sentences together and, hence, helps smooth transition from one discourse element to another (Ellis & Simpson-Vlach, 2009; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010).
- *In addition to* is a transition signal that is composed of a prepositional phrase and is typically followed by a noun or a noun phrase; it is used to frame and establish additive links between textual elements in academic discourse (Biber, 2009; Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008a).

- *For example* is a discourse organizer that has a fixed structure and is extensively used in academic writing as an exemplifier that introduces new elements of discourse (Byrd & Coxhead, 2010; Liu, 2010; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003).
- *For instance* is another frequently used exemplifier in academic writing; similar to *For example*, it is labeled as a linking adverbial, has a fixed structure, and helps introduce new elements in a text (Conrad, 1999; Liu, 2008, 2012; Shaw, 2009).
- *Such as* is an identification and focus formulaic sequence that serves as an expository phrase and is used for exemplification purposes in academic prose (Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003).
- *At the same time* is a salient referential expression that is composed of a prepositional phrase and is highly recurrent in academic writing (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004, 2008; Ellis & Simpson-Vlach, 2009; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). It functions as a connective to signal transition between discourse constituents and is used either literally to refer to two simultaneous events or figuratively as a synonym for *however* (Byrd & Coxhead, 2010; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010).
- *According to (the)* is an epistemic stance expression, also referred to as an intangible framing attribute and a collocation; it is typically followed by a noun or a noun phrase and is utilized in academic writing to report information from sources other than the writer (Durrant, 2009; Hyland, 2008a; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003).

- *As shown in* is a meta-discourse and textual reference formulaic sequence which is highly frequent in academic writing; it is utilized to structure elements of discourse and direct the reader's attention to other elements or figures in a text (Biber et al., 2004; Chen & Baker, 2010; Hyland, 2008a, 2008b; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003).
- *On the one hand/ On the other hand* are two idiomatic discourse organizers that are perceptually salient and highly frequent in academic writing (Biber, 2009; Byrd & Coxhead, 2010; Ellis & Simpson-Vlach, 2009; Liu, 2012; Simpson-Vlach & Ellis, 2010; Stubbs & Barth, 2003). These preposition-based phrases are labeled with different terms including discourse organizers, contrast devices, transition signals, links, and linking adverbials (Byrd & Coxhead, 2010; Chen & Baker, 2010; Conrad, 1999; Erman & Warren, 2000; Hyland, 2008a, 2008b). They are used in academic prose to contrast information in a text and link elements of discourse together (Biber et al., 2004; Byrd & Coxhead, 2010; Hyland, 2008a; Simpson-Vlach & Ellis, 2010).
- *In conclusion/ In sum/ In summary/ To sum up/ To conclude* are labeled with multiple terms in the literature such as discourse organizers, transition signals, linking adverbials, and summative adverbials (Conrad, 1999, 2000; Liu, 2008; Nattinger & DeCarrico, 1992). These phrases have summation functions in that they mark the end of either a section or an entire text (Conrad, 2000; Liu, 2008; Nattinger & DeCarrico, 1992).

Collocations

The last major category includes 26 collocations, which are defined as habitually occurring word combinations whose meaning may be either totally or partially

transparent and whose co-occurrence is relatively restricted (Coxhead & Byrd, 2007; Howarth, 1998a; Laufer & Waldman, 2011). These collocations were extracted from the COCA in order to equip the EAP students with a range of prefabricated lexical items that are typical of describing increasing, decreasing, fluctuating, and stable graphical trends.

Reporting increasing trends. In order to provide the EAP students with a set of collocations that help them describe different types of increasing trends, seven collocations composed of the headword *increase* and its collocates were extracted from the COCA. According to the COCA, the word *increase* is highly frequent in all registers, with the highest frequency in the academic register (270.22 per one million words). As a noun, *increase* most frequently co-occurs with *significant*, *dramatic*, *rapid*, *sharp*, and *slight*. The verb *increase*, in turn, collocates with *dramatically* and *significantly*. Accordingly, *significant increase*, *dramatic increase*, *rapid increase*, *sharp increase*, *slight increase*, *increase dramatically*, and *increase significantly* were added to the list of formulaic sequences for this study.

Reporting decreasing trends. To provide the EAP students with a set of phrases that can help them describe different types of decreasing trends, a set of collocations with *decrease* (both as a noun and as a verb) and *decline* (as a noun) as their stem words were extracted from the COCA based on their high frequency in the academic register (50.94 and 71.77 per one million words, respectively). According to the COCA *Word and Phrase. Info* function*, decrease*, as a noun, co-occurs most frequently with *significant*, *dramatic*, *slight*, and *sharp*, and, as a verb, it co-occurs with *significantly*, *slightly*, and *dramatically*. The noun *decline*, in turn, collocates with *steady*, *sharp*, *significant*, *rapid*, *gradual*, and *dramatic*.

Decrease and decline, as stem words, along with their collocates constituted thirteen collocations, namely, *significant decrease*, *dramatic decrease*, *sharp decrease*, *slight decrease*, *decrease significantly*, *decrease slightly*, *decrease dramatically*, *steady decline*, *sharp decline*, *significant decline*, *rapid decline*, *gradual decline*, and *dramatic decline*. It is worth mentioning here that *decline* as a verb was not included in the list of collocations because none of its adjacent collocates in the COCA (*year*, *rate*, *population*, *price*, etc.) is semantically related to the manner of decrease in prices, percentages, numbers, and so on.

Reporting stable trends. The verb *remain* was used as a stem word for two collocations (i.e., *remain stable* and *remain the same*) which can be implemented when reporting trends that remain relatively invariable over a time period. According to the COCA, the verb *remain* is highly frequent in all registers, but its highest frequency is in the academic register (120.01 per one million words). When checking the collocates that occur with the verb *remain*, only *same* could be extracted; however, doing a reverse search in the *Word and Phrase. Info* function by using *stable* as the stem word helped extract the second collocation (i.e., *remain stable*) because *remain* is listed as the most frequent verb that co-occurs with the adjective *stable*.

Reporting fluctuating trends. Four collocations were extracted from the COCA to help the participants describe instability, namely, *rise and fall (N)*, *rise and fall (V)*, *the fluctuation in the price*, and *prices fluctuate*. The first two collocations were constructed by utilizing one of the sentence builders included in the list of formulaic sequences above, that is *(A) and (B)*. The rationale behind including these two collocations relate to the high frequency of both *rise* and *fall* in the academic register (87.44 and 107.83 per

one million words, respectively) and the fact that they, according to *Word and Phrase. Info*, are collocates of each other. The noun *price*, in turn, has been used as the stem word for the last two collocations due to its high frequency in the academic register (111.48 per one million word). While neither *fluctuate* nor *fluctuation* is among the words that co-occur with the noun *price*, *price* is the most frequent co-occurring noun with *fluctuate* and the second most frequent noun that co-occurs with *fluctuation*.

Appendix B: Analytical Rubrics

Table B1

Analytical Rubric for Graphical Summaries

	Unsatisfactory Possible points: 10 – 40	**Average/Good** Possible points: 50 – 70	**Excellent** Possible points: 80 – 100
Content	Missing and/or inaccurate information Reporting each trend in a separate sentence without any attempt to report similar trends together and/or contrast opposite trends in the text	Most of the information presented in the graph is accurately reported. Some of the similar trends are reported together. Some opposite trends are contrasted in the text	Accurate and proper reporting of the information presented in the graph. Most similar trends are reported together & opposite trends are contrasted in the text.
Format & Paragraph elements	No paragraph format Missing introductory and/or concluding statements **OR** Introductory sentence is too specific (reporting a specific trend rather than the information presented in the graph as a whole) Concluding sentence presents new information that relates to the trends rather than a restatement of the introductory sentence	Paragraph format is satisfactory Introductory and concluding statements are included but could have been better	Appropriate paragraph format (text is written as one chunk and the first sentence is indented) Introductory sentence summarizes the graph properly Concluding sentence restates the introductory statement and may/may not include a final statement by the writer
Vocabulary & Spelling	Poor vocabulary range Informal language Incorrect word usage Many spelling mistakes/errors Lexical mistakes/errors impede comprehension	Some repetition is detected Acceptable word usage with some lexical mistakes/errors Some spelling mistakes/errors	A good vocabulary range Good word usage with very few lexical mistakes/errors A few spelling mistakes/errors
Grammar &	Many grammatical mistakes/errors that relate to subject-verb agreement, prepositions, tenses, parts of speech, or	Acceptable use of grammar, but some errors/mistakes persist Occasional run-on sentences and/or fragments	Very few grammatical mistakes Very few/no run-on sentences and/or fragments Punctuation is

Structure	missing verbs Many run-on sentences and/or fragments Inaccurate/no punctuation Grammatical mistakes/errors impede comprehension	Some punctuation errors/mistakes are still present Errors/mistakes do not impede comprehension	accurate with very few errors/mistakes
Coherence & Cohesion	No smooth transition from one sentence to another No reference to the source of information No coherence and/or cohesion	Acceptable transition from one sentence to another but could have been better Reference to the source of information is made The text is coherent and cohesive but could have been better	Good transition from one sentence to another Reference to the source of information is made The text is coherent and cohesive

Table B2

Analytical Rubric for Essays

	Unsatisfactory Possible points: 10 – 40	**Average/Good** Possible points: 50 – 70	**Excellent** Possible points: 80 – 100
Content	Impoverished content No controlling ideas are presented Controlling ideas are not/hardly supported Some ideas are irrelevant to the topic	The controlling idea of each paragraph is adequately supported with reference to facts and/or personal experience Almost all the supporting ideas/details are relevant to the topic	The controlling idea in each body paragraph is thoroughly supported with reference to facts and/or personal experience All the supporting ideas/details are relevant to the topic

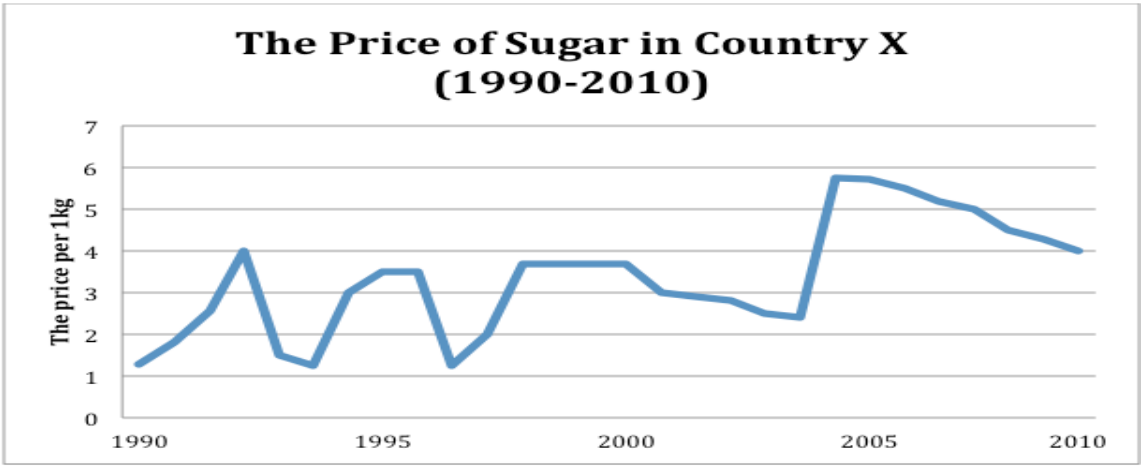
Format & Essay elements	No essay format Missing introduction and/or conclusion Introduction is too short or too specific (with/without a thesis statement) Thesis statement is too general The conclusion presents new information other than the writer's final thought	Essay format is satisfactory (one of the paragraphs may be too short) An introduction with a thesis statement and a conclusion are included but need improvement	Appropriate essay format (4 or more paragraphs of appropriate length) Introduction provides background information and/or has a hook and a clear/strong thesis statement The conclusion reiterates the major points (may/may not include a final statement by the writer)
Vocabulary & Spelling	Poor vocabulary range Informal language Incorrect word usage Many spelling mistakes/errors Lexical mistakes/errors impede comprehension	Some repetition is detected Acceptable word usage with some lexical mistakes/errors Some spelling mistakes/errors	A good vocabulary range Good word usage with very few lexical mistakes/errors A few spelling mistakes/errors
Grammar & Structure	Many grammatical mistakes/errors that relate to subject-verb agreement, prepositions, tenses, parts of speech, or missing verbs Many run-on sentences and/or fragments Inaccurate/no punctuation Grammatical mistakes/errors impede comprehension	Acceptable use of grammar but some errors/mistakes persist Occasional run-on sentences and/or fragments Some punctuation errors/mistakes are still present Errors/mistakes do not impede comprehension	Very few grammatical mistakes Very few/no run-on sentences and/or fragments Punctuation is accurate with very few errors/mistakes
Coherence & Cohesion	No smooth transition from one sentence/paragraph to another The essay is neither coherent nor cohesive	Acceptable transition from one sentence/paragraph to another but needs improvement Overall, the essay is relatively coherent and cohesive	Good transition from one sentence/paragraph to another The essay is coherent and cohesive

Appendix C: Pedagogical Instruments

Instruments for the Focused Instruction Groups

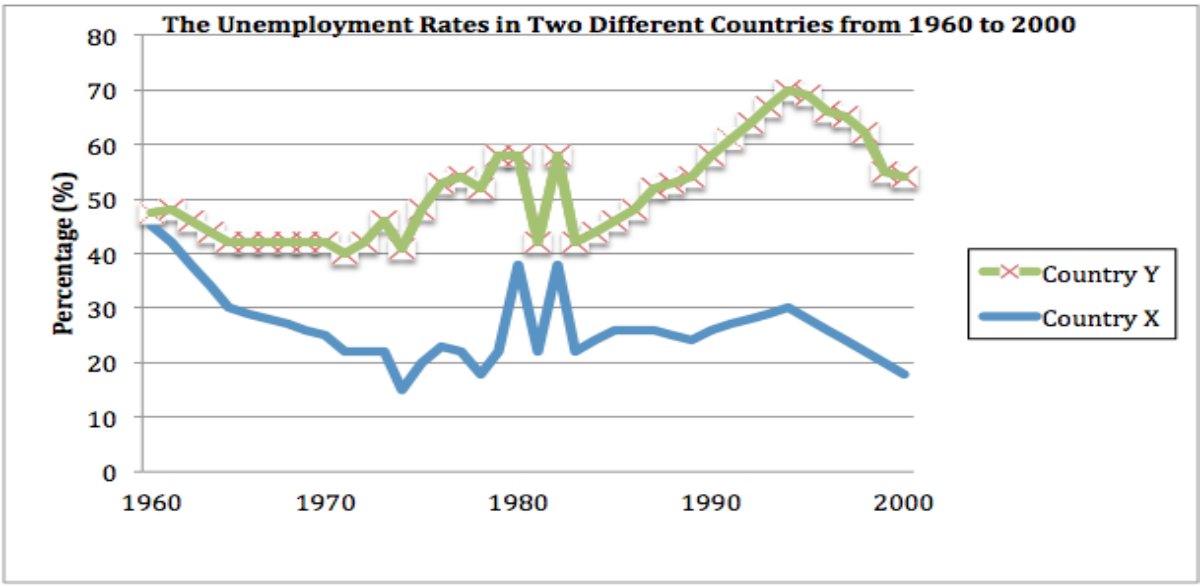
Worksheet 1: Writing a Topic Sentence

A topic sentence is usually the first sentence in a paragraph and is used to introduce the reader to the main points of discussion. When reporting graphical information, make sure to refer to the trends presented in the graph along with the time range, if any.

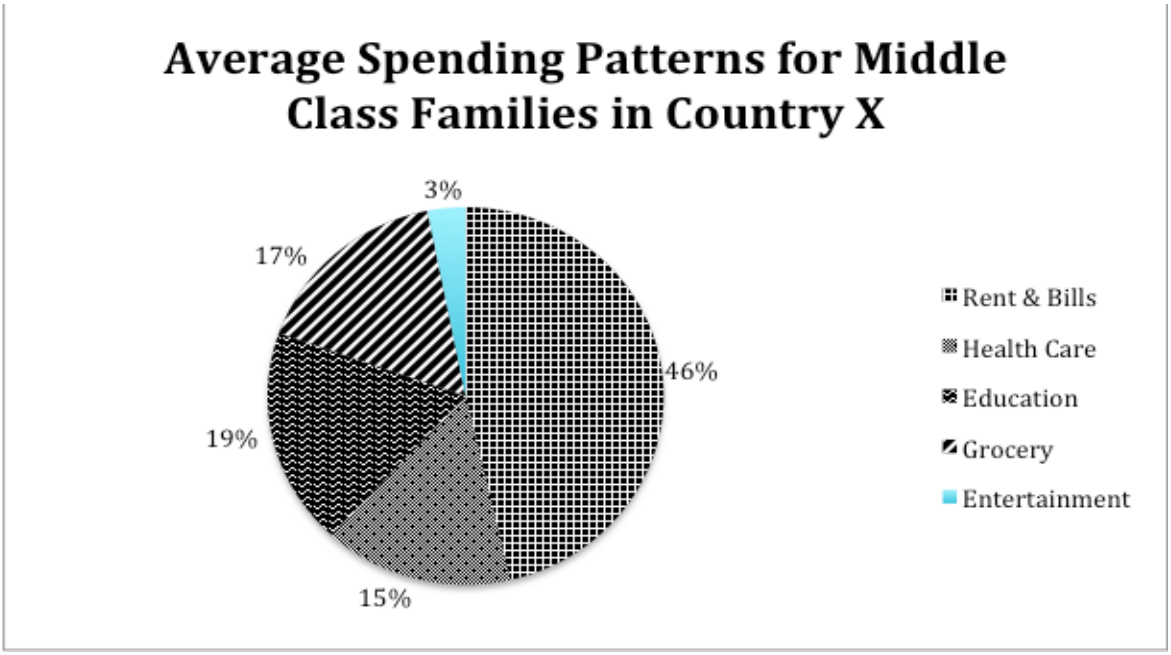


The line graph presents the rise and fall in the price of sugar over a twenty-year period.
OR
The line graph shows the fluctuation in the price of sugar between 1990 and 2010.

The bar graph …… the …… of …….……. in ……..



The line graph ……. the ….…… in …….. from …….. to ……..

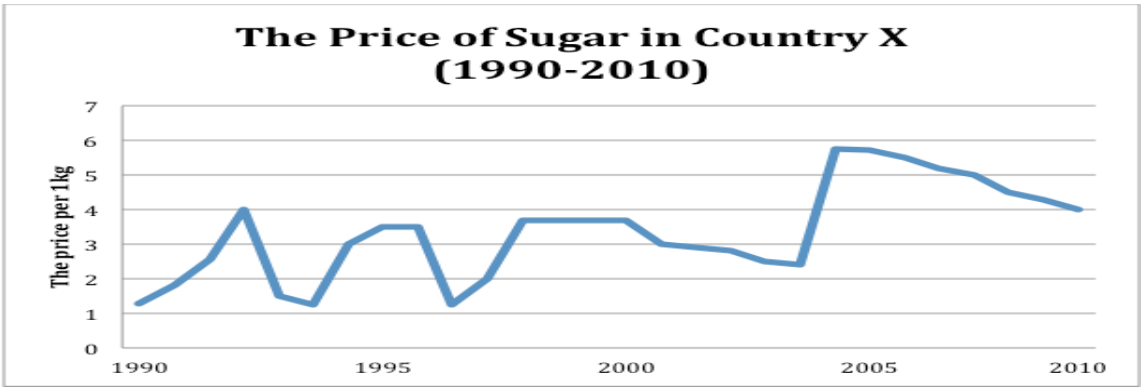


……

……

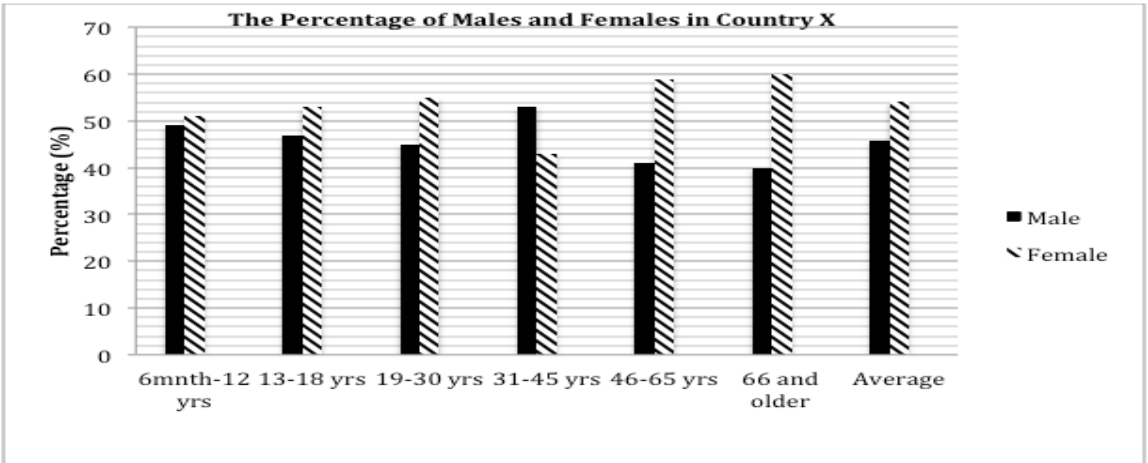
Worksheet 2: Writing a Concluding Sentence

A concluding sentence is usually the last sentence in a paragraph and is used to reiterate the main points of discussion already presented. When reporting graphical information, make sure to restate the general tendency of the trends presented in the graph.



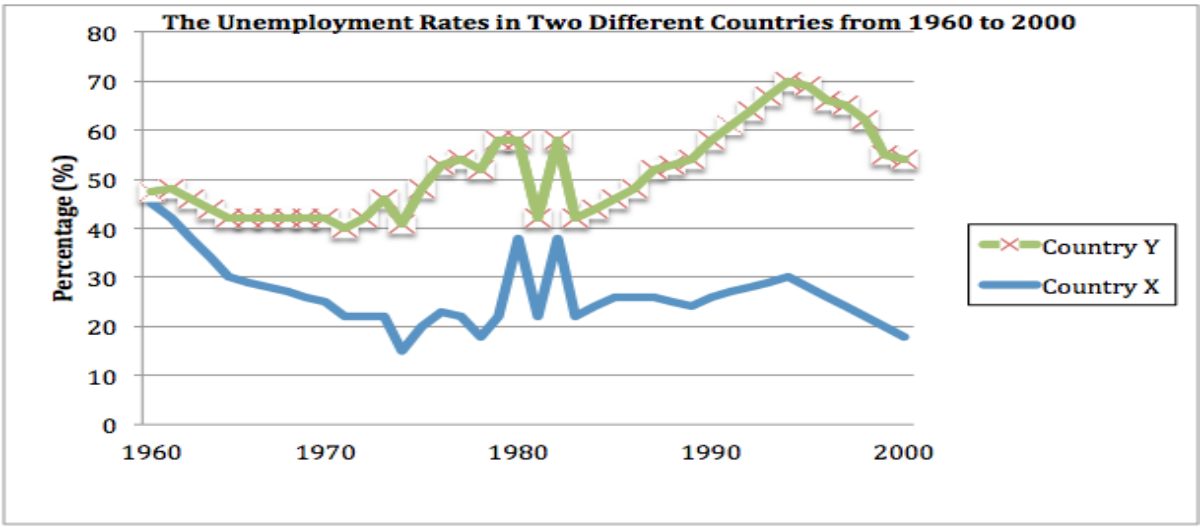
In conclusion, it can be seen from the data given that the price of sugar fluctuated over a twenty-year period. OR

As can be seen, there was a fluctuation in the price of sugar between 1990 and 2010.



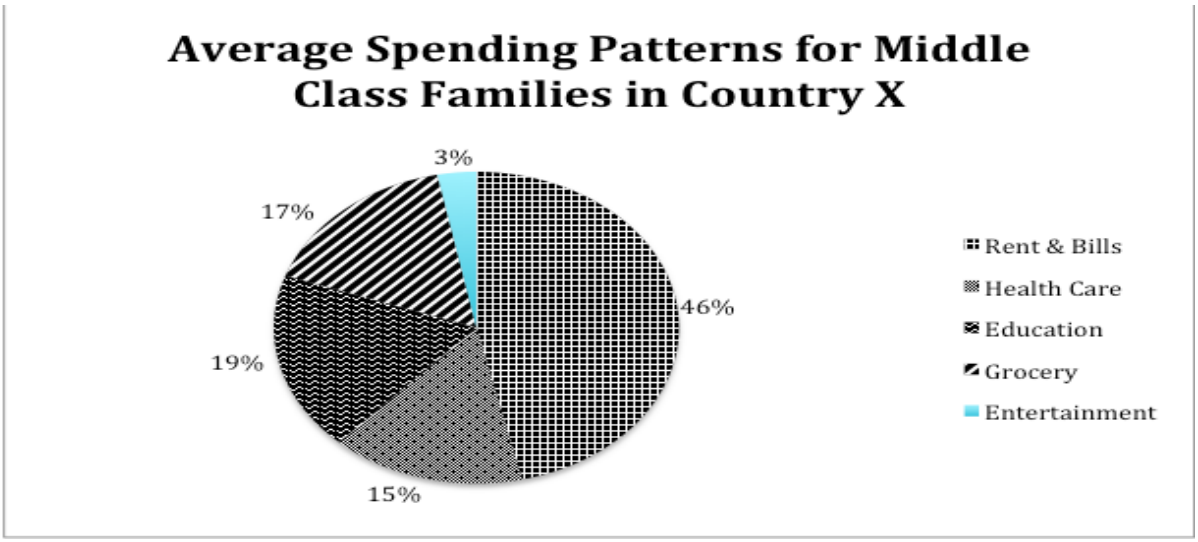
……

……



.....

.....



.....

.....

***Concluding sentences for dictations**

To sum up, it is clear that the percentage of females is higher than that of males in almost all age groups.

To conclude, the rates of unemployment in both country X and country Y fluctuated between 1960 and 2000.

Worksheet 3: Collocations

There are certain collocations that can be used when reporting changes of a particular phenomenon, i.e., prices, rates, percentages, etc. These can be noun phrases or verb phrases and, thus, differ in use and structure.

Noun phrases: There is a + Noun Phrase + the rest of the sentence OR
The Noun phrase + V + the rest of the sentence

Verb phrases: S + Verb Phrase + the rest of the sentence

Increase

In the case of increase (going up), you can use

Noun phrases: significant increase, dramatic increase, sharp increase, slight increase, rapid increase

Verb phrases: increase dramatically, increase significantly

- There is a dramatic increase in the price of rice. (Noun Phrase)
- The dramatic increase in the price of rice affected consumption rates. (Noun Phrase)
- The price of rice increased dramatically. (Verb Phrase)

Decrease

In the case of decrease (going down), you can use

Noun phrases: significant decrease, dramatic decrease, sharp decrease, slight decrease

Verb phrases: decrease significantly, decrease slightly, decrease dramatically

- There is a dramatic decrease in the price of rice. (Noun Phrase)
- The dramatic decrease in the price of rice affected consumption rates. (Noun Phrase)
- The price of rice decreased dramatically. (Verb Phrase)

You can also use “decline” with an adjective to present a decrease

Noun phrases: steady decline, sharp decline, significant decline, rapid decline, gradual decline, dramatic decline

- There is a sharp decline in the price of rice. (Noun Phrase)
- The sharp decline in the price of rice affected consumption rates. (Noun Phrase)

Fluctuation

In the case of fluctuation (rise and fall), you can use

Noun phrases: the rise and fall, the fluctuation in the price

Verb phrases: rise and fall, prices fluctuate

- The rise and fall in the price of rice affected consumption rates. (Noun Phrase)
- The fluctuation in the price of rice affected consumption rates. (Noun Phrase)
- The price of rice rose and fell. (Verb Phrase)
- The price of rice fluctuated. (Verb Phrase)

Stability

If the rates/percentages remain the same, you can use

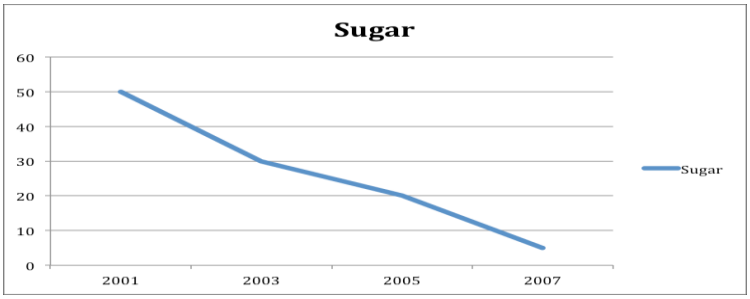
Verb phrases: remain the same, remain stable

- The price of rice remained stable. (Verb Phrase)
- The price of rice remained the same. (Verb Phrase)

Worksheet 4: Collocation Practice

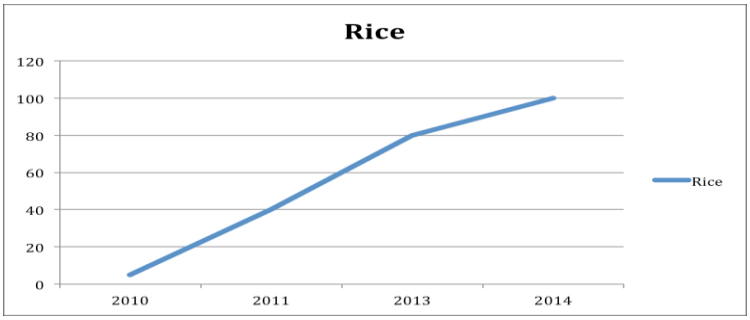
Write two meaningful sentences to describe the trends presented in the charts below. Use the phrases in the box below.

rise and fall (V)	decrease dramatically		remain stable
decrease slightly	sharp increase		significant decline
increase dramatically	gradual decline	the fluctuation in the	remain the same



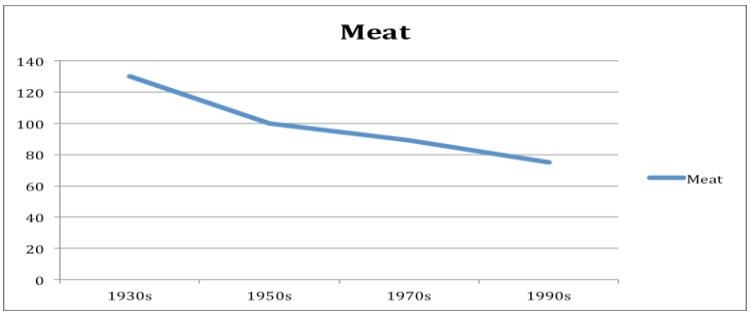
……

……



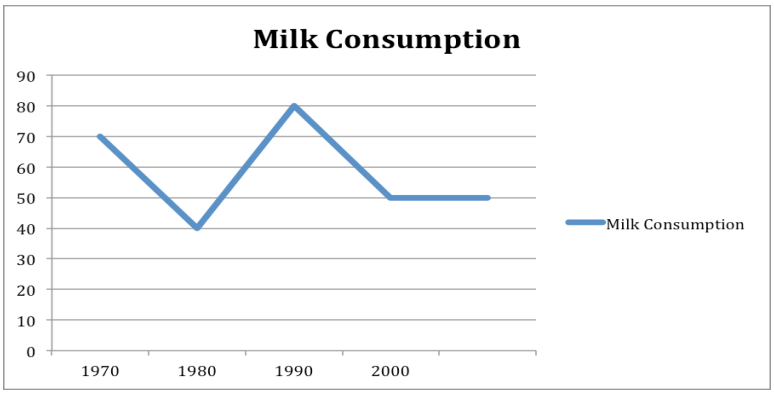
……

……



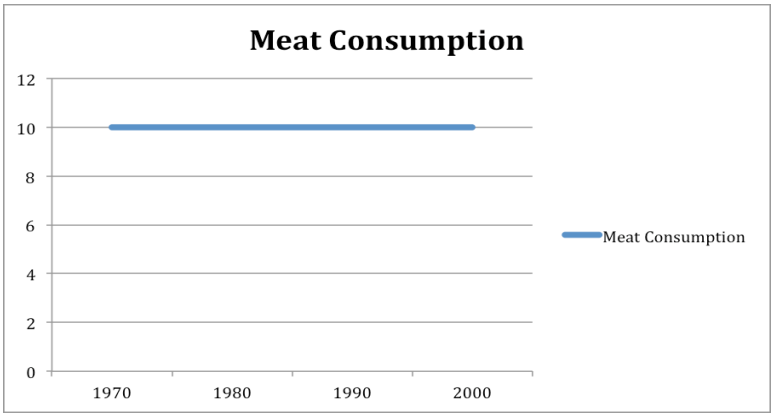
……

……


Milk Consumption
90
80
70
60
50
40
30
20
10
0
1970
1980
1990
2000
Milk Consumption

. . . . . .

. . . . . .


Meat Consumption
12
10
8
6
4
2
0
1970
1980
1990
2000
Meat Consumption

. . . . . .

. . . . . .

Worksheet 5: Sentence Builders

Sentence builders are typically used to introduce the subject of a sentence and, in some cases, introduce your opinion objectively. NOTE!! "S" means a complete sentence, i.e., it includes a subject, a verb, and a complement/an object. "V" means a verb in the base form.

1. (A) and (B)/ Both (A) and (B)/ (A) as well as (B)

These phrases are used to present two subjects/objects that have the same function and/or share the same characteristics in one sentence and, thus, reduce repetition in writing. Instead of writing "Canada is a developed country. The US is a developed country." You can write

- Canada and the US are developed countries. OR
- Both Canada and the US are developed countries. OR
- Canada as well as the US are developed countries.

Note that you can also write: Both of these countries are developed.

2. The (A) of (B)

This is one of the most common phrases used to introduce a complex noun phrase as a subject or an object.

- The advantages of technology outweigh its disadvantages.
- People should not underestimate the advantages of technology.

3. There is a/an/not OR There are a number

These two sentence builders are used to introduce a subject in a sentence. NOTE!! "There is" is used in the case of a singular subject. "There are" is used in the case of a plural subject.

- There is an important reason why smoking should be banned in public places.
- There are a number of reasons why smoking should be banned in public places.

4. It has been generally believed/ noted/ asserted that + S

It is used, <u>typically in introductions</u>, when the writer wants to show that an opinion/ a phenomenon is widely accepted by many people. Thus, it helps writers present their opinion objectively without using such phrases as I think, I believe, in my opinion, etc.

- It has been generally believed that learning a second language is challenging for many adults.

5. It can be seen/observed that + S

This clause can be used to introduce your opinion, typically in conclusions, with reference/in relation to the information you have already presented in your body paragraphs.

- It can be seen that advances in technology have both positive and negative effects on humanity at large.

6. It is important/necessary to + V

These phrases are typically used to introduce suggestions which the writer thinks are essential.

- It is necessary to exercise everyday.
- It is important to exercise everyday.

7. It is difficult to + V

This phrase is used to refer to things that cannot be accomplished/done easily.

- It is difficult to jog in winter.

8. It is possible

This phrase is used to introduce a possibility and can be used in two ways:

- It is possible to + V: It is possible to buy books online.
- It is possible that + S: It is possible that he is so busy.

9. (A) is/are likely to + V

This phrase is used to hedge and avoid generalization; that is, to acknowledge any possible exceptions for something.

- Using phrases is likely to alleviate grammatical errors in writing.
- Salt and sugar are likely to cause many health problems.

Worksheet 6: Practice

Put the following words in meaningful sentences. Use the clues provided between brackets.

1. Both (A) and (B): (hard work, persistence, success)

…….

2. It has been generally believed that…: (car-share programs, pollution)

…….

3. It is possible to: (drive, 18 years old)

…….

4. There is a/an: (penalty, late submissions)

…….

5. The (A) of (B): (cereal price, increased)

…….

6. It can be observed that (registration numbers, increased dramatically)

…….

7. It is important to (practice, the exam)

…….

8. It is possible that (human actions, global warming)

…….

9. There are a number (using phones, driving)

…….

10. is likely to (listening to songs, listening skills)

…….

11. It is difficult to (independent)

…….

Worksheet 7: Cause/Effect & Compare/Contrast

Some phrases/clauses are frequently used to compare and contrast OR to reveal cause and effect in writing. These phrases, for the most part, come in a fixed structure. NOTE!! "S" means a complete sentence.

1. (A) play$_{(s)}$ an important role in (B)

It is mainly used in writing to show how one thing (the subject of the sentence) can positively affect something else (the object of the sentence).

- Practice plays an important role in language learning.

2. S + due to the fact that + S

This is mainly used to show the causal relationship between the two parts of the sentence. Note!! This phrase can be used in two different ways.

- John got full marks due to the fact that he studied hard.
- Due to the fact that he studied hard, John got full marks.

3. As a result of

This phrase is typically used in writing to show that one thing is the result of something else. Note!! This phrase can be used in two different ways.

- She needed support as a result of her nervousness. OR
- As a result of her nervousness, she needed support.

4. The reason for (A) is/was (B) OR (B) is/was the reason for (A)

This phrase is used to show how one thing (A) can be caused by something else (B)

- The reason for infection is germs.
- Germs are the reason for infection.

5. On the basis of/Based on the

These two phrases are used in academic writing to show that one thing has happened because of/as a result of something else.

- The decision was made on the basis of the current situation.
- The decision was made based on the current situation.

6. On the one hand/On the other hand

These two phrases are mainly used when you write two contrasting ideas or paragraphs. In other words, the second sentence/paragraph presents information that is opposite to the first sentence/paragraph. Note!! You can use “on the other hand” without “on the one hand.” However, you CANNOT use “on the one hand” without including “on the other hand.”

- On the one hand, advances in technology have made the world a small village. On the other hand, they have many negative effects on humanity at large. ✓
- Advances in technology have made the world a small village. On the other hand, they have many negative effects on humanity at large. ✓
- ~~On the one hand, advances in technology have made the world a small village. They have many negative effects on humanity at large.~~ ✗

7. Not only … but also

This correlative conjunction is used to show that the head noun has two characteristics. Note!! If you start the sentence with ‘not only,’ remember to invert (Question structure).

- The car is not only economical, but it also feels good to drive.
- Not only is the car economical, but it also feels good to drive.

Worksheet 8: Mingle & Sentence Writing Activities

1. Simplicity is the reason for the popularity of these languages.
2. Group work plays an important role in social development.
3. He found the program really difficult due to the fact that he came from a different educational background.
4. The price of several products increased as a result of economic depression.
5. On the one hand, I want to buy a big house. On the other hand, I do not have enough money.
6. Not only did he invent the Internet, but he also invented the calculator.
7. He was promoted to manager on the basis of his hard work and commitment.

Put the following phrases in meaningful sentences.

The reason for:……...……

……

Play an important role in:…….…….....

……

Due to the fact that:…….....

……..…..

As a result of:…….....

……

On the one hand/on the other hand:……...

……

Not only…but also:…….....

……

Based on:…….....

……

Worksheet 9: Discourse Organizers

Some phrases are used in academic writing to help writers organize their texts and establish smooth transitions between sentences and/or paragraphs. In addition to text organization, they tell your readers what to expect next.

1. In other words,
This phrase is typically used when writers present already introduced information in different words due to the complexity and/or lack of clarity of the information in the previous sentence. Note!! This phrase comes at the beginning of a sentence and is followed by a comma.

- The listening test is really difficult. In other words, you need to practice a lot in order to pass the test.

2. In addition to (A),
This phrase is used when writers add information. It can be used between sentences or at the beginning of your second body paragraph to connect ideas together and help smooth transition from one sentence/paragraph to another. Note!! This phrase comes at the beginning of a sentence and is followed by a comma.

- In addition to lung cancer, smoking is the main reason behind many heart diseases.

3. For example, For instance,
These phrases are mainly used to list examples in a text. You can use these phrases in different ways. Note!! If you use these phrases at the beginning of a sentence, add a comma after them.

- You can practice listening in various ways. For example, you can watch movies or listen to the radio.
- You can practice listening in various ways. For instance, you can watch movies or listen to the radio.

You can also use these phrases at the end of the second sentence. Note!! They are preceded by a comma.

- You can practice listening in various ways. You can watch movies or listen to the radio, for example.
- You can practice listening in various ways. You can watch movies or listen to the radio, for instance.

4. Such as
This phrase, like for example and for instance, is typically used to introduce examples. Note!! It is followed by nouns NOT a sentence.

- I like bright colors such as red and yellow.

5. At the same time,
This phrase, in addition to its literal meaning, is used figuratively as a synonym for however/nevertheless. Note!! This phrase comes at the beginning of a sentence and is followed by a comma.

- We agree with your suggestions. At the same time, we object to your methods.

6. Between (A) and (B)/ From (A) to (B)

These two phrases can be used to introduce time frames and/or price range. Since these phrases come at the beginning or at the end of a sentence, they help produce various sentence structures.

- From 1999 to 2009, John lived in France. Or
- John lived in France from 1999 to 2009.
- Between 1999 and 2009, John lived in France. Or
- John lived in France between 1999 and 2008.

You can also use over a period of ….

- This chart presents the fluctuation in the prices of oil over a period of ten years.

7. At the beginning of…/At the end of….

As their literal meaning indicate, these two phrases are used to refer to the beginning or the end of a thing or a time period. They can be used at the beginning of a sentence with a comma or at the end of a sentence without a comma.

- We had to identify the weaknesses of each design at the beginning of the project. Or
- At the beginning of the project, we had to identify the weaknesses of each design.
- The price of rice decreased sharply at the end of 2010. OR
- At the end of 2010, the price of rice decreased sharply.

8. According to the (A),/As shown in (A),

These phrases are typically used to make reference to the source of information (a chart, an article, or a writer). Note!! These phrases come at the beginning of a sentence and are followed by a <u>comma.</u>

- According to the graph, there is a sharp increase in the number of people using the subway between 6:00 and 8:00 am.
- As shown in the graph, there is a sharp increase in the number of people using the subway between 6:00 and 8:00 am.

9. In conclusion/ In sum/ To sum up/ To conclude/ In summary,
These phrases are used at the beginning of a concluding statement/paragraph. Thus, they mark the beginning of your conclusion. Note!! These phrases come at the beginning of a sentence and are followed by a <u>comma</u>.

- In conclusion, it can be noted that advances in technology have both advantages and disadvantages.

Worksheet 10: Practice

I. Rewrite the following sentences using the transition phrases in brackets.

1. John has a lot of homework to do. He is too busy these days. (In other words)

……

2. The change in the atmospheric temperature affects wild animals. Vegetables are adversely affected by global warming. (In addition to A)

……

3. The migration of various animals and birds is one of the indicators of increasing temperatures. Foxes and whales are now found in the areas away from their natural habitat. (For instance)

……..

4. I like historical movies. I like Brave Heart and The Patriot. (Such as)

…….

5. John paid for the damage. He was very angry about the water leak. (At the same time)

……

6. It can be seen that global warming has negative effects on both animals and plants. (In sum)

……

7. The graph shows how medical practices changed worldwide. (Add a period of time using from… to…./ 1900-2000)

……

8. The prices of several diary products increased between 2005 and 2007. (According to the graph)

……

9. The number of accounting students is significantly higher than the number of engineering students. (As shown in the table below)

……

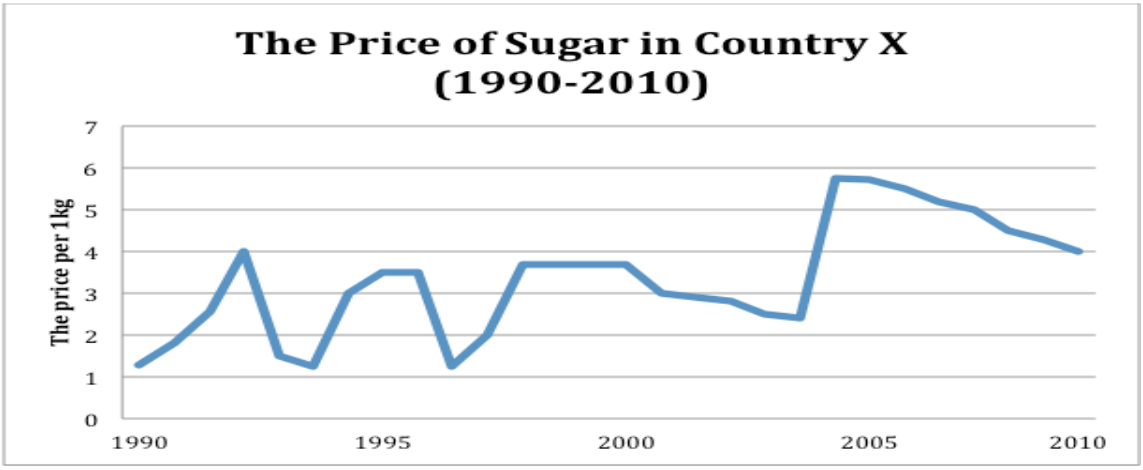
10. You need to submit a detailed report. (at the end of this month)

……

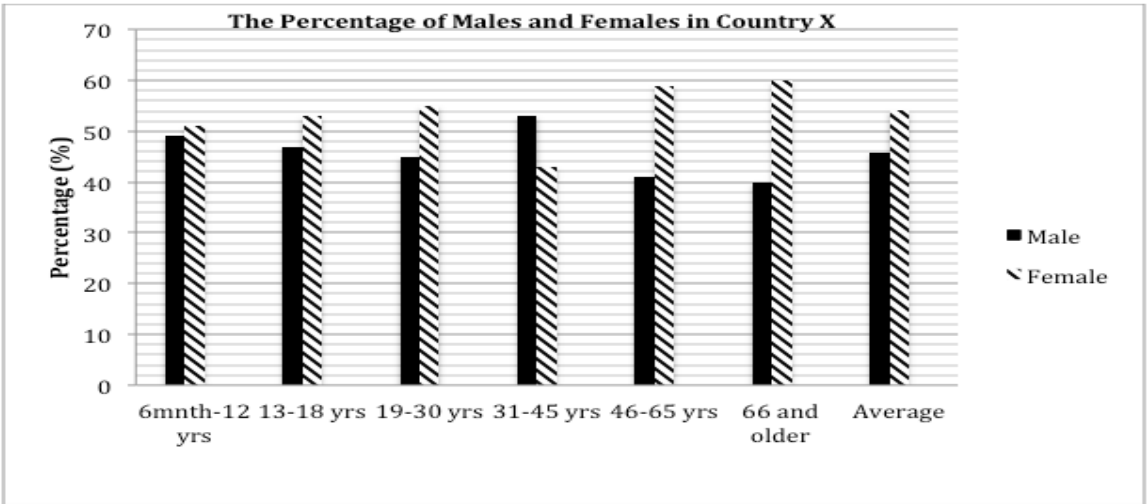
Instruments for the Saliency Groups

Worksheet 11: Writing a Topic Sentence

A topic sentence is usually the first sentence in a paragraph and is used to introduce the reader to the main points of discussion. When reporting graphical information, make sure to refer to the trends presented in the graph along with the time range, if any.

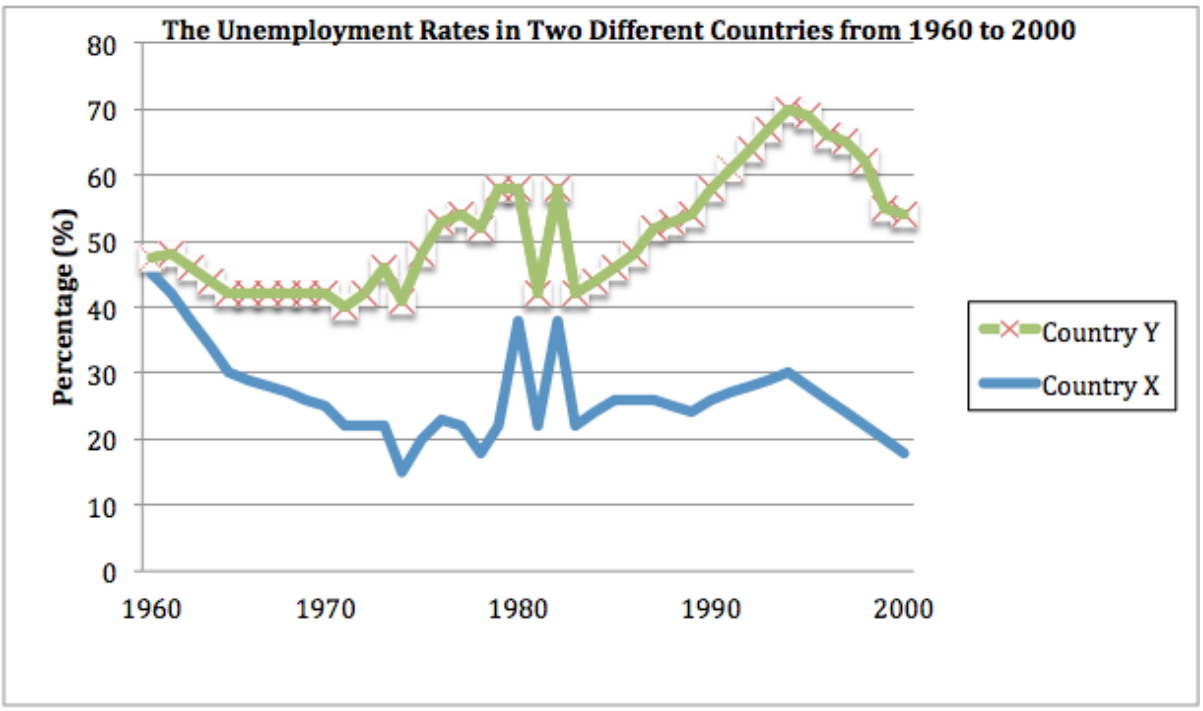


The **line graph presents the rise and fall** in the price of sugar **over a sixty-four-year period.**
OR
The line graph shows the fluctuation in the price of sugar between 1960 and 2006.



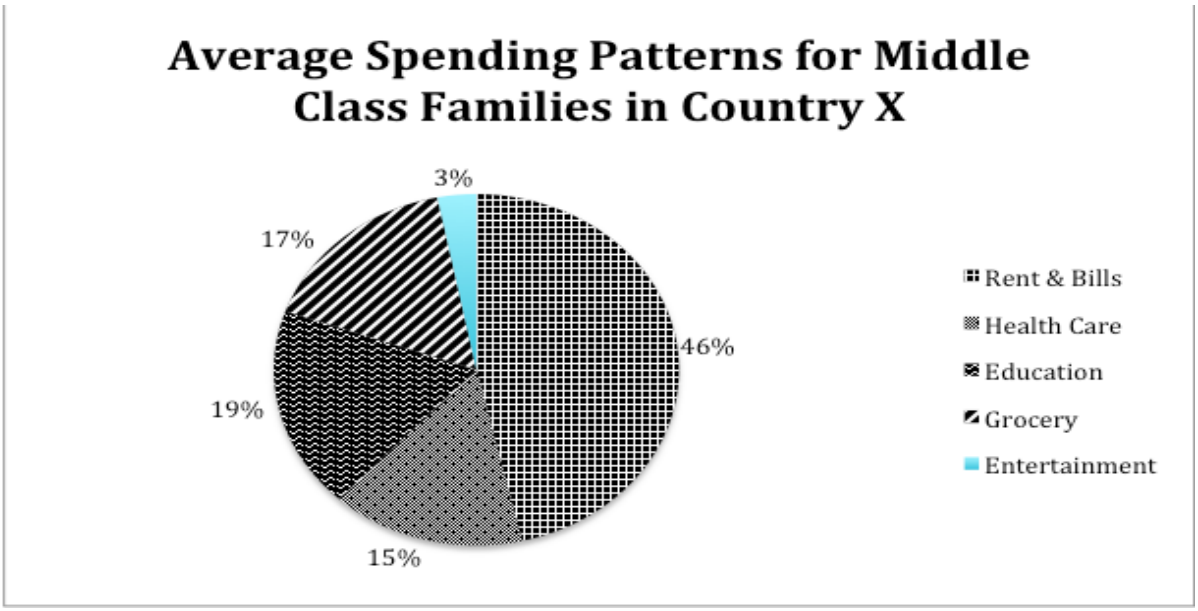
The bar graph …… the …… of ……..…… in

……..



The line graph ……. the …..…… in …….. from …….. to

……..

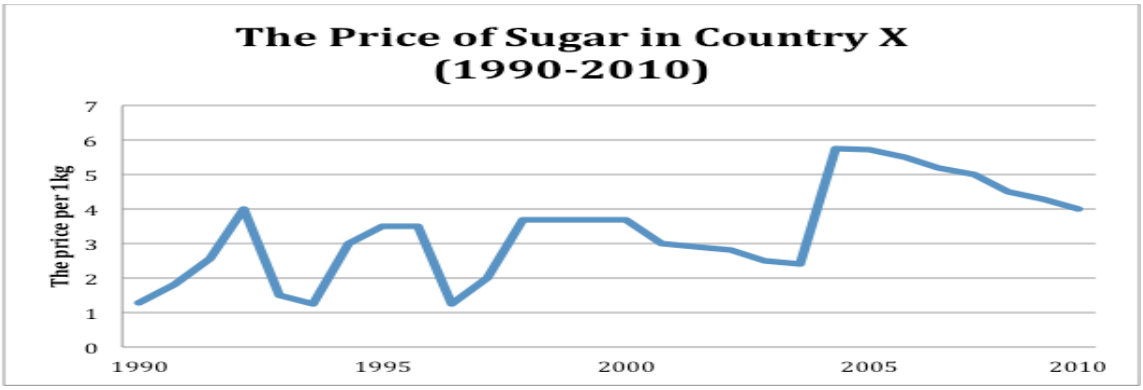


……

……

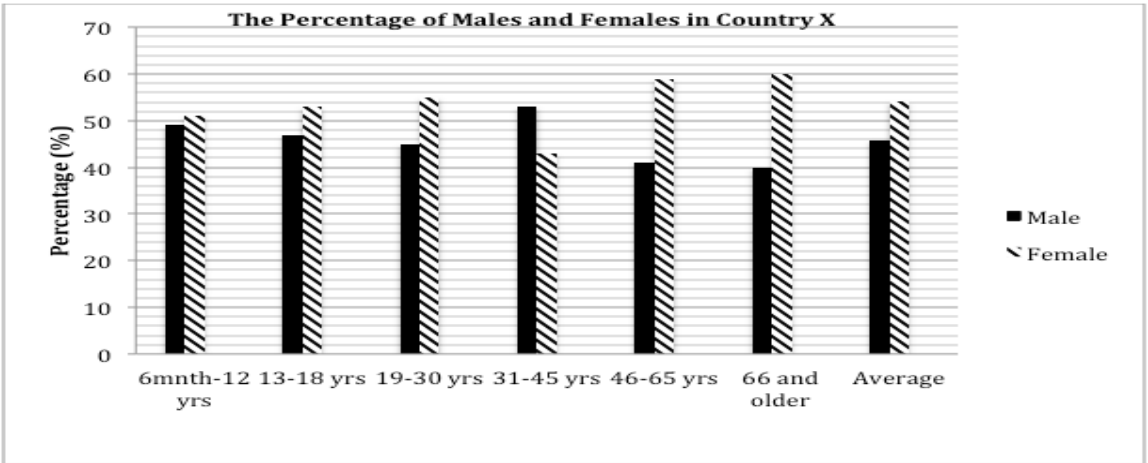
Worksheet 12: Writing a Concluding Sentence

A concluding sentence is usually the last sentence in a paragraph and is used to reiterate the main points of discussion already presented. When reporting graphical information, make sure to restate the general tendency of the trends presented in the graph.



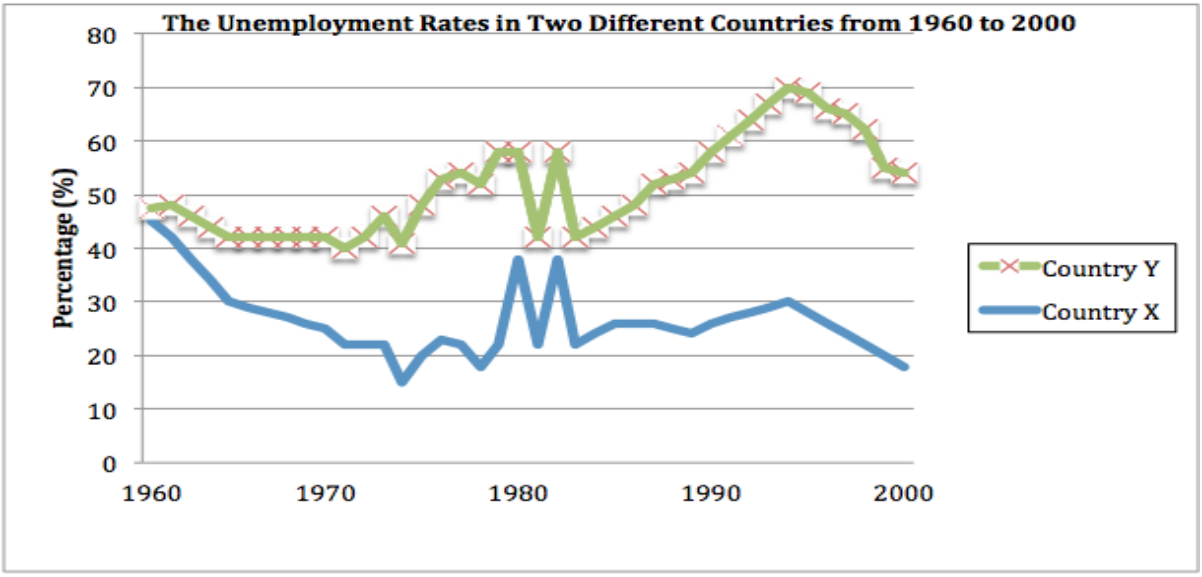
In conclusion, **it can be seen from the data given** that **the price of sugar fluctuated** over a twenty-year period. OR

As can be seen, **there was a fluctuation** in the price of sugar .



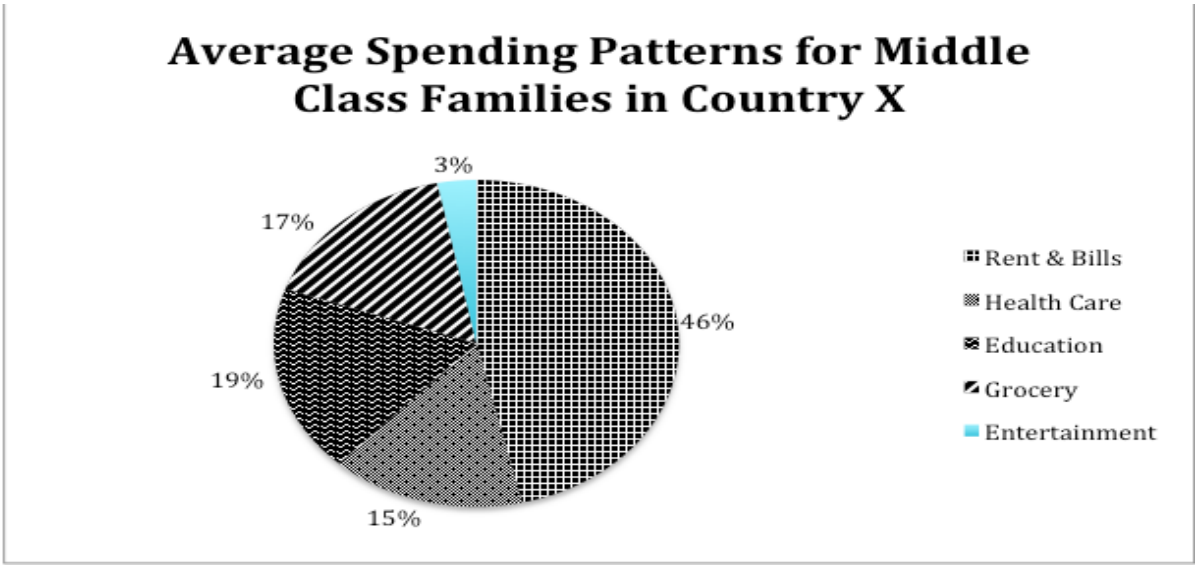
……

……



.....

.....



.....

.....

***Concluding sentences for dictations**

To sum up, it is clear that the percentage of females is higher than that of males in almost all age groups.

To conclude, the rates of unemployment in both country X and country Y fluctuated between 1960 and 2000.

Worksheet 13: Collocations

There are certain collocations that can be used when reporting changes of a particular phenomenon, i.e., prices, rates, percentages, etc. These can be noun phrases or verb phrases and, thus, differ in use and structure.

Noun phrases: There is/was a + Noun Phrase + the rest of the sentence OR
The Noun Phrase + V + the rest of the sentence

Verb phrases: S + Verb Phrase + the rest of the sentence

Increase

In the case of increase (going up), you can use

Noun phrases: significant increase, dramatic increase, sharp increase, slight increase, rapid increase

Verb phrases: increase dramatically, increase significantly

- There was a dramatic increase in the price of rice. (Noun Phrase)
- The dramatic increase in the price of rice affected consumption rates. (Noun Phrase)
- The price of rice increased dramatically. (Verb Phrase)

Decrease

In the case of decrease (going down), you can use

Noun phrases: significant decrease, dramatic decrease, sharp decrease, slight decrease

Verb phrases: decrease significantly, decrease slightly, decrease dramatically

- There was a dramatic decrease in the price of rice. (Noun Phrase)
- The dramatic decrease in the price of rice affected consumption rates. (Noun Phrase)
- The price of rice decreased dramatically. (Verb Phrase)

You can also use "decline" with an adjective to present a decrease

Noun phrases: steady decline, sharp decline, significant decline, rapid decline, gradual decline, dramatic decline

- There was a sharp decline in the price of rice. (Noun Phrase)
- The sharp decline in the price of rice affected consumption rates. (Noun Phrase)

Fluctuation

In the case of fluctuation (rise and fall), you can use

Noun phrases: the rise and fall, the fluctuation in the price

Verb phrases: rise and fall, prices fluctuate

- The rise and fall in the price of rice affected consumption rates. (Noun Phrase)
- The fluctuation in the price of rice affected consumption rates. (Noun Phrase)
- The price of rice rose and fell. (Verb Phrase)
- The price of rice fluctuated. (Verb Phrase)

Stability

If the rates/percentages remain the same, you can use

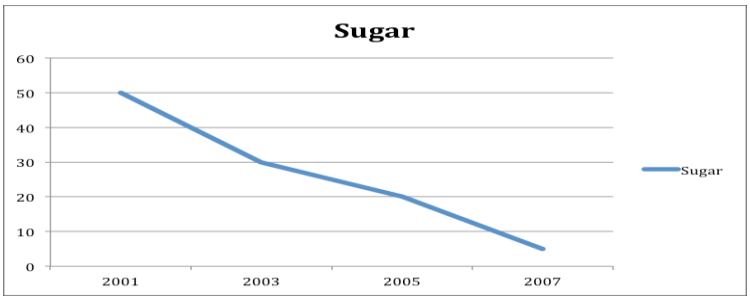
Verb phrases: remain the same, remain stable

- The price of rice remained stable. (Verb Phrase)
- The price of rice remained the same. (Verb Phrase)

Worksheet 14: Collocation Practice

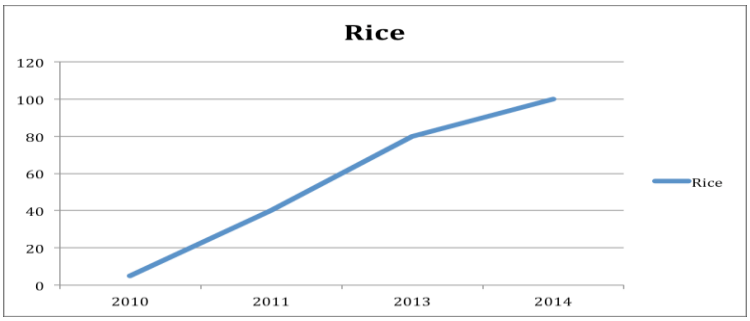
Write two meaningful sentences to describe the trends presented in the charts below. Use the phrases in the box below.

rise and fall (V)	decrease dramatically	remain stable	decrease slightly
sharp increase	significant decline	increase dramatically	gradual decline
the fluctuation in the	remain the same		



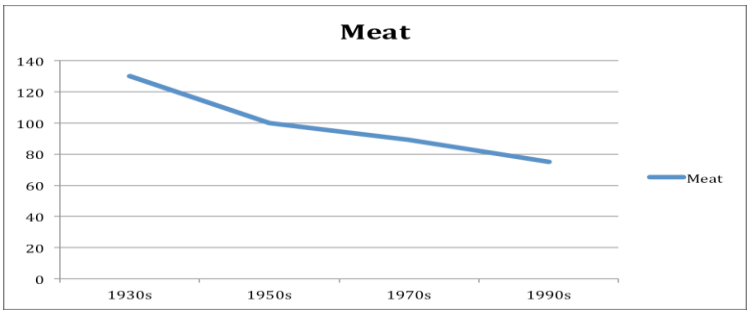
……

……

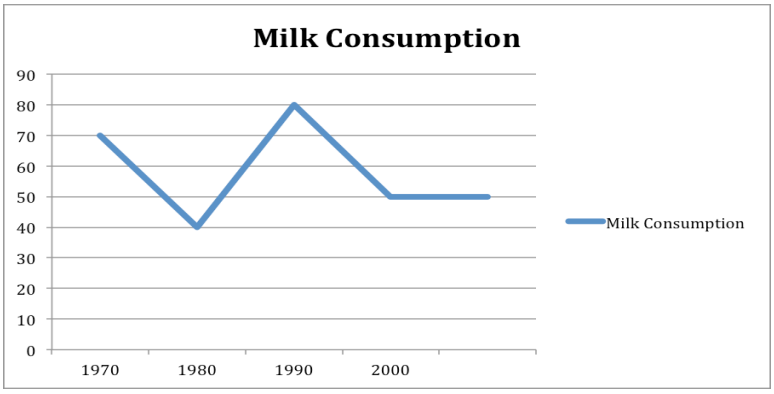


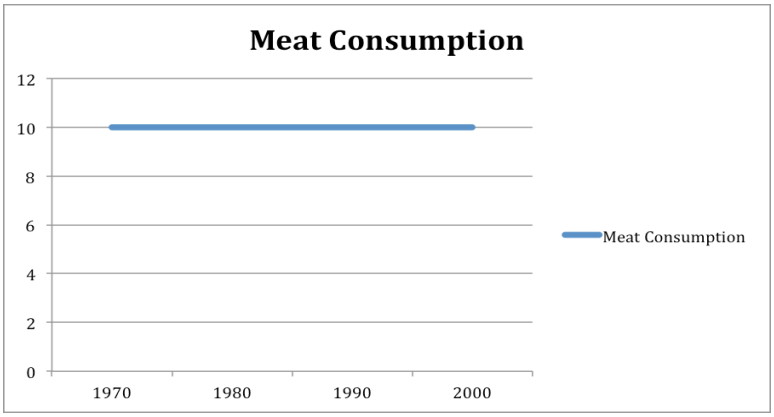
……

……



……


Milk Consumption
90
80
70
60
50
40
30
20
10
0
1970
1980
1990
2000
Milk Consumption


Meat Consumption
12
10
8
6
4
2
0
1970
1980
1990
2000
Meat Consumption

Worksheet 15: Sentence Builders

Sentence builders are typically used to introduce the subject of a sentence and, in some cases, introduce your opinion objectively. NOTE!! **"S"** means a complete sentence, i.e., it includes a subject, a verb, and a complement/an object. **"V"** means a verb in the base form.

1. (A) and (B)/ Both (A) and (B)/ (A) as well as (B)

These phrases are used to present two subjects/objects that have the same function and/or share the same characteristics in one sentence and, thus, reduce repetition in writing. Instead of writing "*Canada is a developed country. The US is a developed country."* You can write

- ***Canada and the US*** are developed countries. OR
- ***Both Canada and the US*** are developed countries. OR
- ***Canada as well as the US*** are developed countries.

Note that you can also write: ***Both of these*** countries are developed.

2. ***The (A) of (B)***

This is one of the most common phrases used to introduce a complex noun phrase as a subject or an object.

- ***The advantages of technology*** outweigh its disadvantages.
- People should not underestimate ***the advantages of technology.***

3. ***There is a/an/not*** OR ***There are a number***

These two sentence builders are used to introduce a subject in a sentence. NOTE!! "***There is"*** is used in the case of a singular subject. "***There are"*** is used in the case of a plural subject.

- ***There is an*** important reason why smoking should be banned in public places.
- ***There are a number*** of reasons why smoking should be banned in public places.

4. ***It has been generally believed/ noted/ asserted that + S***

It is used, **<u>typically in introductions</u>**, when the writer wants to show that an opinion/ a phenomenon is widely accepted by many people. Thus, it helps writers present their opinion objectively without using such phrases as *I think, I believe, in my opinion*, etc.

- ***It has been generally believed that*** learning a second language is challenging for many adults.

5. ***It can be seen/observed that + S***

This clause can be used to introduce your opinion, **typically in conclusions**, with reference/in relation to the information you have already presented in your body paragraphs.

- ***It can be seen that*** advances in technology have both positive and negative effects on humanity at large.

6. ***It is important/necessary to + V***

These phrases are typically used to introduce suggestions which the writer thinks are essential.

- ***It is necessary to*** exercise everyday.
- ***It is important to*** exercise everyday.

7. ***It is difficult to + V***

This phrase is used to refer to things that cannot be accomplished/done easily.

- ***It is difficult to*** jog in winter.

8. ***It is possible***

This phrase is used to introduce a possibility and can be used in two ways:

- It is possible to + V: ***It is possible to*** buy books online.
- It is possible that + S: ***It is possible that*** he is so busy.

9. ***(A) is/are likely to + V***

This phrase is used to hedge and avoid generalization; that is, to acknowledge any possible exceptions for something.

- Using phrases ***is likely to*** alleviate grammatical errors in writing.
- Salt and sugar **are likely to** cause many health problems.

Worksheet 16: Practice

Put the following words in meaningful sentences. Use the clues provided between brackets.

1. Both (A) and (B): (hard work, persistence, success)

…….

2. It has been generally believed that…: (car-share programs, pollution)

…….

3. It is possible to: (drive, 18 years old)

…….

4. There is a/an: (penalty, late submissions)

…….

5. The (A) of (B): (cereal price, increased)

…….

6. It can be observed that (registration numbers, increased dramatically)

…….

7. It is important to (practice, the exam)

…….

8. It is possible that (human actions, global warming)

…….

9. There are a number (using phones, driving)

…….

10. is likely to (listening to songs, listening skills)

…….

11. It is difficult to (independent)

……

Worksheet 17: Cause/Effect & Compare/Contrast

Some phrases/clauses are frequently used to compare and contrast OR to reveal cause and effect in writing. These phrases, for the most part, come in a fixed structure. NOTE!! "**S**" means a complete sentence.

1. ***(A) play(s) an important role in (B)***

It is mainly used in writing to show how one thing (the subject of the sentence) can positively affect something else (the object of the sentence).

- Practice ***plays an important role in*** language learning.

2. ***S + due to the fact that + S***

This is mainly used to show the causal relationship between the two parts of the sentence. **Note!!** This phrase can be used in two different ways.

- John got full marks ***due to the fact that*** he studied hard.
- ***Due to the fact that*** he studied hard, John got full marks.

3. ***As a result of***

This phrase is typically used in writing to show that one thing is the result of something else. **Note!!** This phrase can be used in two different ways.

- She needed support ***as a result of*** her nervousness. OR
- **As a result of** her nervousness, she needed support.

4. ***The reason for (A) is/was (B)*** OR ***(B) is/was the reason for (A)***

This phrase is used to show how one thing (A) can be caused by something else (B)

- ***The reason for*** infection is germs.
- Germs are ***the reason for*** infection.

5. ***On the basis of/Based on the***

These two phrases are used in academic writing to show that one thing has happened because of/as a result of something else.

- The decision was made ***on the basis of*** the current situation.
- The decision was made ***based on the*** current situation.

6. ***On the one hand/On the other hand***

These two phrases are mainly used when you write two contrasting ideas or paragraphs. In other words, the second sentence/paragraph presents information that is opposite to the first sentence/paragraph. **Note!!** You can use ***on the other hand*** without ***on the one hand***. However, you **CANNOT** use ***on the one hand*** without including ***on the other hand***.

- ***On the one hand,*** advances in technology have made the world a small village. ***On the other hand,*** they have many negative effects on humanity at large. ✓
- Advances in technology have made the world a small village. ***On the other hand,*** they have many negative effects on humanity at large. ✓
- ~~***On the one hand,*** advances in technology have made the world a small village. They have many negative effects on humanity at large.~~ ✗

7. ***Not only … but also***

This correlative conjunction is used to show that the head noun has two characteristics. Note!! If you start the sentence with ***not only***, remember to invert (Question structure).

- The car is ***not only*** economical, ***but*** it ***also*** feels good to drive.
- ***Not only*** is the car economical, ***but*** it ***also*** feels good to drive.

Worksheet 18: Mingle Activity

1. Simplicity is THE REASON FOR the popularity of these languages.

2. Group work PLAYS AN IMPORTANT ROLE IN social development.

3. He found the program really difficult DUE TO THE FACT THAT he came from a different educational background.

4. The price of several products increased AS A RESULT OF economic depression.

5. ON THE ONE HAND, I want to buy a big house. ON THE OTHER HAND, I do not have enough money.

6. NOT ONLY did he invent the Internet, BUT he ALSO invented the calculator.

7. He was promoted to manager ON THE BASIS OF his hard work and commitment.

Put the following phrases in meaningful sentences.

THE REASON FOR: ……

……

PLAY AN IMPORTANT ROLE:...…….....

……

DUE TO THE FACT THAT: …….…….....

……

AS A RESULT OF: ……..…….....

……

ON THE ONE HAND/ON THE OTHER HAND: …….……

……

NOT ONLY/BUT ALSO: …….....

……

BASED ON: ……...…….....

……

Worksheet 19: Discourse Organizers

Some phrases are used in academic writing to help writers organize their texts and establish smooth transitions between sentences and/or paragraphs. In addition to text organization, they tell your readers what to expect next.

1. ***In other words,***

This phrase is typically used when writers present already introduced information in different words due to the complexity and/or lack of clarity of the information in the previous sentence. **Note!!** This phrase comes at the beginning of a sentence and is followed by a comma.

- The listening test is really difficult. ***In other words,*** you need to practice a lot in order to pass the test.

2. ***In addition to (A),***

This phrase is used when writers add information. It can be used between sentences or at the beginning of your second body paragraph to connect ideas together and help smooth transition from one sentence/paragraph to another. **Note!!** This phrase comes at the beginning of a sentence and is followed by a **comma.**

- ***In addition to*** lung cancer, smoking is the main reason behind many heart diseases.

3. ***For example, For instance,***

These phrases are mainly used to list examples in a text. You can use these phrases in different ways. **Note!!** If you use these phrases at the beginning of a sentence, add a comma after them.

- You can practice listening in various ways. ***For example,*** you can watch movies or listen to the radio.
- You can practice listening in various ways. ***For instance,*** you can watch movies or listen to the radio.

You can also use these phrases at the end of the second sentence. ***Note!!*** They are preceded by a **comma**.

- You can practice listening in various ways. You can watch movies or listen to the radio***, for example.***
- You can practice listening in various ways. You can watch movies or listen to the radio***, for instance.***

4. ***Such as***

This phrase, like ***for example*** and ***for instance***, is typically used to introduce examples. ***Note!!*** It is followed by nouns **NOT** a sentence.

- I like bright colors ***such as*** red and yellow.

5. ***At the same time,***

This phrase, in addition to its literal meaning, is used figuratively as a synonym for however/nevertheless. **Note!!** This phrase comes at the beginning of a sentence and is followed by a comma.

- We agree with your suggestions. ***At the same time,*** we object to your methods.

6. ***Between (A) and (B)/ From (A) to (B)***

These two phrases can be used to introduce time frames and/or price range. Since these phrases come at the beginning or at the end of a sentence, they help produce various sentence structures.

- ***From*** 1999 ***to*** 2009, John lived in France. Or
- John lived in France ***from*** 1999 ***to*** 2009.
- ***Between*** 1999 ***and*** 2009, John lived in France. Or
- John lived in France ***between*** 1999 ***and*** 2008.

You can also use ***over a period of ….***

- This chart presents the fluctuation in the prices of oil ***over a period of*** ten years.

7. ***At the beginning of…/At the end of****….*

As their literal meaning indicate, these two phrases are used to refer to the beginning or the end of a thing or a time period. They can be used at the beginning of a sentence with a comma or at the end of a sentence without a comma.

- We had to identify the weaknesses of each design ***at the beginning of*** the project. Or
- ***At the beginning of*** the project, we had to identify the weaknesses of each design.
- The price of rice decreased sharply ***at the end of*** 2010. OR
- ***At the end of*** 2010, the price of rice decreased sharply.

8. ***According to the (A)/As shown in (A),***

These phrases are typically used to make reference to the source of information (a chart, an article, or a writer) in a written text. **Note!!** These phrases come at the beginning of a sentence and are followed by a **comma.**

- **According to the graph,** there is a sharp increase in the number of people using the subway between 6:00 and 8:00 am.
- **As shown in the graph,** there is a sharp increase in the number of people using the subway between 6:00 and 8:00 am.

9. ***In conclusion/ In sum/ To sum up/ To conclude/ In summary,***

These phrases are used at the beginning of a concluding statement/paragraph. Thus, they mark the beginning of your conclusion. **Note!!** These phrases come at the beginning of a sentence and are followed by a **comma**.

- ***In conclusion,*** it can be noted that advances in technology have both advantages and disadvantages.

Worksheet 20: Practice

I. Rewrite the following sentences using the transition phrases in brackets.

1. John has a lot of homework to do. He is too busy these days. (*In other words*)

……

2. The change in the atmospheric temperature affects wild animals. Vegetables are adversely affected by global warming. (*In addition to A*)

……

3. The migration of various animals and birds is one of the indicators of increasing temperatures. Foxes and whales are now found in the areas away from their natural habitat. (*For instance*)

……..

4. I like historical movies. I like Brave Heart and The Patriot. (*Such as*)

…….

5. John paid for the damage. He was very angry about the water leak. (*At the same time*)

……

6. It can be seen that global warming has negative effects on both animals and plants. (*In sum*)

……

7. The graph shows how medical practices changed worldwide. (*Add a period of time using from… to…./ 1900-2000*)

……

8. The prices of several diary products increased between 2005 and 2007. (*According to the graph*)

……

9. The number of accounting students is significantly higher than the number of engineering students. (*As shown in the table below*)

……

10. You need to submit a detailed report. (*at the end of this month*)

……

Table C1

Discussion Questions for the Second Set of Instruments

Worksheet #	Discussion Questions
Worksheet 1	1. What is the function of a topic sentence? 2. When reporting graphical information, which elements should be included in the topic sentence? 3. Which phrases did the writer use to a. introduce the graph? b. introduce the changing trends? c. introduce the time range?
Worksheet 2	1. What is the function of a concluding sentence? 2. Which phrases did the writer use to start the concluding sentence? 3. Which phrases did the writer use to refer to the graph?
Worksheet 3	1. What is the difference between the phrases highlighted in yellow and the ones highlighted in green? 2. Which phrases would you use to describe the trends drawn on the board?
Worksheet 5	1. Which phrases can be used to introduce a singular/plural subject in a sentence? 2. Which phrases can be used to avoid repetition in writing? 3. Which phrase can be used in an introduction to introduce your opinion objectively? 4. Which phrase can be used in a conclusion to restate your opinion objectively? 5. Which phrases can be used to a. avoid generalization? b. tell the reader that something is important? c. tell the reader that something is possible? d. construct a complex noun phrase?
Worksheet 7	1. Which phrases are used to show that one thing is caused/the result of the other? 2. Which phrases are used to contrast ideas? 3. Which phrase can be used to show the positive effects of one thing on the other? 4. Which phrase is used to show that the subject has two characteristics?
Worksheet 9	1. Which phrases can be used to list examples? 2. What is the difference between ‘for example’ and ‘such as’? 3. Which phrase can be used to restate the same thing in different words? 4. Which phrases can be used to present a time range? 5. Which phrases can be used to make a reference to already presented information? 6. Which phrases can be used a. to add information? b. to start a conclusion? c. to refer to the beginning/end of a time period?

	d. as a synonym for ‘however’

Note. These questions were used during the training period in the experimental saliency groups

Instruments for the Control Groups

Worksheet 21: Writing a Topic Sentence

In academic writing, a topic sentence is usually the first sentence in a paragraph and is used to introduce the reader to the main points of discussion. A topic sentence should be general enough to include all the points that the writer will discuss in the subsequent sentences. In academic essays, a topic sentence—the first sentence of each body paragraph—has two functions. It introduces the reader to the main idea of the body paragraph; it also links the body paragraph to the thesis statement (the last sentence in the introduction) and to other body paragraphs.

Example:

Pollution is the main reason behind global warming.

In the example above, the topic sentence informs the reader that the entire paragraph will discuss the ways in which pollution increases global warming. Therefore, all the subsequent sentences (i.e., supporting sentences/details) should support the writer's claim that pollution is the major cause for global warming.

Practice

For each of the following prompts, write a topic sentence.

Parents should supervise their children's use of the Internet. Why? Why not?

_____

_____

_____

What is the best way to learn a new language?

_____

_____

_____

What is the main reason behind childhood obesity?

_____

_____

_____

Worksheet 22: Writing a Concluding Sentence

In academic writing, a concluding sentence is usually the last sentence in a paragraph and is used to reiterate the main points of discussion that have already been presented. A concluding sentence may also include a final thought. In academic essays, a concluding sentence, which is typically the last sentence of a body paragraph, has two functions: restating the main points of discussion (i.e., a paraphrase of the topic sentence) and helping smooth transition to the subsequent body paragraphs.

Example:

In conclusion, one of the major causes of global warming is pollution.

In conclusion, since pollution is one of the major causes of global warming, the government should come up with new policies to reduce it.

In the first example, the writer has reiterated his/her major point of discussion which he/she has proposed in the topic sentence (see the example in Worksheet 1). In the second example, the writer restated the main idea of the paragraph and added a final thought (an action to be taken by the government).

Practice

For each of the following prompts, write a concluding sentence.

Parents should supervise their children's use of the Internet. Why? Why not?

_____

_____

_____

What is the best way to learn a new language?

_____

_____

_____

What is the main reason behind childhood obesity?

_____

_____

_____

Worksheet 23: Types of Sentences

Writing is an expressive skill whose main function is to share and communicate ideas with others in written texts. When you write academic texts, you construct dependent and independent clauses to form a variety of sentence types.

Clauses

A clause is a group of words that has a subject and a verb. A clause can be either complete (independent) or incomplete (dependent) in meaning.

Independent Clauses (Complete in Meaning)

An independent clause consists of a subject, a verb and an object or a complement. It is **complete** in meaning and, thus, **can stand alone** as a complete sentence.

Examples:

Global warming is a serious issue.

Learning a second language is not easy.

Note! Each of the above-listed clauses is complete in meaning and, therefore, can stand alone as a complete sentence.

Dependent Clauses (Incomplete in Meaning)

A dependent clause consists of a subject, a verb, and sometimes an object or a complement. It is **incomplete** in meaning and, hence, **cannot stand alone** as a complete sentence.

Examples:

Because global warming is a serious issue…..

Although learning a second language is not easy…..

Note! These two clauses are incomplete in meaning; therefore, each clause needs an independent clause to be complete in meaning.

Sentences

An English sentence is composed of one or more clauses and is complete in meaning. There are three major sentence types in English: simple, compound, and complex.

Remember! Every English sentence must have at least one main verb and must be complete in meaning regardless of its type.

Simple Sentences

A simple sentence consists of **ONE** independent clause. A simple sentence can have one subject and one verb, two subjects and one verb, one subject and two verbs, etc.

Examples:

Jane goes to school everyday. (one subject and one verb)

Jane and Arthur go to school everyday. (two subjects and one verb)

Jane goes to school on weekdays and visits her friends on weekends. (one subject and two verbs)

Jane and Arthur go to school on weekdays and visit their friends on weekends. (two subjects and two verbs)

Note! The above-listed examples are all simple sentences because they consist of **one independent clause**.

Worksheet 24: Practice

I. Underline the independent clause in each of the following sentences.

1. Because of bad management, the company has lost many clients.
2. She is still unemployed although she has been looking for a job for two years.
3. The building collapsed because there were some flaws in the design.
4. While I was watching TV, the telephone rang.
5. I loved physics when I was in grade 10.
6. She has been the manager of this company since she was 35.

II. Decide whether the following strings of words are complete sentences (C) or incomplete sentences (IC).

1. The lady who is sitting next to John_____
2. His professionalism and work ethics are the main reason behind his success_____
3. Spring is my favorite season_____
4. Because he studied really hard _____
5. The car that he bought from his cousin last year _____

III. Form simple sentences using the instructions provided.

1. One subject and one verb

_____

2. One subject and two verbs

_____

3. Three subjects and one verb

_____

4. Two subjects and two verbs

_____

5. One subject, one verb and two objects

_____

Worksheet 25: Types of Sentences (Continued)

In addition to simple sentences, English sentences can be compound or complex.

Compound Sentences

A compound sentence is composed of **TWO or more** independent clauses that are joined using a coordinating conjunction (FAN BOYS), a conjunctive adverb (i.e., however, therefore, moreover, etc.), or a semicolon.

Compound Sentences with Coordinating Conjunctions

A compound sentence can be formed with two or more independent clauses and a coordinating conjunction (i.e., ***and***, ***but***, ***or***, ***so***, ***nor***, ***yet***, ***for***).

Examples:

I love coffee**, so** I drink it every day.

Jack has been studying English for three months**, yet** he cannot spell his name in English.

Remember! Coordinating conjunctions are preceded by a comma.

Compound Sentences with Conjunctive Adverbs

You can also form a compound sentence using a transition signal (i.e., ***moreover***, ***furthermore***, ***however***, ***thus***, ***otherwise***, etc.). These transition signals function as conjunctive adverbs.

Examples:

Rebecca studied really hard**; however,** she failed the test.

Students must arrive on time**; otherwise,** they will be marked absent.

Remember! A conjunctive adverb is used with a semicolon (or a full stop) before it and a comma after it. Do not use a comma before a conjunctive adverb when joining two sentences.

~~*Rebecca studied really hard**, however,** she failed the test.~~

Compound Sentences with a Semicolon

In addition to coordinating conjunctions and conjunctive adverbs, you can use a semicolon between two sentences that are related in meaning to form a compound sentence.

Examples:

John reads poetry**;** his sister reads novels.

The weather is good; we should go for a walk.

Complex Sentences

A complex sentence consists of an independent clause and a dependent clause (an adverb or an adjective clause).

Complex Sentences with Adverb Clauses

A complex sentence consists of an independent clause and a dependent clause that starts with a subordinating conjunction (***when***, ***while***, ***before***, ***after***, ***although***, ***if***, ***because***, ***even though,*** etc.).

Examples:
When you become 18, you can drive a car.
If I were the king, I would change the world.

Note! The dependent clause can be the second part of the sentence. In this case, DO NOT use a comma.
You can drive a car when you become 18.
I would change the world if I were the king.

Complex Sentences with Adjective Clauses
You can also form a complex sentence by using one independent clause and one dependent clause that functions as an adjective and starts with a relative pronoun (***who***, ***which***, or ***that***). The adjective clause must follow the noun it describes.
Examples:
The boy who broke the window was punished by his parents.
William Shakespeare, who was a famous dramatist in the 16th century, wrote *Romeo and Juliet*.
Note! If the adjective clause is essential in defining who the subject is, do not use commas.

Worksheet 26: Practice

I. Decide whether the following sentences are Simple, Compound, or Complex.

1. The house which William bought was very expensive._____

2. John, Beth, and Jack took the bus. _____

3. I took a shower, brushed my teeth, and went to bed at 10:30 pm._____

4. Jack booked a table for 15, but only 3 guests attended his birthday party._____

5. Tom took the wallet he found to the police station, for it was the right thing to do._____

6. When Mike saw the lion, he passed out._____

II. Complete the following sentences to form simple, compound or complex sentences. Add punctuation as needed.

1. We were able to win the game because_____

2. Mary and Jane bought_____

3. After the teacher introduced the topic_____

4. When Tom graduated_____

5. _____ arrived on time.

6. Some of the students were using their phones_____

III. Use the words in parentheses to form compound or complex sentences.

1. (for)_____

2. (after)_____

3. (however)_____

4. (which) _____

5. (or) _____

6. (although)_____

Worksheet 27: English Nouns and Articles

A noun is a word that refers to people, places, or things. There are two types of nouns in English: countable and uncountable.

Countable Nouns

Countable nouns are nouns that refer to things (tables, chairs, etc.) people (women, men, students, etc.) and places (streets, cities, countries) that can be counted using numbers. Countable nouns can be either singular or plural.

Singular Countable Nouns

Singular countable nouns are used to refer to one person (a man, a student, etc.), one thing (a chair, a pen, etc.), or one place (a country). **Remember!** Singular countable nouns ***MUST*** be preceded by an article (***a***, ***an***, ***the***).

- **A/An**: Use a singular countable noun with the indefinite articles ***a*** or ***an*** when the singular noun refers to a general thing (not specific) or when the reader/listener does not know what/who you are referring to.

Examples:

I eat ***an apple*** everyday. (green, yellow, red, etc.)

He buys ***a car*** every year. (we do not know which model, color, etc.)

- ***The***: Use a singular countable noun with the definite article ***the*** when the singular noun refers to a specific thing (the reader/listener knows what/who you are referring to).

Examples:

The apple I ate was very delicious. (the listener/reader knows which apple).

The car Jack bought was very expensive. (the listener/reader knows which car).

- You can also use ***one, this, that, my,*** etc. before a singular noun (e.g. one book). In this case, **do not** use ***a/an*** (*~~a one book~~)

Plural Nouns

Plural nouns are used to refer to several people (men, children, students, etc.), things (pens, chairs, books, etc.), or places (a cities, countries, malls, etc.). Plural nouns can be used with or without the article ***the***.

- Use a plural noun **without** an article (i.e., ***the***) if the noun refers to a general group of people, things or places.

Examples:

Apples are good for your health. (red, green, yellow, etc.)

Students should study hard before exams. (any students at any level)

- Use a plural noun with the article ***the*** if the noun refers to a specific group or when the reader/listener knows who/what you are referring to.

Examples:

The apples I bought from the supermarket were really delicious. (a reference is made to specific apples)

The students who I teach study hard before exams. (a specific group of students)

Uncountable Nouns

Uncountable nouns are nouns that refer to concrete (e.g. water, rice, hair, etc.) or abstract things (happiness, peace, wisdom, etc.) that cannot be counted using numbers.

Uncountable nouns can be used with or without the article ***the***.

- Use an uncountable noun without the article ***the*** when the noun refers to a general thing.

Examples:

Water is important for humans to survive. (any water that people drink)

Hard work is the main reason behind success. (hard work in general)

- Use an uncountable noun with the article ***the*** when the noun refers to a specific thing.

Examples:

The water in this city is drinkable. (the water in a specific city)

The hard work you put into this course was the main reason behind your success. (your hard work)

Remember!!

Using Articles with Nouns		
	General	Specific
Singular Nouns	a/an	The
Plural Nouns	X	The
Uncountable Nouns	X	The

Worksheet 28: Practice

I. Add "*a*," "*an*" or "*the*" to complete the following sentences.

1. _____ laptop I bought is really good.
2. John has _____ meeting at 3:00 pm.
3. There are _____ number of reasons why using mobile phones while driving should be banned.
4. _____ deadline for this assignment is next week.
5. This sentence has _____ subject, _____ verb, and _____ object.

II. The following sentences have errors in the use of articles. Find the errors and correct them.

1. Price of rice has increased dramatically since 2005.

_____

2. Canada and the US are the developed countries.

_____

3. Learning second language is challenging for adults.

_____

4. We have a assignment everyday.

_____

5. The exercise helps people live a healthy life.

_____

III. Read the following sentences and add "*a*," "*an*" or "*the*" where needed. If no article is needed, add *X*.

1. _____ practice plays _____ important role in language learning.
2. John got _____ full marks because he studied hard.
3. Infection is caused by _____ germs.
4. Advances in technology have made _____ world _____ small village.
5. _____ listening test was really difficult.

Worksheet 29: Prefixes and Suffixes

In English, prefixes and suffixes are added to words in order to derive new words. Please note that there is no rule that applies 100% to adding prefixes and suffixes. Thus, you should always check a dictionary if you are not sure.

Prefixes

A prefix is a group of letters (e.g. im-, un-, ir-, re-, mono-, etc.) that is placed at the beginning of a word in order to derive a new word.

- Some prefixes are added to the beginning of a word to derive a new word that is opposite in meaning. Among these prefixes are ***dis-***, ***un-***, ***im-***, ***in-***, ***ir-***, and ***il-***

Examples:

Possible → **Im**possible Important → **Un**important Responsible → **Ir**responsible

Like → **Dis**like Legal → **Il**legal Correct → **In**correct

- Other prefixes are added to the beginning of a word to derive a new word that has a different meaning.

Examples:

Write → **Re**write (***re-*** means again)

Ordinary → **Extra**ordinary (***Extra-*** means beyond)

Test → **Pre**test (***pre-*** means before)

Nourishment → **Mal**nourishment (***Mal-*** means bad)

Suffixes

A suffix is a group of letters that is placed at the end of a word to change its part of speech, that is, noun to adjective, verb to noun, adjective to adverb, etc.

- Add ***–ation***, ***–sion***, ***-ment***, ***-er***, or ***-or*** to derive a noun from a verb.

Examples:

Inform → Inform**ation** Possess → Posses**sion** Govern → Govern**ment**

Teach → Teach**er** Instruct → Instruct**or**

- Add ***–al*** or ***–ful*** to derive adjectives from nouns.

Examples:

Nation → Nation***al*** Beauty → Beauti***ful***

- Add ***–ize*** to adjectives ending in ***-al*** to derive verbs.

Examples:

National → National***ize*** Verbal → Verbal***ize***

- Add ***–ly*** to adjectives in order to derive adverbs.

Examples:

Accurate → accurate**ly** Effective → Effective**ly** Clear → Clear**ly**

- Add ***–ing*** or ***-ed*** to verbs in order to derive present and past participles that function as adjectives.

Examples:

Amaze → Amaz**ed**/Amaz**ing**

Surprise → Surpris**ed**/Surpris**ing**

Worksheet 30: Practice

I. Add prefixes to the following words.

____resistible	____healthy	____convenience
____able	____send	____honest
____regular	____advantage	____cycle
____obey	____lucky	____mature
____necessary	____popular	____order

II. Complete the following table with the correct forms of the words.

Noun	**Verb**	**Adjective**	**Adverb**
	Inform		
Nation			
		Perfect	
			Correctly
	Affect		
Symbol			

Appendix D: Coding Schema, Analyzed Texts, and Elicited Variables

Table D1

A Predefined Categories List

Sentence Builders	Discourse Markers	Collocations
The (A) of (B)	In other words	Significant increase
In the (A) of	In addition to	Dramatic increase
Over a period of	In conclusion	Rapid increase
Between (A) and (B)	In sum	Sharp increase
From (A) to (B)	In summary	Slight increase
At the beginning of	To sum up	Increase dramatically
At the end of	To conclude	Increase significantly
Play$_{(s)}$ an important role	*In short	Significant decrease
(Due to) the fact that	*In a nutshell	Dramatic decrease
As a result of	*In brief	Sharp decrease
The reason for	*All in all	Slight decrease
It can be seen/observed	For example	Decrease significantly
As can be seen	For instance	Decrease slightly
(A) and (B)	Such as	Decrease dramatically
(A) as well as (B)	At the same time	Steady decline
Both (A) and (B)	According to the	Sharp decline
Both of these	As shown in	Significant decline
There is a/an/not	On the one hand	Rapid decline
There are a number (of)	On the other hand	Gradual decline
It has been (adverb)		Dramatic decline
noted/believed/asserted		Remain stable
It is important to		Remain the same
It is necessary		Rise and fall
It is difficult to		Rise and fall
It is possible to		The fluctuation in the price
The (X) shows		Prices fluctuate
The (X) present(s)		
Not only… but also		
Is/are likely to		
On the basis of		
Based on the		

Note. This predefined categories list is based on the target formulaic sequences that were extracted from two sources (i.e., published lists and the COCA) and covered during the training period; *refers to formulaic sequences that were orally introduced during the presentation stage; formulaic sequences that were partially identical to the ones provided in this list were also counted as distinct types

Analyzed Texts

The following written texts were produced by six participants for the six groups at the three stages of data collection and were included to illustrate the coding method that was applied to the textual data. The different typographical cues represent different categorical counts, that is, frequency (formulaic sequences highlighted in both yellow and green); occurrence (formulaic sequences highlighted in green); incorrect formulaic sequences (superscript round brackets with an asterisk and descriptive letters); errors (superscript numbers). Other categorical counts were manually elicited, as explicated in the variables section below. These six participants were randomly selected from the six different groups by choosing the first participant in Tables D2 to D7 who completed all rounds of data collection (i.e., Victoria: upper-intermediate focused instruction group, Mia: advanced focused instruction group, Mike: upper-intermediate saliency group, Nart: advanced saliency group, Alex: upper-intermediate control group, and Charlie: advanced control group).

1. Victoria's Pretest (Task 1: Prompt 3): This graph shows the percentage of homelessness between 1940 and 2000 in four countries. Country B held the [1]most percentage among the four countries during the 60 [2]years period, and its homeless rates kept a steady increase from 12% to 16% [(*st)]between 1940 [3]to 2000. Country A's rates of homelessness declined from 1940 to 2000. In 1940, 9% of country A was homeless people and after a sharp drop, the rate[4] down to 3% in 1960. Then [6]the homeless people [7]took [8]over [9]less [10]and [11]less percentage in country A. The homeless rate of country A gradually decreased to 2% in 2000. The conditions in country C and country D [12]was a little complex. Country D kept a higher rate of homeless people than country C during the whole period. In 1940, the rate in country D was 2% with a sharp increase to 10% in about 1945, and the rate exceeded country A in about 1941. Then, country D's rates[13] up to 13% in 2000 after [(*st.)]a few years'[14] fluctuation.

Victoria's Posttest (Task 1: Prompt 1): The line graph above shows the proportion of international students in [1]Four universities from 1999 to 2005.

In 1999, both university A and University B had 40 percent of students from other countries. Then, the percentage of international students in A increaced steadily to 60 percent [2]In 2005. At the same time, the rate in B fluctuated from 2000 to 2003, and the proportion grew rapidly from 30 to 50 between 2003 and 2005. D and C kept a similar trend from 1999 to 2005. Their percentage stayed at 15 and 13 [(*st)3]during 1999 to 2003. Later, the rate in University D and University C [4]kept a [(*st)5]slightly decline from 2003 to 2005.

Victoria's Delayed Posttest (Task 1: Prompt 2): The corn production in X country was the lowest one among the four crops [1], and [2]also kept a stable trend at around 4 tons from 1995 to 2010. The wheat production was about 5 tons higher than the corn production the whole period between 1995 and 2010. They kept a similar [3]tendency from the lowest production in 1995 to their peak in 2005[4], and had a slight drop from 2005 to 2010. The production of sugar fluctuated from 1995 to 2010. The [5]number dramatically decreased from 25 tons in 1995 to 6 tons in 2010. After that, the production increased to 16 tons in 2004 and kept steady until 2005. Then, the production of sugar declined to 11 tons in 2010.

Victoria's Pretest (Task 2: Prompt 3): The levels of intelligence are various [1]from people. Some people believe [2]It [3]Is decided when we [4&5]was born, while others hold an opposite [6]idea. In my opinion, everyone's intelligence is different when [7]given birth, but it can be influenced in people's later [8]life.

In the early years, [9]the babies may have little consciousness about the outside world. All the things they have an awareness of may come [10]from their mothers' [11]pregnancies. At the [12]baby periods, how long children need to learn walking and speaking may be decided by their original intelligence, so there may be little things which can improve [13]babies intelligence.

However, when babies grow up, they [14]aquire more things, including some skills, knowledge, opinions and so on. All the things they [15]aquire in the later life can stimulate the development of their brains, which means the teenagers may get more intelligent than their childhoods. After they grow up to their adult time, [(*st)]the pace of [16]brain [17]growing may [18] slow down, and some people probably feel more and more stupid as time goes by. This tendency may because [19] the development of people's cells. Also, there are some other aspects that can influence people's intelligence, such as the nutrition they [20]absorbing or some stimulations from [21]outside world.

[(*st)]In [22]a conclusion, intelligence can be improved or influenced. There isn't a role for people's [23]brain to [24]stay [25]still.

Victoria's Posttest (Task 2: Prompt 1): Technology is a double-edged sword. There are definitely some negative effects coming with [1]their advantages as technology is developing so fast.

First, [2]environment is [3]distroyed somehow as a result of [(*st)4]development of technology. People invented cars to make it more [5]convinient to travel around. However, the gas let out by cars [6]caused air pollution which may cause [(*st)]several diseases [7]of [8]human. Petrol also comes from natural petroleum[9]; which is limited. We can continue [10]using [11]car, and the resources from nature continue to be consumed.

Second, some high-tech products may also [12]do [13]to people's health. For instance, people now use computers to do most of the jobs that they used to do by [14]hands. It improved the efficiency of dealing [15] tasks but computers could also hurt people's eyes. Also, staying in front of computers for too much time could also be harmful to people's health because of radiation. Moreover, there are also some weapons [16]can damage people's homeland, such as [(*st)17]atomics and nuclear weapons. All of these weapons [18]are resulted from technology.

All in all, [19]negative effects always come with advantages. We should be aware of them and limit [20]the [21]effect to be small.

Victoria's Delayed Posttest (Task 2: Prompt 2): Every country has its own education system according to [1]their [2]countries' conditions. [3]Educations in China and in Canada [4]are really different in education costs and methods.

In China, the cost of education sometimes [5]depended [6]by the quality of a student. For example, if a person is good enough to go to the top university, his or her tuition may be paid by the government or other sponsors. Moreover, the better a university is, the less its tuition is. Additionally, the education methods in China [7]is [8]a kind of stubborn. Most classes are [9]teached by the [10]teachers [11]words without [(*st)]a [12]reaction of students. We always get lectures from teachers and little interaction is included. Students also have a lot of homework and exams.

On the contrary, conditions are different in Canada. [(*use)]As shown in some materials [13]. The tuition fees in Canada are different in every province. Normally the better universities charge more money [14]from [15]students. Furthermore, Canadian teachers pay more attention to students' [16]participations. They give students more opportunities to express [17] in classes. Comparing with Chinese students, students in Canada have [(*st)]less homework and [18]exams, which give them more free time.

Both education systems are suitable for [(*st)]the condition of [19]two countries. There is not a better one. Whatever education system is the most suitable for a country is the best system.

2. Mia's Pretest (Task 1: Prompt 2): [(*use)]The chart shows [1]that [(*sp)]the trends of country X [2]Production from 1995 to 2010. There are four lines in the chart which are [(*st)]wheat,

sugar and corn [3]. In general, wheat, corn and rice all increased slowly from 1995 to 2010 [4], [5]Only sugar [6]lines decreased rapidly from 1995 to 2001. Then [7]It [(*st)8]quick increased from 2001 to 2005. There are two similar trends [9]grew up all the time, which are wheat and corn. Rice [10]line was the most stable trend (below 5 tons) from 1995 to 2010. The highest quantity was 25 tons of sugar in 1995. Overall, the trend of country X production [11]is growing up.

Mia's Posttest (Task 1: Prompt 1): The line chart shows the percentage of international students in University A, B, C, and D between 1995 and 2005. In general, the trend of international students' percentage in University A and B [1]was going up from 1995 to 2005. [(*st)]The percentage of [2]immigranters in A increased slowly [3]about 55%. The line of University B [4]contained [(*st)5]fall and [6]grow together from 1995 to 2005. [(*st)]Another trend of [7]percentage of international students [(*st)8]were stable. College C and D both declined [9]one or two [10]percentages between 1999 and 2005. Overall, [(*st)]the percentage of [11]immigranters in University A and B [12]raised slowly from 1999 to 2005, and [13]another two [14]line kept stable during [15] [16]20 [17]contury.

Mia's Delayed Posttest (Task 1: Prompt 3): This line chart presents the percentage of homelessness in 4 [1]Countries from 1940 to 2000. The overall trend was that [(*use)]two line [2]charts grew up between 1940 and 2000 [3], they are Country B and Country D. On the other hand, Country C and Country A decreased [4]in 60 years. Country D and Country A changed [5]in some percentage (7%), but they [6]were changed in opposite [7]trends. The percentage of country B increased slowly by almost 3% from 1940 to 2000. The trend of country C was divided into two parts; [8]It increased dramatically in [9] first 30 years and went down between 1970 and 2000. In general, this line chart shows [(*st)]a complex trend of [11] percentage of homelessness in 4 countries.

Mia's Pretest (Task2: Prompt 2): [1]Diversity education is a new trend around the world. [2]Different country has [3]different cultural background, so the education system in each country [4]are different. [5]Compared China and Canada, there are two main differences [6] [7]regulations between China and Canada, which are [(*st)8]discussion and materials.

The most obvious difference is that Canada [9]allow students [10] discuss with [11]partner anytime and [12]interupt the class. If students want to ask some questions, they can ask [13] teacher during the class. However, Chinese [14]student just [15]have questions after class. Also[16] they [17]would get punishment when they [18]discussed in a group in class.

The second [19]differences between China and Canada are [20]material. Chinese students only buy one book for one course[21], they do not need to read other books or materials to study [22]the course. However, [23]Canada schools [24]requires students [25]read widely. They have to find different [(*st)]sources and [26]material to finish one course. Therefore, [27]different country has [28]their own regulation to teach students.

In conclusion, …

Mia's Posttest (Task 2: Prompt 1): Nowadays, with the development of technology, more and more people [1]are relying on [2]the [3]high-tech, such as (*st)[4]cell-phone, [5]computer, and robots. Although technology provides [6]the [7]convenient and [8]the [9]comfortable for [10]human, it still has two negative effects on (*st)our [11]body and mental health if we do not control (*st)[12]time of using [13]of technology, which are (*st)vision loss and slow [14]in [15]reacting.

First of all, (*sp)the percentage of [16]tenagers who get myopia rapidly [17] in recent years [18], (*st)the main reason [19]of this situation is the overuse of technology. (*st)According to [20]the research, (*st)the flickering screens of [21]cell-phone and [22]computer x2 freq. 23&24 were running children's [25]eye. Also [26]myopia can lead to persistent headaches and muscle pain. Therefore, it is easy to see[27] the overuse of technology hurts our eyes.

In addition, if people [28]uses [29]the technology [30]in a long term, they will feel tired and [31]slow [32]at [33]first to react [34] some [35]situation. For example, many teenagers [36]would like to play computer games all night, [37]then they [38]always cannot focus [39]on class because they are tired and their brains [40] still excited [41]in the [42]game's world. Therefore, the overuse of technology has [43]the negative influences on our physical and mental health.

In conclusion, technology helps people to save time and money because it is convenient. However, the overuse of technology has [44]the [45]bag effects on our brains and physical health, such as (*st)vision loss and slow [46]in [47]reacting. Therefore, people should use technology [49]in a [50]certain period and manage their own [51]time table.

Mia's Delayed Posttest (Task 2: Prompt 3): It's generally believed that some people have certain talents when they are born. However, others are not born intelligent. I totally disagree with [1] that some people are born intelligent. Children can be taught to be smart [2]person when they grow up. Parents and self-requirement are two main factors [3]to influence children's intelligence.

First of all, parents' education can be [4]a effective way to influence one's IQ. It all [5]start from parents when they show their attention to their children and what [6]do they do to develop children's intelligence. For example, in [7] [8]chinese traditional story <<Shong Chang Yang>>, Zhang Yang was a smart boy who could create (*st)[9]poetries and songs when he was 5 years old. However, his parents never [10]teach him and [11]refer to send him to school. Then, Chang Yang became a normal person, [12]even he became [13]a foolish when he [14] 18 years old. Therefore, (*st)parents and [15]environment are important factors to decide one's intelligence.

In addition, self-management is [16]also another element to decide what [17]you are (*st)kind of people. If you work hard and try your best to enrich your vision, such as (*st)reading books and [18]travel [19] you will be a [20]knowledageable [21]people. These all [22] a part of acquired disposition. Take Einstein as a good example; he could speak when he was 8

years old. Compared with other children, it was very late. However, he [23]require himself in a serious way and [24]work hard. Therefore, he [25]becomes [26]a [27] famous scientist in the world.

To sum up, I totally disagree with [28] that people are born intelligent. In my opinion, parents and self-requirement are two necessary factors to decide one's intelligence. People are all equal when they are born [29], As a consequence[30] acquired disposition is very important.

3. Mike's Pretest (Task 1: Prompt 2):

The chart shows the changes (*st)in [1&2]number of country [3] production between 1995 and 2010. First of all, in terms of rice, it (*st)[4]remineded the same from 1995 to 2010 [5]as 5 tons. [6]However, corn [7]has increased from 8 tons in 1995 to 20 tons in 2005 [8], then it (*st)[9]reminded the same until 2010. As well [10]as, wheat [11]has increased from 13 tons in 1995 to 24 tons in 2005. In addition, sugar [12]has [13]discreased from 20 tons in 1995 to 6 tons in 2002 [14], then it [15]has increased to 16 tons in 2003. [16]Suger (*st) [17]has [18]remineded the same from 2003 to 2005 [19]as 16 tons [20], then it [21]had decreased to 9 tons between 2005 and 2010. There were big [22]changed in country [23] production over the past 20 years.

Mike's Posttest (Task 1: Prompt 3): The graph shows (*st)[1]the fall and rise in the percentage of homeless people from 1940 to 2000. First of all, Country A (*st)[2]has reminded the same until 1960 [3]as 12% [4], then it (*sp)increased [5]slatly until 2000 [6]as 16%. Country B (*sp)decreased [7]slatly from 9% to 2% in 2000. However, Country C was 2% in 1940 [8], then it (*sp)increased [9]slatly to 6% in 1965 and in the year of 2000 (*sp)decreased [10]slatly to 3%. Also, country D (*sp)increased [11]drimitecly from 6% to 10% in 1940 [12], then it (*sp)decreased [13]slatly to 6% in 1975, but it (*sp)increased [14]slatly [15]until 13% by the year of 2000. (*st)As it can [16] seen, (*sp)the [17]shart is about the changes (*st)in [18] number of homeless people between 1940 and 2000.

Mike's Delayed Posttest (Task: Prompt 1): The line graph shows the changes (*st)in [1] percentage of international [2]students' in four universities from 1999 to 2005. First of all, university A increased from 40% to 60% over [3]sis years. University B also (*st)[4]fell and rose [5]whcich [6]start from 40% and [7]felled to 30% [8], then it [9]ended up to 50%. University D and C (*st)[10]were [11]reminding the same until 2003 [12], then they (*sp)increased [13]slatly. As it can be seen, international [14]students' have gone in many changes in percentage in four universities for (*sp)the term of [15]sis years.

Mike's Pretest (Task 2: Prompt 2): [1]Life's has changed over the past 20 years (*st)[2]as well as education. [3]Eash country has its education system such as Canada and Saudi

Arabia. They both have different education [4]system [5] than [6]the [7]other . Two of the differences are [(*st)]the cost of schools and the way of [8]chosen programs[x3freq.&x2occ.].

[9]Education system in Canada is different than [10]it in Saudi Arabia. First of all, schools in Canada [(*st)] play [11] biggest [12]rol in [(*use)13]terms of economy. Students have to pay their fees in order to start school. However, this is different in Saudi Arabia because schools are for free. In addition, [14]Students who attend [15]to [16]start university get 350 as a motivation per month for 4 years.

Students have [(*sp)]a dream of their [17]carer from [18]the [19]child. However, they might change their dream if they are Saudis. [20]University in Saudi Arabia require a [21]spisific [22]grade in order to [23]choose a [24]spisific program. Students have to take [(*st)25]first year of university as preparation for their program. [26]Therefore, in Canada students can choose, whatever they like to study.

[(*st)]The cost of schools and the way of [27]chosen programs [x3 freq.] are different in Canada and Saudi Arabia. [28]Education system is essential in order to improve the quality of education.

Mike's Posttest (Task 2: Prompt 3): Each person has [(*sp)]a different [1]pearsonlity as well as a different [2]talients. Most people think that people are born intelligent, and they are the best. [3]Actully, people are born equal, but there are people who work very hard in order to be what they want or what they [4]dreemed to be. [(*st)5]Practicing and [6]hard working enable people to be intelligent.

[7]Practicing is one way to be expert in [8] [9]majer. For example, not every one can play soccer, but by [10]practicing, they will be experts. In other words, [(*st)]keeping yourself [11]focusing [12]in one [13]majer and [14]pricticing a lot more [15] you one of the [16]intelligents people.

Hard workers are [17]one [18]of the people who are intelligent because they are successful in their [19]works or studies. If many people just put a small [20]affert to work hard, they [21]could [22]have considered [23]themselF as [24]an intelligent people. However, many people are not hard [25]worker.

[(*st)26]Practicing and [27]hard working are ways that [28]makes people [29]be successful. I [30]stronglly believe that people [31]do not born intelligent, but they acquire [32]it. [33]People [34]works [35]are not the same which [36]it [37]lead them to success.

Mike: Delayed Posttest (Task 2: Prompt 1): With[(*st)]the [1]wild-spread of technology [2] many people do not use their mental abilities [3]effectfully. Many people count on their phones in order to remind them [4]for their [5]works or meetings. However, that can affect [6]negatively their physical and mental health in two [7]difference ways such as [(*sp)]their behavior and their [8]momary.

In terms of behavior, people tend to be isolated from each other. Technology is one way that can keep people away from reality, so they do not act or deal with other people

outside [9]socil [10]midah because they have not seen the real world or [11]life it. [(*st)12]On other words, some people know about the world from their homes only[13] which makes them [14]afried to see or deal with the [15]out side world. In addition, people who are addicted to technology might not be [16]awear about their [17]family or take care of them. Also, they may not be friendly [18], if everyone [19]try to keep them away of technology.

[20]However, technology [21]affect [22]negatively [23]people [24]momary. For example, most people count on their phone note; [(*use)25]as a result [26] to remind them because they do not want to use their [27]functions. That [28] a way to [29]makes their [30]momary [31]turned [32]to [33]be poor or weak and with [34]times they lose a lot of their abilities such as [35]remimber names or numbers.

To sum up, people can have such [36]an issues in their behavior or their [37]momary [38]by the overuse of technology. I highly [39]recommned to be careful [40]in the way that people use technology to protect their health. Technology has [41]to sides one negative and one positive.

4. Nart's Pretest (Task 1: Prompt 3): [1]The chart demonstrates the percentage of homeless people in four different countries. Overall, the [2] countries [3]have a constant [4]rend that did not change starting from 1976 to 2000.
Furthermore[5] country A had an increase in 1960, while country B had a dramatic decrease in 1943. Country C had a decrease in most of its period until 1985 [6]where the percentage did not vary until 2000. Country D had a constant increase [(*st)7]between 1940 - 1965.

Nart: Posttest (Task 1: Prompt 2): [1]Overall, the production of corn, wheat and rice[x2 freq &x3occ.] increased gradually [(*st)2]between 1995-2010.
[4]The production of sugar decreased dramatically [(*st)3]between 1995-2000.
The production of corn and sugar[x2 freq. x2occ.] was almost the same in 2003.
[5]In contrast[6] [(*st)7]there is a significant difference between the amount of wheat in 1995 [8]in [9]comparison [10]with the amount of rice in 2010.
[11]To conclude, [(*sp)12]Rice, wheat and corn production increased [(*st)13]between 1995-2000, while sugar fluctuated.

Nart's Delayed Posttest (Task 1: Prompt 1): This line graph demonstrates the percentage of international students in 4 universities. Overall, there was a constant increase in the percentage of international students [(*st)1]between 1999-2002.

Between 2000 and 2002, the [2]Line fluctuated in university B. However, it [(*st)3]stayed steady in university C.

(*st)4Both universities 5percentage 6of C and D rice x2 freq. &x3occ. decreased between 2004 and 2005. 7on 8the 9contarary 10 university A, 11B 12had 13their international students' percentage increased after 2003.

In conclusion, the international 14students percentage increased in A and B universities. (*st)15Starting in 1999 and ending in 2005. On the other hand, it decreased in both of universityC and D x2freq. as a final result.

Nart's Pretest (Task 2: Prompt 3): There are many arguments about whether intelligence comes by nature or not. However, other factors may 1overweigh the genetic 2one. (*use)3Such as4, (*st)5the person's background, parenting as well as desire.

Scientifically speaking, people are born with different abilities. (*use)6For example, the strength of their 7memory 8. It is achieved by random selection which is the transfer of genetics from both the father and the mother. In contrast, 9It has been proven that these abilities will not function as 10good as they should do if people do not use them, so saying that someone is smart by nature is 11nor accurately right 12Because smart people work on themselves in order to improve 13 and get to the place where they are. For example, people with a good short-term memory usually practice improving their memory by (*st)writing, reading as well as processing14 to learn.

A person's background has a great impact on his IQ levels. 15IF we 16took a person who studies 17calculas and 18enrolled him in an advanced art course, he 19would probably get a lower result than his 20artistic peers because his knowledge 21was not 22revolving around art.

Parenting is 23a 24major key because the path that 25the child is directed toward will affect the way he thinks in the long run 26. So if a child is given several brainy games from his childhood such as chess and video games27 he will grow up having the ability to solve problems and think out of the box.

(*use)28In conclusion. 29It is difficult to force the 30human's mind to think about something if (*st)there 31was no desire created. I think that (*st)both the genetic factor as well as desire count.

Nart's Posttest (Task 2: Prompt 2): Knowledge plays a major role in the success of nations. It all boils down to how knowledge is acquired, transfered and taught. Every country has it own rules in terms of the education system. Some countries have developed their educational system, while others still follow previous ones. Canada and Syria share many similarities. However, they differ greatly.

Both Syria and canada have primary and secondary education followed by post-secondary education. After that students (*st)1more likely 2 continue their studies and get a (*sp)master's degree and for more specialization a 3PHD degree.

Public universities in Syria are free.4 In which students do not have to pay tuition fees, whereas public universities in Canada have tuition fees which can be waived if a student has a high GPA or gets an internship.

Nart's Delayed Posttest (Task 2: Prompt 1): Technology has improved 1human's lives amazingly. From performing daily activities to complicated tasks, it is always mandatory. One 2can not live without it anymore, but the problem is when people use it more than how they are supposed to do. That is where its disadvantages 3overdo the advantages.

The revolution of technology resulted in the creation of more suitable devices that meet costumers' satisfaction. This innovation brings about a constant relationship between the users and the device itself; therefore, people become attached to them. Some people usually use their mobile phones while walking or driving. Many accidents have had taken place as a result of that,4 and led to death.

Technology (*use)5plays a major role in distraction. The more it is used, the more distraction it creates. This affects the mental status negatively6 which results in an incapability, disorganization and difficulty in one's life.

Looking at a digital screen for a long time 7ruin the eyes. Most addicts suffer from eye problems…

5. Alex's Pretest (Task 1: Prompt 1):

(*st)The 1bar graph shows how many international 2student 3studying in four different 4university from 1999 to 2005.

5The 6A university had (*st)7noticable 8increased from 1999 to 2002, but it 9stopped 10 one year,11 then it 12completed growing until 2005,13 14with 60 15persent.

16The 17B university (*st)18seems not 19stabile from the year 1999 to 2005. In 2000, there were just (*st)20 20persent of 21student,22 then 23it had gone up to 35 24persentage,25 then again 26&27drop off in 2002 and, lastly, it reached 28to 50 29persentage.

Alex's Posttest (Task 1: Prompt 3):

The chart shows (*st)the 1persentage of 2homless 3 4on four different countries from 1940 to 2000.

Country A,[5] had almost [(*st)]ten percent of [6]homless [7] [8]on 1940. After that [9]it went down [10]dobble between 1940 and 1960. Then it decreased five [11]time more till 1960 and [12]continued to 2 percent [13]on 2000.
However, country D,[14] had [(*st)]6 [15]persente of [16]homless [17] [18]on 1940. Then, [19(*sp)] was not [20]stabile [(*st)]between 1940 [21]till 1975. After that[22] it [(*sp)]went up [23]dramitically to reach [(*st)]14 [24]persent of [25]homless [26] [27]on 2000.
In brief, these [28]was [(*st)]the [29]presentage of homeless [30] [31]between four [32]country from 1940 to 2000.

Alex's Delayed Posttest (Task 1: Prompt 2): This chart shows…

Alex's Pretest (Task 2: Prompt 1): Nowadays, technology has become [(*st)]one part of our [1]life, and people use it [2]in everywhere. Even though the new technology is very helpful in our [3]life, it [4]still can affect us in bad ways. People who [5]are using [6]a technology more than [7]the [8]normal [9]way, they [10]would have health [11]problem and lose their social skills.

First of all, if people use [12]the technology more than the limit, they would [13]heart themselves without [14]they [15]know. For example, [16]the [17]teenge use their [18]cellphone before they go to bed and use [19]it until midnight. If they [20]used [21]it [22]lik that, they [23]would damage their [24]eys, and they might get [25]eys [26]shortage. Also, they [27]would not get [28]a [29]full sleep, and they will be so tired in class. In fact, [30]the human [31]being need 6 to 7 hours [32]to sleep.

Second, people will lose their social skills if they spend a lot of time [33]in it. For example, some people [34]use all their shopping online [35] that [36]it is okay when they use it from time to time. If they use it very much, it [37]would affect their [38]personality. To be clear, they [39]would not know how to communicate with people [40]and learn from others how to buy [41]and how to talk [42]and to deal with [(*st)]strangers [43]people and so on. They [44]would become more lazy and just eat and sleep.

To sum up, these are the [45] most [46]effect [47] can [48]heart people and it seems people [49] not aware [50]from that. I wish people can change [(*st)]their [51]using of technology in good [52]way.

Alex's Posttest (Task 2: Prompt 3): There are some people [1] [2]belive people [3] born [4]with intelligent while others do not [5]belive in that. It seems like a [6]conversonal question to many people. For me, I [7]belive that all people are [8]simillar to each other, but every person has [9] own skills to do things. For example, [10]Parac Obama is a normal person who can [11]achmboilishhis [12] goal and [13]being [(*sp)]the [14]prisedent of [15]the U.S.A.

Being intelligent is sometimes great because those [16] do not need to learn new skills. Some children [17] born with courage[18] so they do not feel anxious [19]of [(*st)]standing in front of [20] class and giving [21] presentation, while others [22]are learning these skills [23] many years.

However, being [24]not born intelligent is not a big deal because they can [25]learn them or study [26]them such as Nelson Mandela.

In brief, I think [27]every one [28] the same and [29]every one has [(*st)30]the [31]power skill and [32]the [33]least skills.

Alex's Delayed Posttest (Task 2: Prompt 2): Every country in the world has [1]their own way to deal with education. Some countries invest [2]on people to be educated in order to lead the country to growth. [3]However, there are many differences [4]in [5]education [6]in Saudi Arabia and Canada.

There are two [7]reasons that the learning [8]experence in Saudi Arabia and Canada is [5]diffrent. First of all, the education in Saudi Arabia is free. [9] Saudi [10]Arabia government wants everyone to be educated and help the country in [11] future. In addition, [12]universities' students [13] not need to pay their tuition fees. The second difference is that education in Saudi Arabia depends on [14]memorazation. If people [15&16]momerised [17]the book, they [18]would usually pass the exam. I believe that [19]memorazation is not [20]the key to success in the real world.
[21]However, [22]learning [23]experence in Canada is expensive,[24] and useful. After [25]students graduated from high school, they have a choice [26]that whether they go to university or not. [(*st)]The [27]prices of taking [28&29]cource at university are [30]sometime [31]student [32]cannot [33]afford [34]it. If they [35]wanted to study at university, they probably [36]would work a part job. The learning style is very good in Canada. [37]They [38]the teach children how to think not to [39]memorise. If children [40]learened the good thinking [41]ability, they [42]would not have a problem in their future.

To sum up, [43]the education is different [44]from Canada and Saudi Arabia. I think that [45]on every country has [(*sp)]a positive and [46]negitive [47]way [(*st)]of education[x2 freq].

6. Charlie's Pretest (Task 1: Prompt 2):[(*st)]As [1]it shown [2]on the graph, the production of rice, wheat, sugar and corn[x2 freq. x3 types] [3]has changed [(*st)4]during 1995 to 2010 in country X. Generally, the [5]productions of wheat and corn[x2 freq.x2types] [6]has increased [7]about 5 tons in 1995-2005, and then [(*st)8]tend to be steady in the following 5 years [(*st)9]on nearly 23.5 tons [10] wheat each year and nearly 19.5 tons [11] corn each year. The production of rice [12]has not changed much from 1995 to 2010 and [13&14]keep [(*st)15]in the amount of 4 tons every year. However, sugar production [(*st)16]has [17]been changed dramatically which

[18]drop from 25 tons in 1995 to 6 tons in 2002. Then it increased to 16 tons in 2005 and finally decreased to about 10 tons per year in the next 5 years. All in all, the production of wheat and corn[x2 freq.] [19]has increased; the production of rice [20]hasn't change much, the production of sugar [21]has decreased, increased and decreased [(*st)22]during 1995 to 2010.

Charlie's Posttest (Task 1: Prompt 3): [(*st)]As it can be seen [1]on the graph, the percentages of homelessness of 4 [2]different countries [3]have changing trends [(*st)4]during 1940 to 2000.

As to country D, homelessness [(*st)5]has dramatically [6]raised from 6% to 10% and then dropped to 7% in [7]1950s. After that, it kept increasing from 6% since [8]1970s. [9]However, the percentage of homeless population [(*st)10]was generally increased [(*st)11]in the range of 12% to 16% in country B. In contrast, homelessness in country A [12]was decreased from 9% to 2% [13]in [14]country [15]A. In terms of country C, the percentage of homelessness [16]was increased from 2% to 6% during 1940s-1970s and then decreased to 3% in the following years.

Charlie's Delayed Posttest (Task 1: Prompt 1):[(*st)]As it is shown [1]on the [2]above chart, the changes of the percentage of international students from 1999 to 2005 are different [3]among university A, B, C and D.

[(*st)4]International students' percentage of university A [(*st)5]has steady increased from 40% in 1999 to 60% in 2005. However, the portion of international students in university B [(*st)6]has fluctuated greatly [(*st)7]during [8]late 1999 to 2003 in the range of 40% -20%, following an uptrend from 20% to 50%. [(*st)]The changes of [9]internationals students' percentage in university C and university D [(*st)10]are rather steady. In university D, the percentage of international students [11]has generally [12]kept [13]in 17% from 1999 to 2003,[14] and [(*st)15]has slightly dropped to 12% [16]since 2003.[17] While in 1999 – 2003, the portion of international students in university C kept around 12%,[18] then [19]has decreased to about 5%.

All in all, [(*st)20]in 1999 to 2005, [(*st)]the [21]occupy of international students in the 4 [22]university [13]has changed variously.

Charlie's Pretest (Task 2: Prompt 2):

There are many differences [1]on the education system between China and Canada.

As far as I know, Chinese high schools [2]pay [(*st)]less [3]amount of time to physical education than Canada's high schools. In China, [(*use)4]there is a huge population and limited great universities. Therefore, students in China [5]prepared themselves to get to the top universities when they are very young. The schools in China [6]are also help them to prepare [7] the University Entrance Examination. Usually, schools [8]sacrifice the time on P.E. courses and arrange more "useful" courses like math, [9]chinese, biology, etc. However, in Canada, high schools pay more attention [10]on [(*st)11]health of their students[12] so they put more time on sports.

What is more, the evaluation system of Chinese schools [13]are more [14]emphasis on the final examination scores rather than daily scores [15] which is different from Canadian academic evaluation systems,[16] In high school [17]stage, Chinese schools teach knowledge which is harder than what [18]be taught in Canada. Therefore, many errors or mistakes in [19]chinese students' [20]homeworks is a [21]common situation. Since the final score [21]matter [22] most in their evaluation, students can make mistakes in their [23]homeworks [24] which will not affect their overall scores. But here in Canada every score [25]is mattered which makes people [26]don't dare [27]to make mistakes.

Charlie's Posttest (Task 2: Prompt 3):

Born to [1]be

Every individual has unique genes [2]which [3]inheirated directly from their parents. Those genes contain some biological [4]informations which decide [(*st)]peoples' appearances, genders, characteristics, and [5]intelligents. [6]Intelligents can be [7]specified [8]to the ability to remember [9]. It is a fact that some people are born smarter than others.

To [10]identify whether a person is more intelligent than others, people should start to observe his/her behavior from the moment he/she [11]was born. If people use the same method to teach two toddlers a word, there is a great chance that baby A can remember the word after the first time he [12]was taught while baby B needs to be taught over and over [13] to remember it. As we all know, [(*st)]the ability [14]of memorize things can be trained. However, we can assure that no one will train newborn babies to remember things. Therefore, [(*st)]the difference [15]of toddler A and toddler B[x2freq&x2occ.] when they remember things is [16]born [17]naturally.

Sometimes, [(*st)18]the shortage of [19]intelligent can be made up through hard working. However, [20]it cannot deny the fact that people are born to be intelligent differently.

Charlie's Delayed Posttest (Task 2: Prompt 1):

Nowadays, people's lives [1]are [2]more rely on [3]the advanced technology than [4]we used to be. After Steve Jobs introduced iPhone to the general public, smart phones [5]has [6]became an [(*st)]important part [7]of [8]part of [9]human's daily lives. Despite the [10]convinience that [11]the smart phones have brought to us, they also have [12]left many negative impacts.

Because of the varieties of apps on [13]the smart phone, people [14]become more and more lazy. [15]Thank to [16]the creative programmers[17] now people [18]almost have apps for everything. In today's society, people can order [19]foods by using apps, buy clothes by using apps, read books by using apps, or even book a taxi by using apps. The well-developed app business provides the majority of people [20] a [21]convinient life-style, but also brings a severe problem. Since now people can do almost everything at home [22] clicking [23] button, they [24]are extremely [(*st)25]lack [26]of [27]exercises. People can survive by just [(*st)]staying [28] their home and using smart phones for several years. This kind of life style would [29]distroy their health.

The development of social networking apps [30]"help" people to live alone. In the western society, almost every individual has a [31]facebook account. People can make new friends and get in touch with the outside world by refreshing their [32]facebook. However, this is not like face to face communication [33] which people can observe [(*st)]every detailed [34]reactions of a person and respond according to it. The [35]informations on [36]the [37]facebook do not [38]need people to respond in verbal language. [39]After time goes by, people may feel uncomfortable to communicate with others face to face. This lack of social interaction may lead to some psychological diseases.

The advanced technologies have improved people's lives in many ways. But people should not rely on them too much due to the negative effects that [40]have mentioned above.

List of Variables

The data quantification stage resulted in creating 102 variables that represented different numerical and categorical aspects:

- *Frequency* (×6): These 6 continuous variables represented the total number of formulaic sequences used by the 67 EAP students in their summaries and essays at the three stages of data collection, including the repeated ones.
- *Occurrence* (×6): The six continuous variables (3 for task 1 and 3 for task 2) labeled *Occurrence* included the raw scores for the distinct types of formulaic sequences implemented by the 67 participants in their summaries and essay responses at the three production stages; that is, each distinct type, together with each formulaic sequence that the participants utilized creatively, was counted once, and all the repeated formulaic sequences were excluded in these variables.
- *Correct* (×6): These continuous variables were manually derived from the Frequency variables and contained the total number of correctly constructed formulaic sequences at the three stages of data collection (both task 1 and task 2).
- *Wrong* (×6): These 6 variables (3 for task 1 and 3 for task 2) were also manually extracted from the Frequency variables, but, unlike the former, they included the total

number of the incorrect formulaic sequences at the three stages of data collection. These six variables were further subdivided into three types of errors, namely usage, structure, and spelling.

- Usage (×6): The six continuous variables were manually derived from the variables labeled Wrong and comprised the raw scores for the incorrect formulaic sequences whose errors were outside their boundaries. In other words, these variables presented the raw scores for the formulaic sequences that were correctly constructed but incorrectly used.
- Structure (×6): These continuous variables (3 for task 1 and 3 for task 2) were also derived from the variables labeled Wrong; they comprised the raw scores for the formulaic sequences that had errors within their boundaries other than spelling (e.g. missing words, tenses, articles, subject-verb agreement, word form, etc.).
- Spelling (×6): These six variables constituted the last error subcategory and represented the raw scores for the formulaic sequences that were misspelled by the 67 EAP students at the three stages of data collection (both task 1 and task 2).
- Errors (×6): The variables labeled Errors consisted of the raw scores for the total number of errors and mistakes (i.e., tense, word form, spelling, punctuation, lexical choice, etc.) made by the 67 EAP students in both task 1 and task 2 at the three production stages.
- Words (×6): These six continuous variables (3 for task 1 and 3 for task 2) represented the total number of words generated by the 67 students at each data collection stage.
- Correct% (×6): These continuous variables were manually derived from the Frequency and Correct variables by dividing the number of the correct formulaic

sequences by the total number of formulaic sequences and, hence, included the percentage of the correct formulaic sequences utilized in task 1 and task 2 at the three production stages. The rationale behind the manual derivation of these variables is that the raw scores for the correct formulaic sequences cannot independently reflect the participants' performance at each stage of data collection. For example, 2, 5, and 7 correctly used formulaic sequences in the pretest, posttest, and delayed posttest respectively do not necessarily mean that a participant could successfully use more correct formulaic sequences at the posttest and the delayed posttest stages. In other words, these numbers need to be viewed with reference to the total number of formulaic sequences because a participant may only use 2 formulaic sequences in the pretest which means that he/she has 100% correct formulaic sequences; the same participant may use 20 formulaic sequences in the posttest and the delayed posttest, only 5 of which are correct in the posttest and 7 of which are correct in the delayed posttest; this means the percentage of the correct formulaic sequences dropped from 100% in the pretest to 25% and 35% in the posttest and the delayed posttest, respectively.

- Errors% (×6): These continuous variables were generated based on the Errors and Words variables by manually dividing the total number of errors made in each text by the number of words generated. Similar to the rationale provided for the variables labeled Correct%, error counts in a writing sample cannot independently reflect a participant's performance because making 20 errors in a 100-word text is not the same as making 20 errors in a 200-word text.
- Evaluation (×6): These six continuous variables (3 for task 1 and 3 for task 2)

comprised the EAP judge's evaluation of summaries and essays that the 67 EAP students produced at the three production stages.

In addition to the 72 continuous variables, 18 continuous variables were constructed using the *Compute Variable* function on SPSS by combining the pretest, posttest, and delayed posttest values in order to assess the distribution of scores in each major category. Moreover, 10 continuous variables were developed using the *Compute Variable* function by calculating the average scores for the frequency and occurrence of the target formulaic sequences, the number of words, the evaluation of the EAP judge, and the total number of errors prior to computing *Pearson correlation coefficients*.

Two discrete variables were also created to split the data at the data analysis stage by group and proficiency level. The first discrete variable had three levels that indicated the three different types of pedagogical interventions—that is, 1 for the experimental focused instruction groups, 2 for the experimental saliency groups, and 3 for the control groups. The second discrete variable was created to split the data based on the proficiency level of the 67 EAP students. It was composed of two levels of measurement (i.e., 1 for the upper-intermediate proficiency level & 2 for the advanced proficiency).

Raw Scores for Task 1 and Task 2

Table D2

Focused Instruction Groups: Raw Scores (Task 1)

	Frequency			Occurrence			Correct			Wrong		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Victoria	16	19	12	10	11	8	13	17	12	3	2	0
2. John	14	10	9	11	9	7	7	4	6	7	6	3
3. Zoey	16	16	19	11	13	15	10	11	9	6	5	10
4. Amy	14	17	10	9	9	7	9	11	6	5	6	4
5. Paul	13	10	5	8	10	5	8	4	0	5	6	5
6. Bonapeite	16	17	26	7	13	14	9	14	14	7	3	12
7. Brian	15	19	21	8	11	12	14	15	16	1	4	5

8. George	13	9	7	6	5	5	8	4	3	5	5	4
9. Sara	12	25	11	9	18	9	10	23	8	2	2	3
10. Eva	9	11	6	8	7	6	6	6	4	3	5	2
11. Henry	17	13	13	10	13	11	8	8	10	9	5	3
12. Jasmine	7	18	9	6	12	8	4	12	5	3	6	4
13. Curry	19	16	4	12	12	3	3	5	2	16 9	11	2
14. Elias	18.67	18	27	9.89	9	15	12.83	15	20	5.83	3	7
15. Maram	11	15	16.72	7	11	10.31	7	11	13.44	4	4	3.28
16. Mia	17	20	17	9	11	10	13	15	15	4	5	2
17. Maya	20	22	29	11	13	18	4	14	17	16 14	8	12
18. Dana	15	14	18	12	11	10	10	5	8	5	9	10
19. Silva	12	22	20	9	11	10	11	22	17	1	0	3
20. Amanda	16	22	15	9	14	13	9	18	14	7	4	1
21. Ana	12	27	21	8	16	12	9	19	20	3	8	1
22. Mary	8	18	25	5	10	16	3	10	16	5	8	9
23. Kareem	17	20	21	11	15	16	8	16	16	9	4	5
24. Tom	13	22	30	8	14	13	11	10	22	2	12	8
25. Leo	27	26	32	10	12	14	13	24	28	14	2	4
26. Fill	14	11	16.19	8	7	8.72	9	8	12.92	5	3	3.28
27. Cherry	5	14	13.03	5	9	8.19	1	7	8.17	4	7	4.86
28. Yin	11	11	17	7	8	9	10	8	15	1	3	2
29. Fares Althaqel	10	16	25	7	9	13	8	13	22	2	3	3
30. Mary (2)	10	5	15	6	5	9	8	5	9	2	0	6
31. Jie	25	25	22	13	16	10	22 13	19	20	3	6	2
32. Joker	16	17	22	11	12	11	8	13	20	8	4	2

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D2

Focused Instruction Groups: Raw Scores (Task 1) Continued

	Use			Structure			Spelling		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Victoria	1	0	0	2	2	0	0	0	0
2. John	0	0	0	5	5	2	2	1	1
3. Zoey	1	2	1	5	2	9	0	1	0
4. Amy	1	1	0	3	5	4	1	0	0
5. Paul	0	1	0	3	5	4	2	0	1
6. Bonapeite	0	0	0	7	3	12	0	0	0
7. Brian	0	0	0	1	4	4	0	0	1
8. George	0	0	0	4	5	3	1	0	1
9. Sara	0	0	0	2	2	2	0	0	1

10. Eva	0	0	0	3	3	1	0	2	1
11. Henry	0	0	0	9	5	3	0	0	0
12. Jasmine	0	1	0	3	4	3	0	1	1
13. Curry	2	0	0	14	11	2	0	0	0
14. Elias	.08	0	0	6.06	3	7	-.31	0	0
15. Maram	0	0	-.17	2	3	1.58	2	1	1.86
16. Mia	1	0	1	2	5	1	1	0	0
17. Maya	1	0	1	13	6	7	2	2	4
18. Dana	1	2	1	4	7	9	0	0	0
19. Silva	1	0	0	0	0	2	0	0	1
20. Amanda	0	0	0	7	3	1	0	1	0
21. Ana	0	0	0	3	7	0	0	1	1
22. Mary	1	0	0	2	4	3	2	4	6
23. Kareem	0	0	0	7	4	4	2	0	1
24. Tom	0	1	0	2	9	3	0	2	5
25. Leo	0	0	1	14	1	2	0	1	1
26. Fill	0	0	-.17	2	1	.53	3	2	2.92
27. Cherry	1	1	.89	3	5	3.17	0	1	.81
28. Yin	0	0	0	1	2	1	0	1	1
29. Fares Althaqel	0	0	0	1	2	3	1	1	0
30. Mary (2)	0	0	0	2	0	6	0	0	0
31. Jie	0	3	0	3	3	2	0	0	0
32. Joker	1	0	0	7	4	2	0	0	0

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow

Table D2

Focused Instruction Groups: Raw Scores (Task 1) Continued

	Errors			Words			Correct %		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Victoria	14	5	5	170	113	121	81.25	89.47	100
2. John	24	15	8	133	92	64	50	40	66.66
3. Zoey	23	15	23	111	118	137	62.50	68.75	47.36
4. Amy	25	32	21	142	149	86	64.28	64.70	60
5. Paul	35	27	30	188	130	109	61.53	40	0
6. Bonapeite	16	8	27	128	104	162	56.25	82.35	53.84
7. Brian	9	8	10	126	125	120	93.33	78.94	76.19
8. George	44	32	35	159	149	139	61.53	44.44	42.85
9. Sara	5	3	5	74	98	63	83.33	92	72.72
10. Eva	13	6	5	65	62	30	66.66	54.54	66.66
11. Henry	30	22	21	162	116	112	47.05	61.53	76.92
12. Jasmine	40	25	20	133	123	83	57.14	66.66	55.55
13. Curry	80	34	32	203	114	83	15.78	31.25	50
	44								

14. Elias	24.25	12	23	142.89	113	162	68.72	83.33	74.07
15. Maram	14	12	8.75	122	132	130.14	63.63	73.33	80.38
16. Mia	11	17	11	106	112	120	76.47	75	88.25
17. Maya	55 31	24	22	149	147	153	20	63.63	58.62
18. Dana	31	26	24	99	94	110	66.66	35.71	44.44
19. Silva	8	9	12	118	143	121	91.66	100	85
20. Amanda	12	11	6	111	134	99	56.25	81.81	93.33
21. Ana	15	13	5	126	117	116	75	70.37	95.23
22. Mary	16	21	16	76	115	153	37.50	55.55	64
23. Kareem	16	10	11	112	120	139	47.05	80	76.19
24. Tom	17	17	16	126	116	152	84.61	45.45	73.33
25. Leo	26	10	13	148	152	145	48.14	92.30	87.50
26. Fill	18	9	9.28	102	87	95.83	64.28	72.72	79.80
27. Cherry	22	23	18.78	110	124	119.58	20	50	62.70
28. Yin	11	8	17	74	82	101	90.90	72.72	88.23
29. Fares Althaqel	16	9	9	108	116	175	80	81.25	88
30. Mary (2)	8	1	9	79	29 82	88	80	100	60
31. Jie	15	15	4	167	157	144	88	76	90.90
32. Joker	26	11	10	162	106	97	50	76.47	90.90

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D2

Focused Instruction Groups: Raw Scores (Task 1) Continued

	Error%			Evaluation		
	T1	T2	T3	T1	T2	T3
1.Victoria	8.23	4.42	4.13	77	76	65
2. John	18.04	16.30	12.50	65	67	70
3. Zoey	20.72	12.71	16.78	65	67	65
4. Amy	17.60	21.47	24.41	70	65	62
5. Paul	18.61	20.76	27.52	66	63	52
6. Bonapeite	12.50	7.69	16.66	78	75	60
7. Brian	7.14	6.40	8.33	70	70	76
8. George	27.67	21.47	25.17	67	65	70
9. Sara	6.75	3.06	7.93	78	83	40
10. Eva	20	9.67	16.66	57	55	40
11. Henry	18.51	18.96	18.75	85	82	83
12. Jasmine	30.07	20.32	24.09	70	63	60
13. Curry	39.40	29.82	38.55	65	62	60
14. Elias	16.97	10.61	14.19	69.72	68	62
15. Maram	11.47	9.09	6.72	70	70	67.17
16. Mia	10.37	15.17	9.16	65	67	62

17. Maya	36.91 31.31	16.32	14.37	65	67	70
18. Dana	31.31	27.65 18.54	21.81	65	66	70
19. Silva	6.77	6.29	9.91	73	68	67
20. Amanda	10.81	8.20	6.06	63	70	82
21. Ana	11.90	11.11	4.31	77	80	77
22. Mary	21.05	18.26	10.45	65	72	65
23. Kareem	14.28	8.33	7.91	75	83	90
24. Tom	13.49	14.65	10.52	77	87	75
25. Leo	17.56	6.57	8.96	77	87	76
26. Fill	17.64	10.34	9.68	77	73	72.44
27. Cherry	20	18.54	15.70	68	63	62.42
28. Yin	14.86	9.75	16.83	55	60	60
29. Fares Althaqel	14.81	7.75	5.14	63	65	75
30. Mary (2)	10.12	3.44	10.22	70	25 60	80
31. Jie	8.98	9.55	2.77	82	80	87
32. Joker	16.04	10.37	10.30	75	65	70

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D3

Saliency Groups: Raw Scores (Task 1)

	Frequency			Occurrence			Correct			Wrong		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Aleed	2	12	8	2	10	9	2	2	1.50	0	10	6.50
2.Mike	15	20	10	8	12	9	11	8	5	4	12	5
3.Arthur	18	16	23	12	9	18	9	6	12	9	10	11
4.Eric	5	8	14	3	7	9	2	5	5	3	3	9
5.Moe	14	18	14	5	10	9	3	14	5	11	4	9
6.Nart	6	18	18	6	13	13	5	13	14	1	5	4
7.Franklin	14	12	14	11	8	10	9	9	11	5	3	3
8.Lulu	14	14	17	9	8	11	11	10	11	3	4	6
9.Andrea	17	28	26	10	16	20	7	18	17	10	10	9
10.Lee	8	20	12	5	15	10	7	13	8	1	7	4
11.Jinner	24	14	17	14	9	12	16	12	13	8	2	4
12.Kristen	10	15	16	7	9	10	8	8	12	2	7	4
13.Lucian	10	19	8	4	13	3	1	9	8	9	10	0
14.Hank	7	6	4	5	4	4	3	4	3	4	2	1
15.Zia	10	23	17.29	5	16	11.58	6	19	13.54	4	4	3.75
16.Enzo	21	25.29	21	14	15.96	13	15	15.25	9	6	10.04	12
17.Riki	9	13	12	7	10	8	6	10	6	3	3	6

18.Brun	17	21	13	10	9	7	8	13	8	9	8	5

Note. Participants 1-5 are intermediate; participants 6-18 are advanced
Imputed scores are highlighted in yellow

Table D3

Saliency Groups: Raw Scores (Task 1) Continued

	Use			Structure			Spelling		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Aleed	0	1	.25	0	6	6	0	3	.25
2.Mike	0	0	0	4	4	3	0	8	2
3.Arthur	0	0	0	5	0	8	4	10	3
4.Eric	0	2	0	3	1	9	0	0	0
5.Moe	0	2	1	11	2	7	0	0	1
6.Nart	0	0	0	1	4	4	0	1	0
7.Franklin	1	0	0	4	2	2	0	1	1
8.Lulu	0	0	1	2	3	5	1	1	0
9.Andrea	0	1	1	10	6	5	0	3	3
10.Lee	0	2	0	1	5	3	0	0	1
11.Jinner	1	0	2	5	2	1	2	0	1
12.Kristen	0	0	0	2	4	4	0	3	0
13.Lucian	6	0	0	2	10	0	1	0	0
14.Hank	0	0	0	3	2	1	1	0	0
15.Zia	0	0	-.13	4	1	2.25	0	3	1.63
16.Enzo	0	.71	2	5	7.88	9	1	1.46	1
17.Riki	0	0	0	3	3	4	0	0	2
18.Brun	3	1	0	6	7	5	0	0	0

Note. Participants 1-5 are intermediate; participants 6-18 are advanced
Imputed scores are highlighted in yellow

Table D3

Saliency Groups: Raw Scores (Task 1) Continued

	Errors			Words			Correct %		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Aleed	22	21	22.75	102	61	68.13	100	16.66	18.75
2.Mike	23	18	15	128	122	88	73.33	40	50
3.Arthur	18	24	25	107	123	126	50	37.50	52.17
4.Eric	26	11	34	149	66	129	40	62.50	35.71
5.Moe	57	7	23	170	82	77	21.42	77.77	35.71
6.Nart	7	13	15	75	68	105	83.33	72.22	77.77
7.Franklin	15	11	16	121	97	116	64.28	75	78.57
8.Lulu	24	14	21	138	115	129	78.57	71.42	64.70
9.Andrea	22	18	16	95	136	128	41.17	64.28	65.38
10.Lee	21	19	14	130	141	122	87.50	65	66.66

11.Jinner	39	14	16	169	101	112	66.66	85.71	76.47
12.Kristen	18	19	8	169	87	97	80	53.33	75
13.Lucian	34	24	8	97	97	129	10	47.36	100
14.Hank	29	22	16	105	85	65	42.85	66.66	75
15.Zia	15	8	7.50	95	99	97.71	60	82.60	78.31
16.Enzo	9	9.33	13	89	91.67	98	71.42	60.30	42.85
17.Riki	12	9	8	83	102	73	66.66	76.92	50
18.Brun	25	18	19	110	135	107	47.05	61.90	61.53

Note. Participants 1-5 are intermediate; participants 6-18 are advanced
Imputed scores are highlighted in yellow

Table D3

Saliency Groups: Raw Scores (Task 1) Continued

	Errors%			Evaluation		
	T1	T2	T3	T1	T2	T3
1.Aleed	21.56	34.42	33.39	40	55	47.88
2.Mike	17.96	14.75	17.04	60	60	55
3.Arthur	16.82	19.51	19.84	67	60	60
4.Eric	17.44	16.66	26.35	55	40	65
5.Moe	33.52	8.53	29.87	58	57	50
6.Nart	9.33	19.11	14.28	70	60	62
7.Franklin	12.39	11.34	13.79	75	68	72
8.Lulu	17.39	12.17	16.27	68	65	65
9.Andrea	23.15	13.23	12.50	63	75	70
10.Lee	16.15	13.47	11.47	65	67	75
11.Jinner	23.07	13.86	14.28	75	80	70
12.Kristen	10.65	21.83	8.24	83	73	78
13.Lucian	35.05	24.74	6.20	50	63	10
						50
14.Hank	27.61	25.88	24.61	60	60	50
15.Zia	15.78	8.08	7.68	62	75	67.17
16.Enzo	10.11	10.18	13.26	75	81.79	77
17.Riki	14.45	8.82	10.95	70	65	65
18.Brun	22.72	13.33	17.75	70	70	75

Note. Participants 1-5 are intermediate; participants 6-18 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D4

Control Groups: Raw Scores (Task 1)

	Frequency			Occurrence			Correct			Wrong		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1. Wukong	14	15	11.89	8	10	7.33	7	10	8	7	5	3.89
2. Alex	8	13	1	6	7	1	4	5	1	4	8	0

3. Laila Ahmed	4	9	6	3	4	6	4	9	6	0	0	0
4. Hazem	8	11	11	7	9	7	1	4	5	7	7	6
5. Jonathan	13	15	11	9	10	8	8	9	8	5	6	3
6. Dead Pool	12	13	13	8	7	8	4	10	10	8	3	3
7. Aiden	15	12	17	8	8	9	11	11	10	4	1	7
8. Jessy	8	11	6.33	6	7	4.56	1	3	.78	7	8	5.56
9. Christine	13	16	12	10	7	7	8	11	8	5	5	4
10. Todd	17	13	11	10	6	5	14	10	11	3	3	0
11. Charlie	19	12	22	12	6	12	12	7	12	7	5	10
12. Anita	11	18	21	7	13	12	8	11	15	3	7	6
13. Zee	16	12	15	7	6	9	12	9	5	4	3	10
14. Crizik	13	12	15	11	9	10	8	8	9	5	4	6
15. Madara Uchiha	21	19	25	17	11	15	16	13	22	5	6	3
16. Good	12	12	21	7	5	11	4	10	7	8	2	14
17. Jack	15.33	19	15	8.75	9	8	15.83	19	14	-.5	0	1

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Table D4

Control Groups: Raw Scores (Task 1) Continued

	Use			Structure			Spelling		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1. Wukong	0	0	-.28	6	5	3.72	1	0	.44
2. Alex	0	0	0	4	6	0	0	2	0
3. Laila Ahmed	0	0	0	0	0	0	0	0	0
4. Hazem	0	0	0	3	7	4	4	0	2
5. Jonathan	1	1	1	4	4	2	0	1	0
6. Dead Pool	3	0	0	5	3	3	0	0	0
7. Aiden	1	0	1	3	1	5	0	0	1
8. Jessy	0	0	-.28	7	8	5.94	0	0	-.11
9. Christine	0	2	0	5	3	4	0	0	0
10. Todd	0	1	0	3	2	0	0	0	0
11. Charlie	0	0	0	7	5	10	0	0	0
12. Anita	0	1	0	3	6	6	0	0	0
13. Zee	0	1	2	4	2	7	0	0	1
14. Crizik	0	0	0	5	4	6	0	0	0
15. Madara Uchiha	0	0	1	5	5	1	0	1	1
16. Good	0	0	0	6	2	12	2	0	2
17. Jack	-.42	0	0	0	0	1	-.08	0	0

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Table D4

Control Groups: Raw Scores (Task 1) Continued

	Errors			Words			Correct%		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1. Wukong	30	33	21.67	97	136	88.67	50	66.66	67.28
2. Alex	29	32	0	83	97	3	50	38.46	100
3. Laila Ahmed	2	7	5	79	77	87	100	100	100
4. Hazem	31	24	19	69	79	88	12.50	36.36	45.45
5. Jonathan	14	20	15	109	124	110	61.53	60	72.72
6. Dead Pool	32	41	26	139	134	129	33.33	76.92	76.92
7. Aiden	17	18	16	177	166	157	73.33	91.66	58.82
8. Jessy	35	41	28.89	108	122	87	12.50	27.27	12.32
9. Christine	15	23	23	121	157	106	61.53	68.75	66.66
10. Todd	16	16	14	140	158	139	82.35	76.92	100
11. Charlie	22	15	23	156	111	156	63.15	58.33	54.54
12. Anita	19	19	22	122	152	189	72.72	61.11	71.42
13. Zee	17	17	30	116	110	146	75	75	33.33
14. Crizik	18	15	29	93	115	154	61.53	66.66	60
15. Madara Uchiha	14	17	13	126	168	180	76.19	68.42	88
16. Good	20	25	26	91	130	189	33.33	83.33	33.33
17. Jack	9.42	18	6	122.33	153	157	103	100	93.33

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Table D4

Control Groups: Raw Scores (Task 1) Continued

	Errors%			Evaluation		
	T1	T2	T3	T1	T2	T3
1. Wukong	30.92	24.26	24.44	57	60	46.22
2. Alex	34.93	32.98	0	55	60	0
3. Laila Ahmed	2.53	9.09	5.74	65	65	83
4. Hazem	44.92	30.37	21.59	60	68	67
5. Jonathan	12.84	16.12	13.63	77	62	65
6. Dead Pool	23.02	30.59	20.15	72	73	68
7. Aiden	9.60	10.84	10.19	75	75	76
8. Jessy	32.40	33.60	33.21	55	63	46.78
9. Christine	12.39	14.64	21.69	77	67	65
10. Todd	11.42	10.12	10.07	67	63	65
11. Charlie	14.10	13.51	14.74	77	75	75
12. Anita	15.57	12.50	11.64	70	67	78
13. Zee	14.65	15.45	20.54	73	72	70
14. Crizik	19.35	13.04	18.83	68	73	70
15. Madara Uchiha	11.11	10.11	7.22	88	95	90
16. Good	21.97	19.23	13.75	80	77	76
17. Jack	7.70	11.76	3.82	80.5	82	80

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Table D5

Focused Instruction Groups: Raw Scores (Task 2)

	Frequency			Occurrence			Correct			Wrong		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Victoria	11	10	12	8	7	6	9	7	8	2	3	4
2. John	0	8	6	0	7	4	0	3	5	0	5	1
3. Zoey	8	6	11	4	5	7	3	3	7	5	3	4
4. Amy	5	3	4	5	2	1	1	3	1	4	0	3
5. Paul	13	9	7	4	6	6	8	6	4	5	3	3
6. Bonapeite	11	13	11	5	8	6	4	13	10	7	0	1
7. Brian	5	15	8	5	6	4	3	10	6	2	5	2
8. George	8	10	9	8	7	6	3	4	5	5	6	4
9. Sara	2	11	9	2	7	7	1	8	4	1	3	5
10. Eva	6	2	9	5	2	4	2	2	4	4	0	5
11. Henry	26	31	16	12	8	9	15	16	13	11	15	3
	13	15									6	
12. Jasmine	10	2	3	8	2	2	5	1	2	5	1	1
13. Curry	5	3	1	4	2	1	3	1	1	2	2	0
14. Elias	9	15	21	5	13	11	4	8	9	5	7	12
15. Maram	16	20	17.69	7	12	9.17	8	14	11.64	8	6	6.06
16. Mia	7	24	12	4	10	7	5	15	8	2	9	4
17. Maya	25	22	16	14	16	9	9	8	7	16	14	9
	16			8						8		
18. Dana	6	5	3	4	4	3	4	4	3	2	1	0
19. Silva	10	12	11	6	7	7	7	6	9	3	6	2
20. Amanda	9	13	17	5	5	11	5	7	14	4	6	3
21. Ana	7	15	19	6	10	11	5	9	10	2	6	9
22. Mary	3	13	19	2	5	11	3	5	14	0	8	5
23. Kareem	13	11	19	8	9	9	10	8	16	3	3	3
24. Tom	13	10	8	7	7	5	6	8	4	7	2	4
25. Leo	7	18	15	5	8	8	3	9	10	4	9	5
26. Fill	5	6	4.50	2	5	2.83	2	4	3.19	3	2	1.31
27. Cherry	16	6	10.31	8	6	6.53	11	6	9	5	0	1.31
28. Yin	10	14	7	7	6	6	2	9	5	8	5	2
29. Fares Althaqel	9	33	9	4	17	5	6	29	9	3	4	0
		24						15				
30. Mary (2)	3	10	5.56	3	9	5.47	1	7	4.25	2	3	1.31
31. Jie	9	12	12	5	10	9	6	6	9	3	6	3
32. Joker	4	5	11	4	4	7	3	3	6	1	2	5

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D5

Focused Instruction Groups: Raw Scores (Task 2) Continued

	Use			Structure			Spelling		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Victoria	0	0	1	2	3	3	0	0	0
2. John	0	3	0	0	0	1	0	2	0
3. Zoey	1	0	0	4	3	4	0	0	0
4. Amy	1	0	0	3	0	3	0	0	0
5. Paul	0	0	0	2	3	2	3	0	1
6. Bonapeite	0	0	0	6	0	1	1	0	0
7. Brian	0	2	1	2	3	1	0	0	0
8. George	1	2	0	4	3	4	0	1	0
9. Sara	1	0	1	0	3	3	0	0	1
10. Eva	0	0	0	3	0	4	1	0	1
11. Henry	1	0	0	8	14	2	2	1	1
12. Jasmine	1	0	0	3	0	1	1	1	0
13. Curry	0	0	0	2	2	0	0	0	0
14. Elias	0	0	1	4	7	9	1	0	2
15. Maram	0	0	-.11	8	6	6.06	0	0	.11
16. Mia	0	0	0	2	8	4	0	1	0
17. Maya	1	3	1	13	7	4	2	4	4
18. Dana	1	0	0	1	1	0	0	0	0
19. Silva	1	0	1	2	6	0	0	0	1
20. Amanda	0	0	0	2	5	1	2	1	2
21. Ana	0	0	1	1	5	7	1	1	1
22. Mary	0	0	0	0	7	3	0	1	2
23. Kareem	0	1	0	1	1	2	2	1	1
24. Tom	0	2	1	7	0	3	0	0	0
25. Leo	1	0	1	2	9	4	1	0	0
26. Fill	0	2	.94	3	0	.25	0	0	.11
27. Cherry	3	0	1.47	1	0	-.81	1	0	.64
28. Yin	0	1	0	7	4	2	1	0	0
29. Fares Althaqel	1	0	0	1	2	0	1	2	0
30. Mary (2)	1	0	.42	0	3	.25	1	0	.64
31. Jie	1	1	0	2	5	2	0	0	1
32. Joker	1	1	3	0	1	2	0	0	0

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow

Table D5

Focused Instruction Groups: Raw Scores (Task 2) Continued

	Errors			Words			Correct %		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Victoria	25	21	19	226	193	221	81.81	70	66.66
2. John	8	30	34	31	196	192	0	37.50	83.33
3. Zoey	80	37	58	292	168	200	37.50	50	63.63
4. Amy	65	32	52	275	143	159	20	100	25
5. Paul	90	52	35	318	220	150	61.53	66.66	57.14
6. Bonapeite	25	18	22	201	228	233	36.36	100	90.90
7. Brian	31	37	29	235	233	236	60	66.66	75
8. George	65	58	33	231	250	196	37.50	40	55.55
9. Sara	32	24	31	173	144	122	50	72.72	44.44
10. Eva	61	9	37	236	62	162	33.33	100	44.44
11. Henry	46	47	23	347	305	188	57.69	51.61	81.25
12. Jasmine	60	32	30	289	145	201	50	50	66.66
13. Curry	87	57	57	369	223	210	60	33.33	100
14. Elias	46	30	55 41	220	166	314	44.44	53.33	42.85
15. Maram	41	41	32.22	276	277	283.75	50	70	65.80
16. Mia	28	51	30	158	246	287	71.42	62.50	66.66
17. Maya	99 64	72 51	36	299	269	216	36	36.36	43.75
18. Dana	29	41	21	225	198	184	66.66	80	100
19. Silva	17	25	25	203	232	315	70	50	81.81
20. Amanda	18	30	32	177	224	383	55.55	53.84	82.35
21. Ana	23	39	32	166	319	287	71.42	60	52.63
22. Mary	11	37	37	51	183	275	100	38.46	73.68
23. Kareem	30	29	29	187	211	289	76.92	72.72	84.21
24. Tom	45	31	41	245	229	286	46.15	80	50
25. Leo	37	46	25	191	315	296	42.85	50	66.66
26. Fill	27	29	18.50	172	254	216.72	40	66.66	70.89
27. Cherry	34	37	26.42	296	254	282.17	68.75	100	87.29
28. Yin	64	31	25	262	181	197	20	64.28	71.42
29. Fares Althaqel	20	27	21	243	422 319	278	66.66	87.87	100
30. Mary (2)	9	32	10.58	71	250	161.31	33.33	70	76.44
31. Jie	17	25	25	194	273	276	66.66	50	75
32. Joker	35	47	32	199	230	255	75	60	54.54

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D5

Focused Instruction Groups: Raw Scores (Task 2) Continued

	Error%	Evaluation

	T1	T2	T3	T1	T2	T3
1.Victoria	11.06	10.88	8.59	65	55	65
2. John	25.80	15.30	17.70	5 50	60	60
3. Zoey	27.39	22.02	29	50	50	55
4. Amy	23.63	22.37	32.70	60	45	40
5. Paul	28.30	23.63	23.33	67	67	50
6. Bonapeite	12.43	7.89	9.44	70	80	70
7. Brian	13.19	15.87	12.28	70	70	80
8. George	28.13	23.20	16.83	65	57	67
9. Sara	18.49	16.66	25.40	55	60	55
10. Eva	25.84	14.51	22.83	60	10 45	50
11. Henry	13.25	15.40	12.23	67	65	65
12. Jasmine	20.76	22.06	14.92	73	55	65
13. Curry	23.57	25.56	27.14	65	50	40
14. Elias	20.90	18.07	17.51	67	72	77
15. Maram	14.85	14.80	11.36	70	73	63.83
16. Mia	17.72	20.73	10.45	63	73	86
17. Maya	33.11 24.42	26.76	16.66	67	73	62
18. Dana	12.88	20.70	11.41	60	63	57
19. Silva	8.37	10.77	7.93	62	55	65
20. Amanda	10.16	13.39	8.35	60	67	75
21. Ana	13.85	12.22	11.14	68	76	72
22. Mary	21.56	20.21	13.45	40	65	77
23. Kareem	16.04	13.74	10.03	70	67	73
24. Tom	18.36	13.53	14.33	65	62	65
25. Leo	19.37	14.60	8.44	67	77	75
26. Fill	15.69	11.41	8.54	65	67	58.03
27. Cherry	11.48	14.56	9.36	73	65	61.19
28. Yin	24.42	17.12	12.69	63	55	57
29. Fares Althaqel	8.23	6.39	7.55	70	85	73
30. Mary (2)	12.67	12.80	6.56	30 40	75	43.78
31. Jie	8.76	9.15	9.05	75	75	70
32. Joker	17.58	20.43	12.54	60	57	55

Note. Participants 1-13 are intermediate; participants 14-32 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D6

Saliency Groups: Raw Scores (Task 2)

	Frequency			Occurrence			Correct			Wrong		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3

1.Aleed	5	11	7.88	4	6	5.88	2	7	6	3	4	1.88
2.Mike	16	7	12	8	5	9	9	3	8	7	4	4
3.Arthur	8	15	17	4	9	11	5	10	10	3	5	7
4.Eric	16	27	8	8	15	5	8	16	7	8	11	1
5.Moe	11	13	19	9	9	12	2	7	11	9	6	8
6.Nart	15	8	7	11	6	6	8	6	6	7	2	1
7.Franklin	19	10	8	9	9	6	11	8	3	8	2	5
8.Lulu	8	12	7	4	5	2	4	5	5	4	7	2
9.Andrea	10	9	7	6	7	3	6	7	4	4	2	3
10.Lee	6	16	13	3	11	9	5	11	9	1	5	4
11.Jinner	21	6	12	8	6	9	14	4	8	7	2	4
12.Kristen	4	9	11	3	7	8	1	7	9	3	2	2
13.Lucian	11	11	2	6	9	1	4	5	1	7	6	1
14.Hank	2	8	7	2	8	4	0	5	3	2	3	4
15.Zia	8	16	10	3	11	6.04	3	11	6.04	5	5	3.95
16.Enzo	13	6	14	8	5	10	9	4	9	4	2	5
17.Riki	15	22	12	6	13	9	9	14	8	6	8	4
18.Brun	7	4.92	0	5	4.92	0	6	4.46	0	1	.46	0

Note. Participants 1-5 are intermediate; participants 6-18 are advanced
Imputed scores are highlighted in yellow

Table D6

Saliency Groups: Raw Scores (Task 2) Continued

	Use			Structure			Spelling		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Aleed	0	1	.38	3	2	1.88	0	1	-.38
2.Mike	1	0	1	5	3	2	1	1	1
3.Arthur	0	0	1	3	1	5	0	4	1
4.Eric	2	4	1	2	6	0	4	1	0
5.Moe	2	2	2	6	3	5	1	1	1
6.Nart	3	0	1	4	1	0	0	1	0
7.Franklin	1	1	0	6	1	3	1	0	2
8.Lulu	0	1	0	4	6	2	0	0	0
9.Andrea	0	1	1	4	1	2	0	0	0
10.Lee	0	0	0	1	5	4	0	0	0
11.Jinner	0	1	0	7	1	4	0	0	0
12.Kristen	1	0	0	2	0	2	0	2	0
13.Lucian	0	2	0	5	4	1	2	0	0
14.Hank	0	1	0	1	2	4	1	0	0
15.Zia	0	2	.63	5	3	3.46	0	0	-.13
16.Enzo	0	1	1	2	1	3	2	0	1
17.Riki	0	2	0	6	6	4	0	0	0
18.Brun	0	.67	0	1	-.08	0	0	-.13	0

Note. Participants 1-5 are intermediate; participants 6-18 are advanced

Imputed scores are highlighted in yellow

Table D6

Saliency Groups: Raw Scores (Task 2) Continued

	Errors			Words			Correct %		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1.Aleed	37	42	43.38	142	211	210.25	40	63.63	76.14
2.Mike	28	37	41	201	186	264	56.25	42.85	66.66
3.Arthur	42	76	75	256	232	349	62.50	66.66	58.82
4.Eric	44	43	19	249	266	175	50	59.25	87.50
5.Moe	32	51	57	127	225	218	18.18	53.84	57.89
6.Nart	31	4	7	276	129	162	53.33	75	85.71
7.Franklin	53	18	38	322	191	196	57.89	80	37.50
8.Lulu	27	29	31	232	267	244	50	41.66	71.42
9.Andrea	36	19	18	196	160	164	60	77.77	57.14
10.Lee	32	46	46	169	315	238	83.33	68.75	69.23
11.Jinner	49	47	40	339	301	268	66.66	66.66	66.66
12.Kristen	55	15	29	241	208	266	25	77.77	81.81
13.Lucian	57	68	12	206	327	68	36.36	45.45	50
14.Hank	27	41	42	80	153	155	0	62.50	42.85
15.Zia	40	33	30.04	177	196	160.50	37.50	68.75	60.40
16.Enzo	13	12	19	182	60	195	69.23	66.66	64.28
17.Riki	39	36	27	226	217	199	60	63.63	66.66
18.Brun	25	12.67	3	179	110.29	21	85.71	90.65	0
									37.50

Note. Participants 1-5 are intermediate; participants 6-18 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D6

Saliency Groups: Raw Scores (Task 2) Continued

	Errors%			Evaluation		
	T1	T2	T3	T1	T2	T3
1.Aleed	26.05	19.90	20.63	58	55	42
2.Mike	13.93	19.89	15.53	70	65	70
3.Arthur	16.40	32.75	21.48	73	68	59.50
4.Eric	17.67	16.16	10.85	65	72	75
5.Moe	25.19	22.66	26.14	52	67	63
6.Nart	11.23	3.10	4.32	85	78	80
7.Franklin	16.45	9.42	19.38	80	70	75
8.Lulu	11.63	10.86	12.70	72	76	75
9.Andrea	18.36	11.87	10.97	70	67	68
10.Lee	18.93	14.60	19.32	55	70	67

11.Jinner	14.45	15.61	14.92	72	70	75
12.Kristen	22.81	7.21	10.90	68	67	70
13.Lucian	27.66	20.79	17.64	60	85	10
14.Hank	33.75	26.79	27.09	55	60	63
15.Zia	22.59	16.83	18.72	67	70	64.46
16.Enzo	7.14	20	9.74	63	10 32.63	57
17.Riki	17.25	16.58	13.56	65	68	67
18.Brun	13.96	11.49	14.28	55	32.63	5

Note. Participants 1-5 are intermediate; participants 6-18 are advanced
Imputed scores are highlighted in yellow; adjusted outliers are highlighted in green

Table D7

Control Groups: Raw Scores (Task 2)

	Frequency			Occurrence			Correct			Wrong		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
1. Wukong	8	6	6.72	6	4	4.56	5	3	4.67	3	3	2.06
2. Alex	7	7	10	5	6	6	4	4	7	3	3	3
3. Laila Ahmed	4	5	4	4	3	4	1	4	3	3	1	1
4. Hazem	6	9	5	4	5	3	1	0	3	5	9	2
5. Jonathan	4	8	12	4	7	6	2	7	5	2	1	7
6. Dead Pool	10	13	12	7	6	8	7	7	7	3	6	5
7. Aiden	12	15	9	8	12	6	8	9	5	4	6	4
8. Jessy	8	9	8.39	6	6	5.67	2	3	3	6	6	5.39
9. Christine	4	5	5	4	3	5	3	3	4	1	2	1
10. Todd	5	4	8	4	4	7	2	1	6	3	3	2
11. Charlie	7	6	11	5	5	4	4	2	6	3	4	5
12. Anita	3	11	9	3	8	8	3	7	6	0	4	3
13. Zee	11	14	6	7	11	4	7	5	3	4	9	3
14. Crizik	12	19	15	5	5	9	7	12	8	5	7	7
15. Madara Uchiha	22	16	17	14	13	10	20	15	12	2	1	5
16. Good	9	15	25	4	8	10	3	7	17	6	8	8
17. Jack	4.50	14	1	2.92	8	1	4.50	10	1	0	4	0

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Table D7

Control Groups: Raw Scores (Task 2) Continued

	Use			Structure			Spelling		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1. Wukong	0	0	-.22	3	2	1.94	0	1	.33
2. Alex	0	0	0	3	3	2	0	0	1
3. Laila Ahmad	1	0	0	2	0	1	0	1	0

4. Hazem	0	1	0	5	4	2	0	4	0
5. Jonathan	0	0	1	1	1	5	1	0	1
6. Dead Pool	0	0	0	3	5	4	0	1	1
7. Aiden	0	0	0	3	5	4	1	1	0
8. Jessy	2	1	1.44	3	4	3.06	1	1	.89
9. Christine	0	0	0	1	2	1	0	0	0
10. Todd	0	1	0	3	2	1	0	0	1
11. Charlie	1	0	0	2	4	4	0	0	1
12. Anita	0	0	0	0	3	2	0	1	1
13. Zee	0	3	0	3	4	3	1	2	0
14. Crizik	0	0	0	4	6	7	1	1	0
15. Madara Uchiha	0	0	1	2	1	4	0	0	0
16. Good	1	1	2	4	7	6	1	0	0
17. Jack	-.25	0	0	.25	4	0	0	0	0

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Table D7

Control Groups: Raw Scores (Task 2) Continued

	Errors			Words			Correct%		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1. Wukong	50	78	49.44	238	348	254.94	62.50	50	69.49
2. Alex	52	33	47	257	148	242	57.14	57.14	70
3. Laila Ahmed	21	12	12	150	141	161	25	80	75
4. Hazem	90	74	47	208	160	182	16.66	0	60
5. Jonathan	26	40	36	184	234	196	50	87.50	41.66
6. Dead Pool	60	71	55	337	317	219	70	53.84	58.33
7. Aiden	70	41	37	387	323	267	66.66	60	55.55
8. Jessy	72	55	48.89	216	187	153.27	25	33.33	35.76
9. Christine	46	39	38	168	198	194	75	60	80
10. Todd	45	41	41	221	241	235	40	25	75
11. Charlie	27	20	40	210	184	301	57.14	33.33	54.54
12. Anita	25	25	35	180	220	195	100	63.63	66.66
13. Zee	50	39	49	222	223	189	63.63	35.71	50
14. Crizik	36	42	45	244	243	264	58.33	63.15	53.33
15. Madara Uchiha	15	12	14	229	208	207	90.90	93.75	70.58
16. Good	81	50	75	246	337	394	33.33	46.66	68
17. Jack	11.83	13	7	92.25	177	58	100	71.42	100

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Table D7

Control Groups: Raw Scores (Task 2) Continued

	Errors%			Evaluation		
	T1	T2	T3	T1	T2	T3
1. Wukong	21	22.41	19.39	50	55	51.78
2. Alex	20.23	22.29	19.42	55	50	72
3. Laila Ahmed	14	8.51	7.45	60	70	67
4. Hazem	43.26	46.25	25.82	65	62	63
5. Jonathan	14.13	17.09	18.36	60	70	73
6. Dead Pool	17.80	22.39	25.11	68	67	65
7. Aiden	18.08	12.69	13.85	63	75	80
8. Jessy	33.33	29.41	31.90	45	50	46.22
9. Christine	27.38	19.69	19.58	60	57	55
10. Todd	20.36	17.01	17.44	55	45	57
11. Charlie	12.85	10.86	13.28	68	72	60
12. Anita	13.88	11.36	17.94	72	85	80
13. Zee	22.52	17.48	25.92	75	70	70
14. Crizik	14.75	17.28	17.04	65	72	73
15. Madara Uchiha	6.55	5.76	6.76	90	88	83
16. Good	32.92	14.83	19.03	73	77	85
17. Jack	12.82	7.34	12.06	39.08	68	15

Note. Participants 1-10 are intermediate; participants 11-17 are advanced
Imputed scores are highlighted in yellow

Appendix E: Screening for Outliers

Outliers: Mixed Factorial ANOVA

The variables that included the raw scores for frequency, occurrence, correct, wrong, correct%, errors, errors%, words, and evaluation in both Task 1 and Task 2 data sets were first screened graphically in order to identify outlying scores within each of the six groups at the three production stages. The *z scores* for all the potential outliers identified in each boxplot were then inspected, and all cases whose *z scores* were greater than +/-2.58 were replaced with the next extreme score whose *z score* did not exceed +/-2.58 within a variable (Fidell & Tabachnick, 2003; Tabachnick & Fidell, 2007a).

Frequency (task 1). A visual representation of the scores for the frequency of the target formulaic sequences showed six potential outlying scores in the upper-intermediate focused instruction and advanced saliency groups (i.e., two each) as well as the upper-intermediate saliency and advanced focused instruction groups (i.e., one each), as shown in Table E1 below. However, none of the detected extreme scores in Table E1 or any other cases within the frequency variables were actual outliers because none of them had *z scores* that exceeded +/-2.58 (see Table E2 below). Accordingly, none of these scores was altered.

Table E1

Boxplots: Frequency (Task 1)

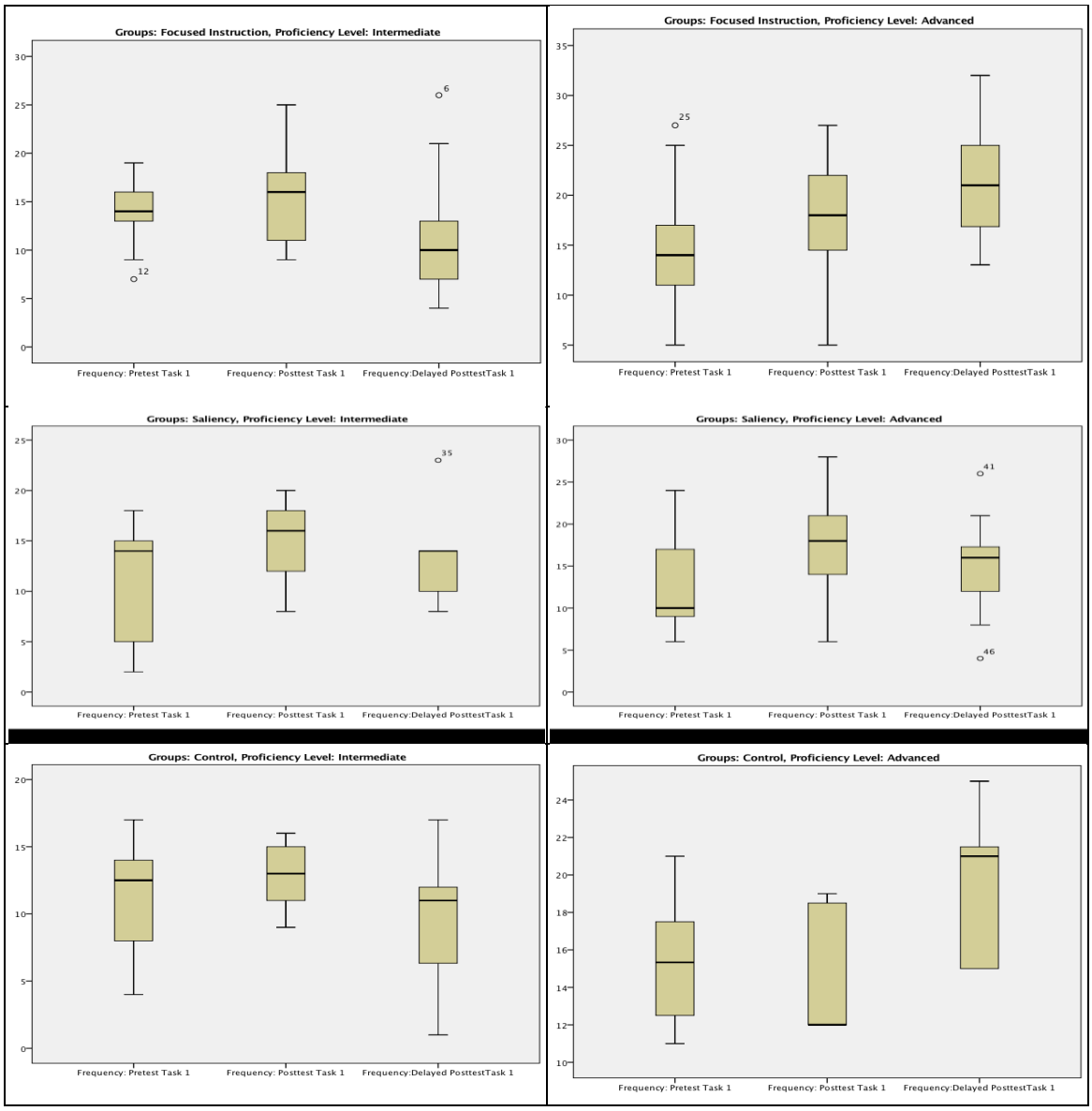


Table E2

Z Scores: Frequency (Task 1)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-2.13*(12)	-1.38	-1.17	-1.27	-1.41	-1.01	-1.79	-1.77	-2.02
	Max.	1.56	2.09	2.17*(6)	1.04	1.08	1.60*(35)	1.44	1.49	1.56
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.75	-2.32	-1.46	-1.23	-1.92	-1.98*(46)	-1.18	-.80	-1.01
	Max.	2.25*(25)	1.56	1.95	2.00	1.73	1.97*(41)	1.54	1.16	1.43

*(n) indicates the number and the *z score* for each extreme case

Occurrence (task 1). As for the occurrence of the target formulaic sequences, the *boxplots* in Table E3 below revealed 4 extreme scores in the saliency groups, three of which were in the upper-intermediate saliency group. However, none of these scores, as shown in Table E4 below, or the minimum and maximum scores for the other cases had a *z score* that was greater than +/-2.58. Therefore, none of these potential outlying scores or any other scores within the occurrence variables was replaced.

Table E3

Boxplots: Occurrence (Task 1)

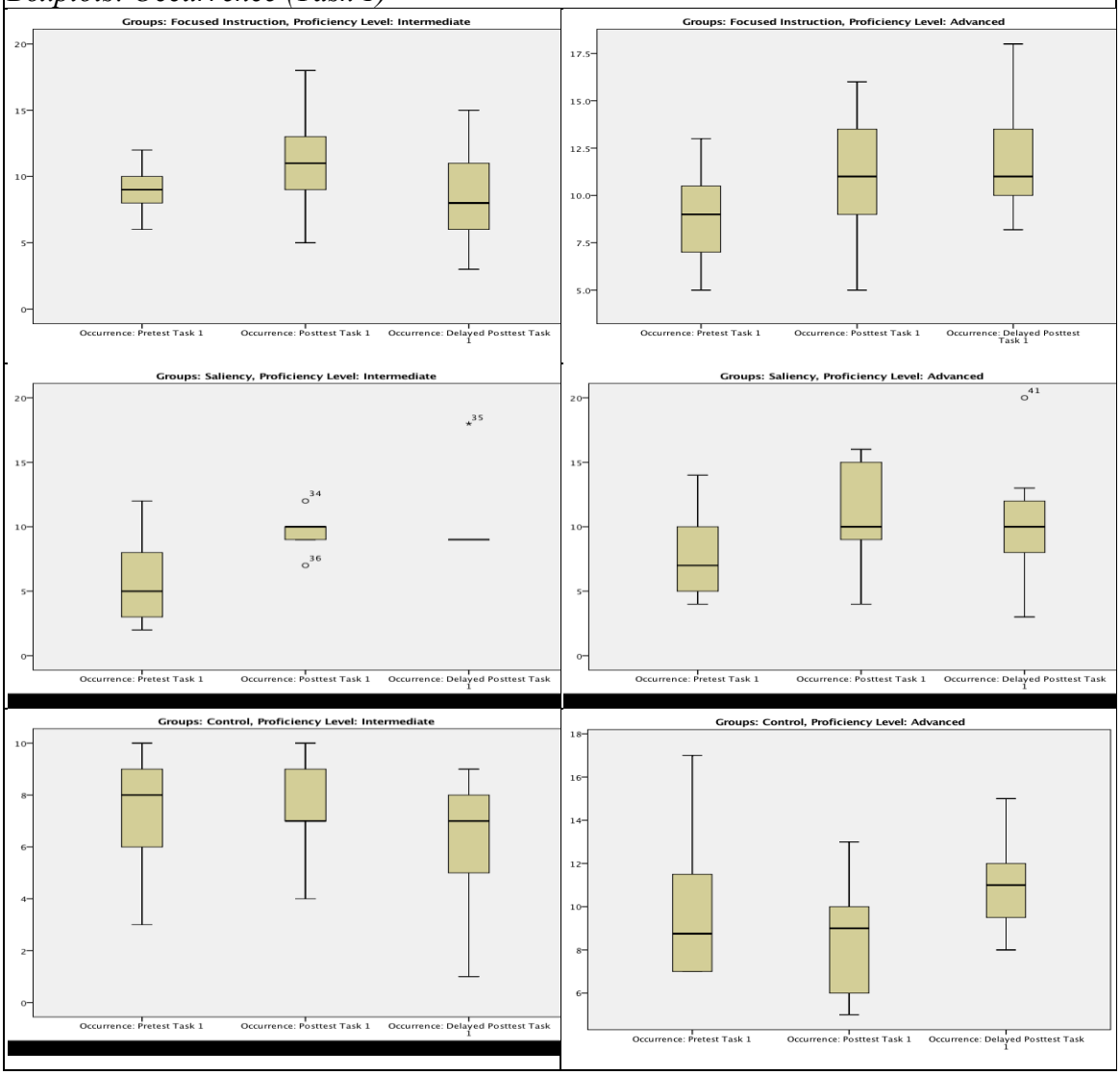


Table E4

Z Scores: Occurrence (Task 1)

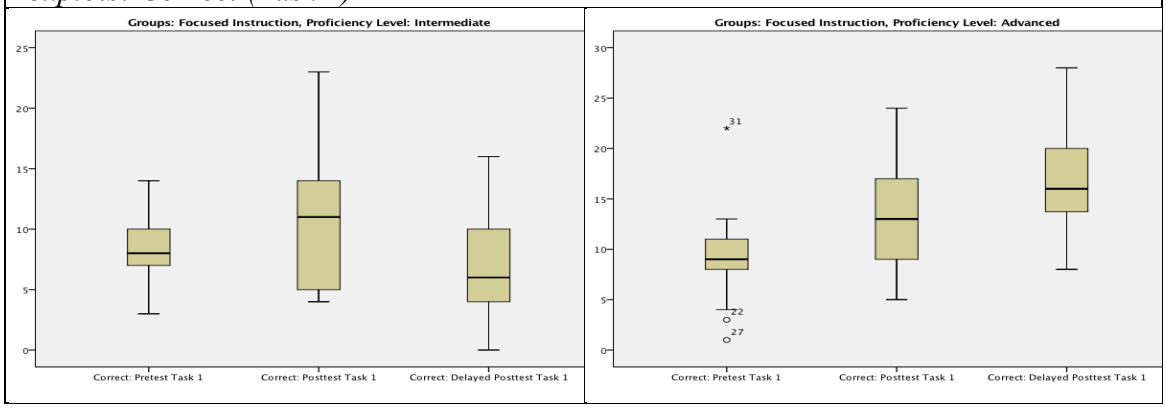
		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.49	-1.87	-1.51	-.98	-1.43*(36)	-.45	-2.12	-1.90	-2.29
	Max.	1.65	2.18	1.81	1.48	1.32*(34)	1.79*(35)	1.18	1.36	1.17
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.64	-2.06	-1.30	-1.25	-1.88	-1.67	-.80	-1.17	-1.30
	Max.	1.87	1.59	2.13	1.70	1.24	2.27*(41)	1.90	1.56	1.73

*(n) indicates the number and the *z score* for each extreme case

Correct (task 1). Ten potential outliers in the variables that represented the number of correct formulaic sequences used by the six groups at the three production stages were identified (see Table E5 below). Six of these outlying scores were detected in the upper-intermediate and advanced saliency groups, three each. The other four potential outliers were detected in the advanced focused instruction and control groups, three and one respectively.

Table E5

Boxplots: Correct (Task 1)



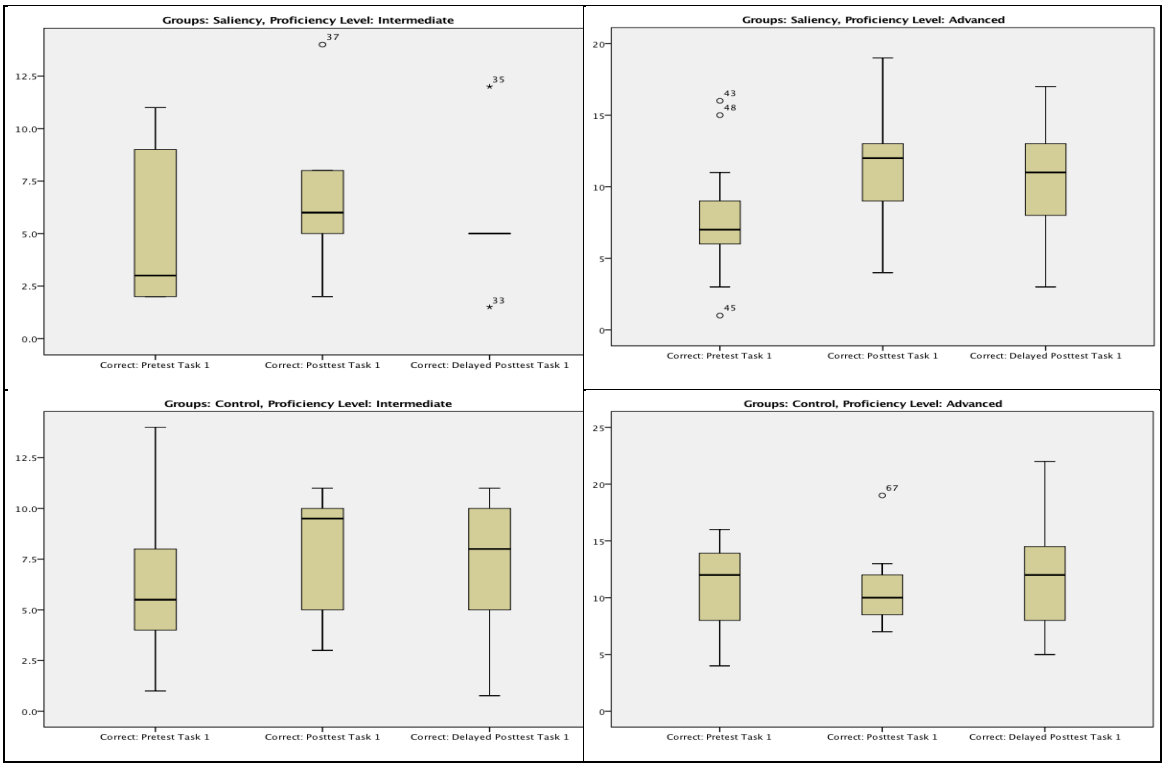


Table E6 below shows that the *z score* for case#31 pretest exceeded +2.58. Therefore, the value for this outlier (i.e., 22) was replaced with the second highest score within the advanced focused instruction group's pretest variable (i.e., 13). The *z scores* of the other potential outlying scores—including the *z scores* for case#22 (i.e., -1.42) and case#48 (i.e., 1.69)—as well as the *z scores* for all other cases within these variables fell within the +/-2.58 range.

Table E6

Z Scores: Correct (Task 1)

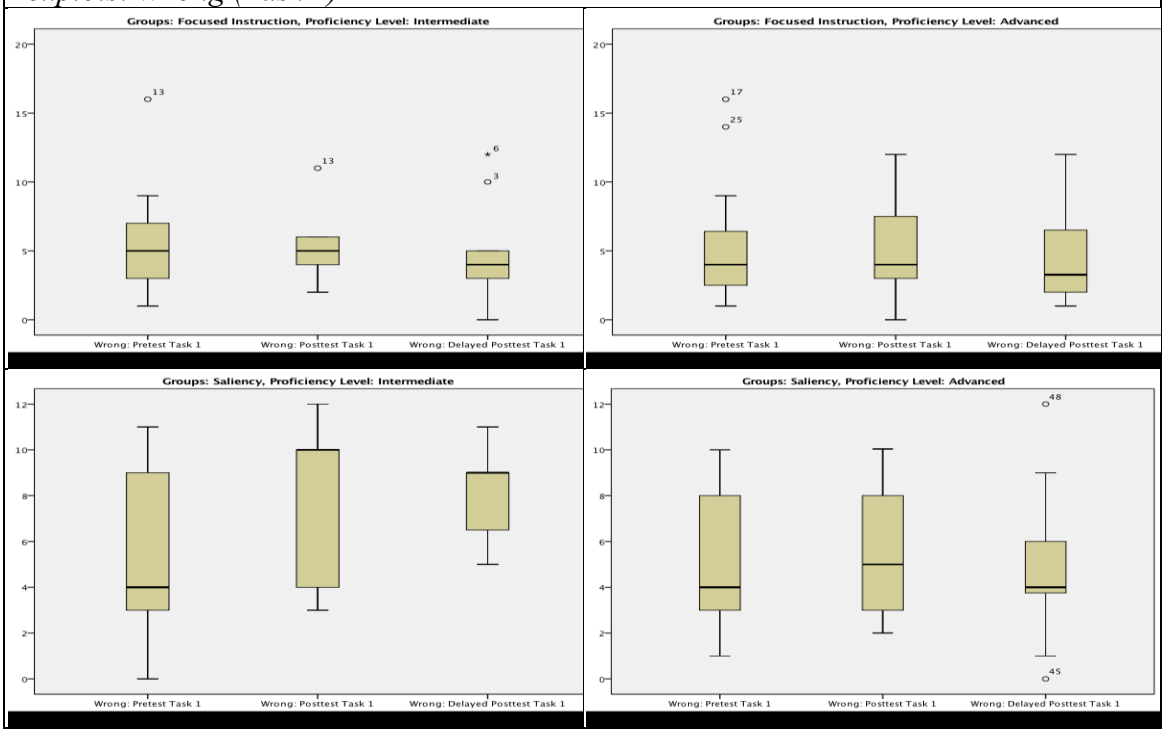
		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.74	-1.07	-1.53	-.79	-1.12	-1.10*(33)	-1.23	-1.73	-1.67
	Max.	1.81	2.15	1.82	1.31	1.57*(37)	1.64*(35)	1.85	.93	1.18
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.87*(27)	-1.49	-1.66	-1.61*(45)	-1.89	-1.93	-1.55	-.99	-1.22
	Max.	2.85*(31)	1.94	2.24	1.92*(43)	1.75	1.79	1.17	1.98*(67)	1.75

*(n)indicates the number and *z score* of an extreme case; green represents an actual outlier

Wrong (task 1). The *boxplots* generated to detect potential outlying scores within the variables that included the raw numbers of incorrect formulaic sequences revealed the presence of nine extreme scores (see Table E7 below). Four of these extreme scores were identified in the upper-intermediate focused instruction group. The other five outlying scores were in the advanced focused instruction group's pretest (i.e., two), the advanced saliency group's delayed posttest (i.e., two), and the advanced control group's pretest (i.e., one). Of these potential outliers, the *z scores* for three cases (i.e., case#13 pretest, case#13 posttest, and case#17 pretest) were greater than +2.58, as shown in Table E8 below.

Table E7

Boxplots: Wrong (Task 1)



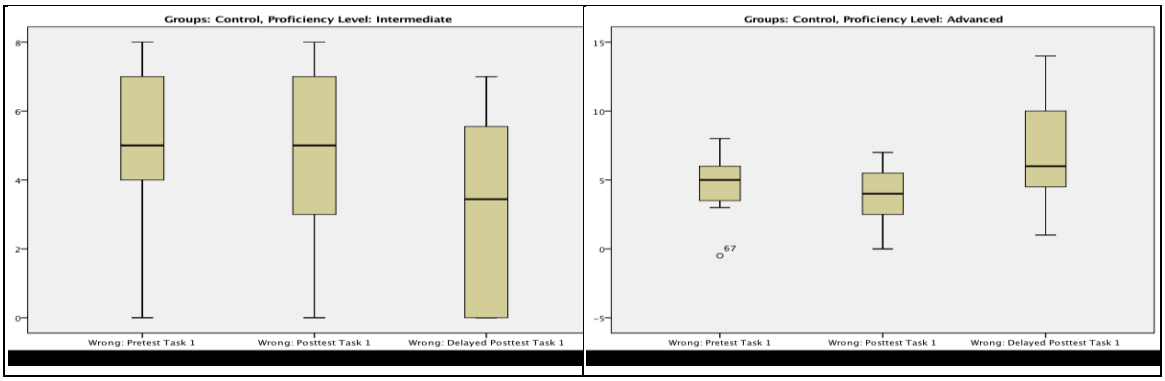


Table E8

Z Scores: Wrong (Task 1)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.17	-1.34	-1.35	-1.20	-1.19	-1.32	-2.08	-1.64	-1.26
	Max.	2.71*(13)	2.59*(13)	2.34*(6)	1.24	1.04	1.23	1.25	1.22	1.46
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.06	-1.57	-1.14	-1.28	-1.23	-1.52*(45)	-1.80*(67)	-1.60	-1.37
	Max.	2.62*(17)	2.27	2.29	1.59	1.39	2.32*(48)	1.26	1.30	1.53

*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

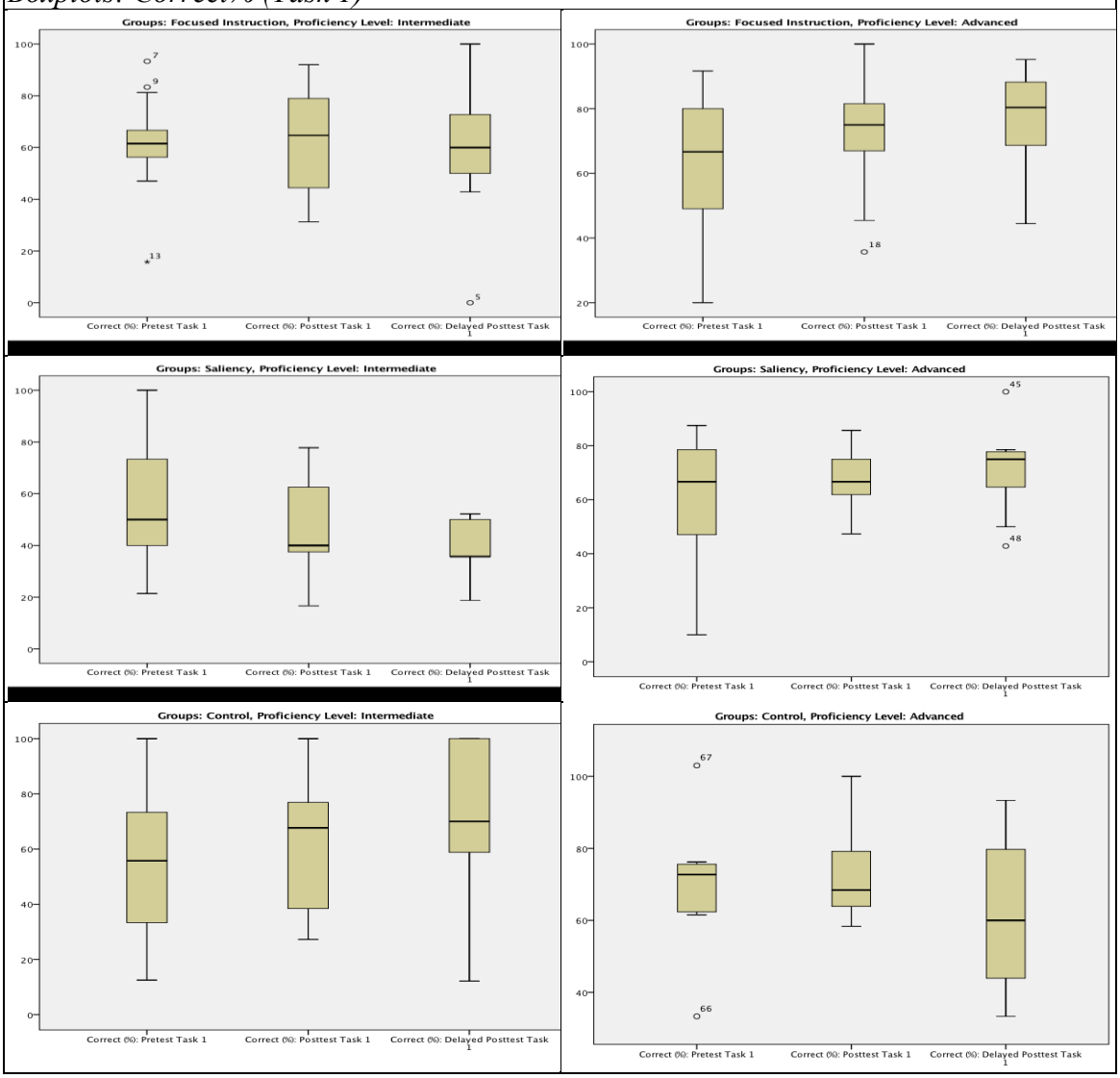
The raw score for the pretest of case#13 (i.e., 16) was replaced with 9, the second highest score in the pretest variable for the upper-intermediate focused instruction group. The raw score for case#17 pretest (i.e., 16), in turn, was replaced with the second highest score within the advanced focused instruction group's pretest, that is, 14. Since the *z score* for case#13 posttest exceeded the cut-off for this study by .01, this case was not expected to affect the results and, hence, was not altered. The *z scores* for all other cases—including the other potential outliers shown in Table E8 above and the *z scores* for case#25 pretest (*z score*=2.13) and case#3 delayed posttest (*z score*=1.73)—did not exceed +/-2.58.

Correct% (task 1). Nine potential outliers were detected in the *boxplots* generated for the percentage of the correct formulaic sequences that the six groups used

(see Table E9 below). Five of these outliers were identified in the upper-intermediate and advanced focused instruction groups. The other four potential outliers were equally divided between the advanced saliency group and the advanced control group.

Table E9

Boxplots: Correct% (Task 1)



The z scores presented in Table E10 below demonstrated that none of these scores were actual outliers because they did not exceed the *z score* threshold set for this study. Nor did the *z scores* for case#9 (*z score*=1.14) or case#67 (*z score*=1.61) or the z scores for all other cases.

Table E10

Z Scores: Correct% (Task 1)

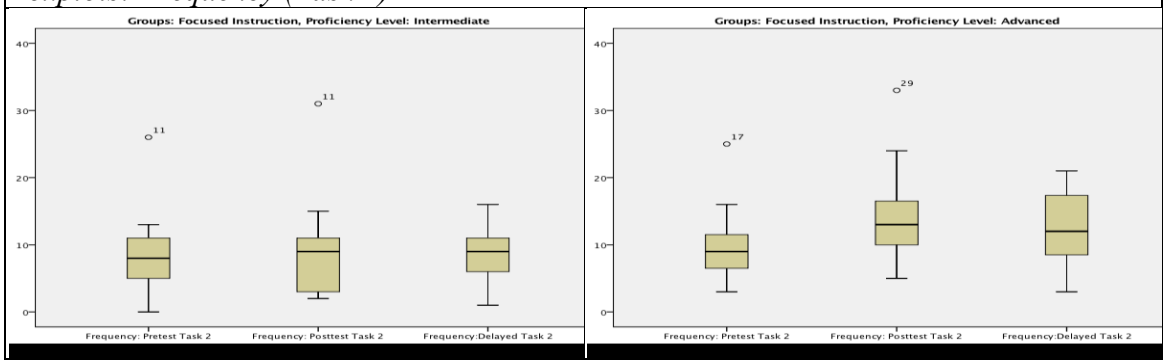
		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-2.39*(13)	-1.59	-2.52*(5)	-1.17	-1.28	-1.46	-1.45	-1.54	-2.09
	Max.	1.66*(7)	1.48	1.74	1.41	1.30	1.02	1.62	1.48	1.09
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.99	-2.18*(18)	-2.36	-2.39	-1.87	-1.91*(48)	-1.72*(66)	-1.03	-1.20
	Max.	1.28	1.59	1.22	1.21	1.62	2.08*(45)	1.61	1.85	1.31

*(n) indicates the number and the *z score* for each extreme case

Frequency (task 2). Table E11 revealed eight potential outlying scores in the variables representing the frequency of the target formulaic sequences that were used in essay responses. Six of these extreme scores were in the upper-intermediate focused instruction group, the advanced focused instruction group, and the advanced control group, two each. The other two extreme scores were detected in the upper-intermediate and advanced saliency groups. Of these eight scores, 4 cases were actual outliers because their *z scores*, as shown in Table E12 below, exceeded +2.58.

Table E11

Boxplots: Frequency (Task 2)



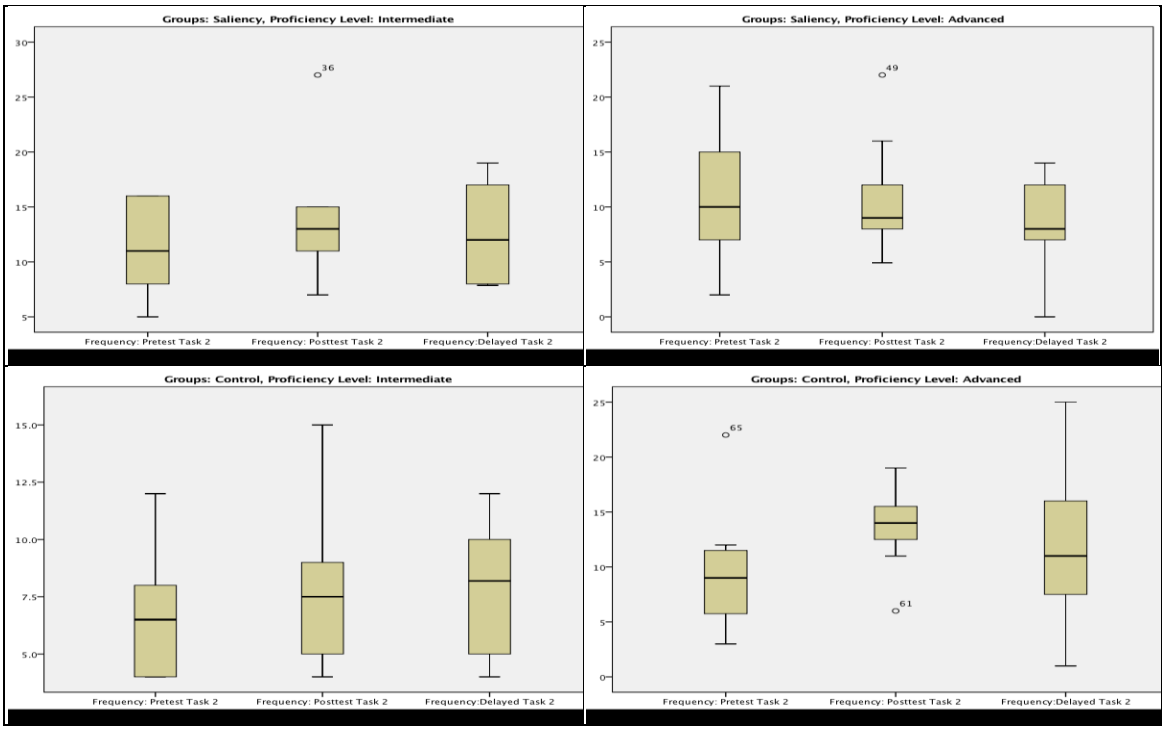


Table E12

Z Scores: Frequency (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.31	-.96	-1.77	-1.27	-1.01	-.96	-1.02	-1.15	-1.41
	Max.	2.72*(11)	2.78*(11)	1.94	.99	1.65*(36)	1.22	1.90	1.93	1.40
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.22	-1.24	-1.73	-1.53	-1.17	-2.04	-1.08	-1.84*(61)	-1.40
	Max.	2.89*(17)	2.67*(29)	1.56	1.81	2.34*(49)	1.33	1.94*(65)	1.32	1.66

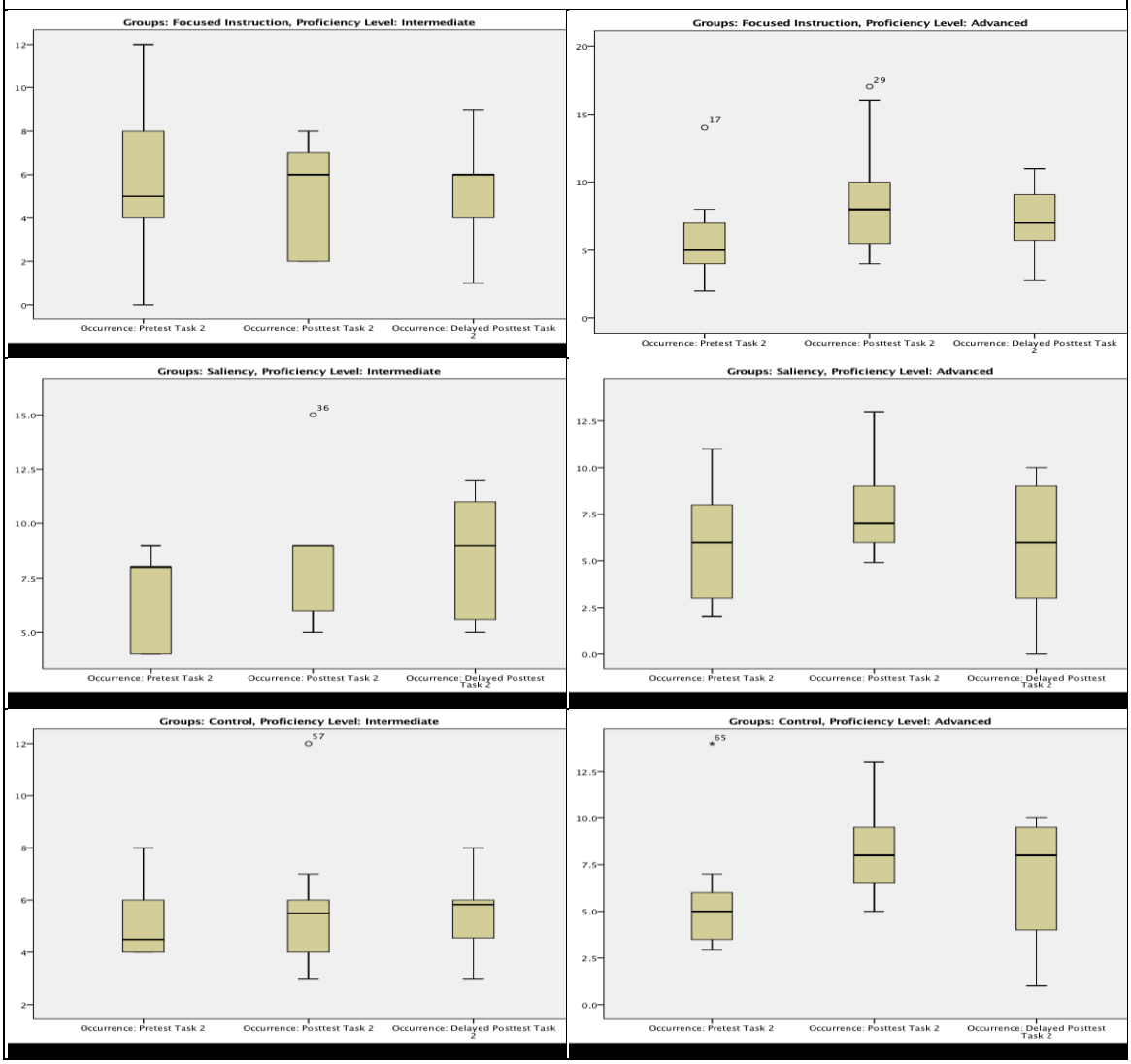
*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

The frequency raw scores for case#11 pretest (i.e., 26) and posttest (i.e., 31) were respectively replaced with 13 and 15, the second highest frequency scores for within each variable. The raw score for case#17 pretest (i.e., 25) and case#29 posttest score (i.e., 33), in turn, were respectively changed to 16 and 24, the second highest score within the advanced focused instruction group's pretest and posttest. The other four scores were not altered because their *z scores* did not exceed the threshold for this study, as shown in Table E12 above. Nor did the *z scor*es for any other cases.

Occurrence (task 2). An exploration of the *boxplots* generated for the distinct types the six groups used in their essay responses at the three production stages revealed five potential outlying scores. These scores, as shown in Table E13, were scattered in the upper-intermediate saliency and control groups as well as the advanced focused instruction and control groups.

Table E13

Boxplots: Occurrence (Task 2)



Of these scores, only the *z score* for case#17 pretest exceeded +2.58, as shown in Table E14 below. Therefore, the raw score for case#17 pretest (i.e., 14) was replaced with 8, the second highest score within the advanced focused instruction group's prestest.

Table E14

Z Scores: Occurrence (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.77	-1.36	-1.57	-1.08	-.97	-1.16	-.81	-.99	-1.74
	Max.	2.17	1.11	1.70	1.00	1.59*(36)	1.11	1.90	2.43*(57)	1.71
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.32	-1.20	-1.81	-1.36	-1.12	-1.67	-.76	-1.12	-1.57
	Max.	3.10*(17)	2.20*(29)	1.34	1.95	1.97	1.31	2.11*(65)	1.61	.97

*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

The other four scores along with all other cases, on the other hand, had *z scores* between +2.58 and -2.58 and, thus, were not altered.

Correct (task 2). Nine potential outlying scores were detected in the variables representing the raw scores for the number of correct formulaic sequences, most of which were in the posttest variables (see Table E15 below). Almost all the identified extreme values, as shown in Table E16 below, had *z scores* that were between +2.58 and -2.58. The *z scores* for the posttests of case#15 (i.e., .92) and case#16 (i.e., 1.09) were also within the same range. Only case#29, as shown in Table E16 below, exceeded the cut-off score for this study.

Table E15

Boxplots: Correct (Task 2)

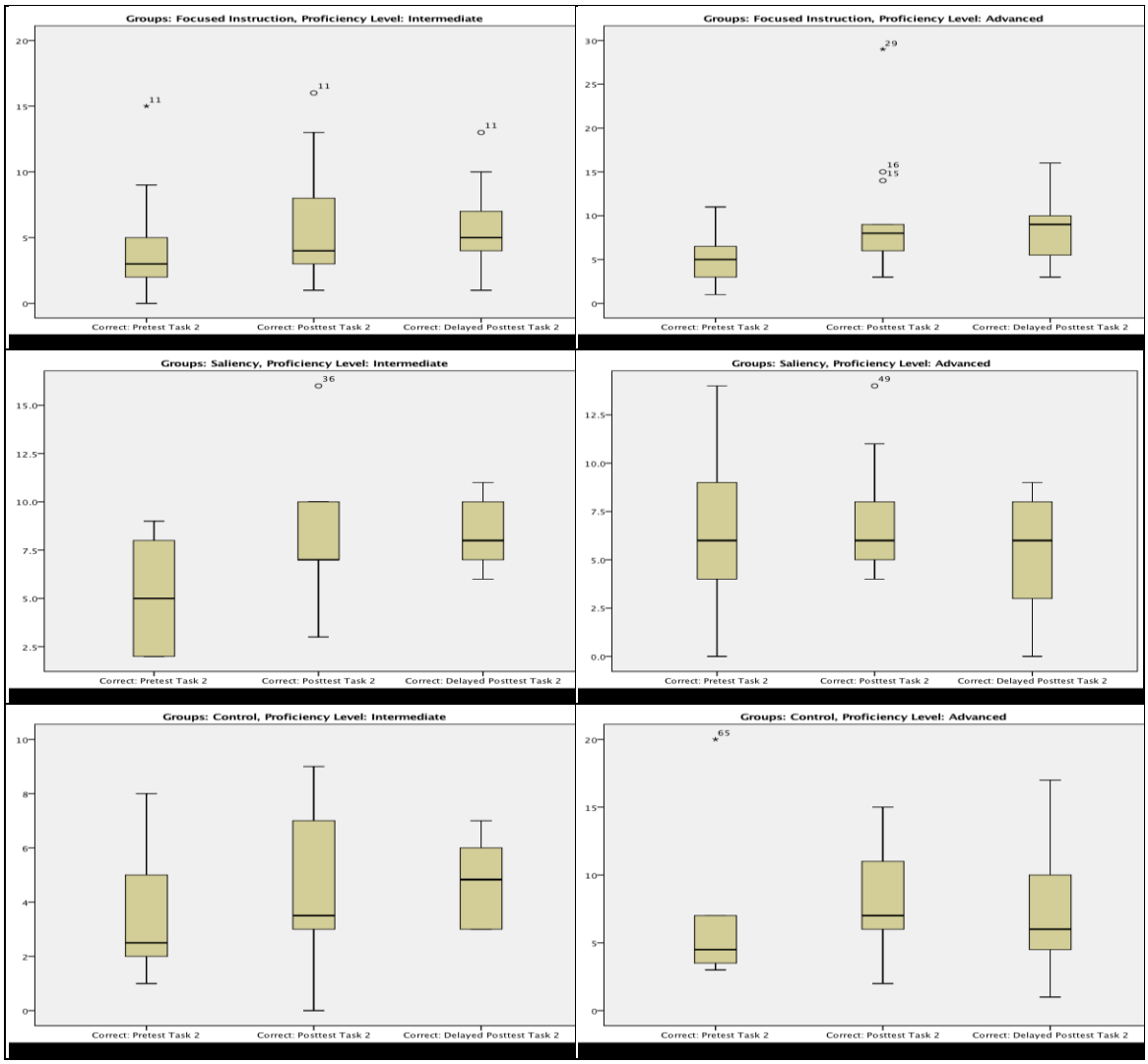


Table E16

Z Scores: Correct (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.07	-1.04	-1.26	-.98	-1.16	-1.16	-1.02	-1.46	-1.14
	Max.	2.58*(11)	2.14*(11)	2.19*(11)	1.16	1.53*(36)	1.25	1.83	1.75	1.44
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.53	-.98	-1.47	-1.55	-.96	-1.76	-.65	-1.43	-1.21
	Max.	2.06	3.52*(29)	2.02	1.97	2.21*(49)	1.14	2.18*(65)	1.53	1.73

*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

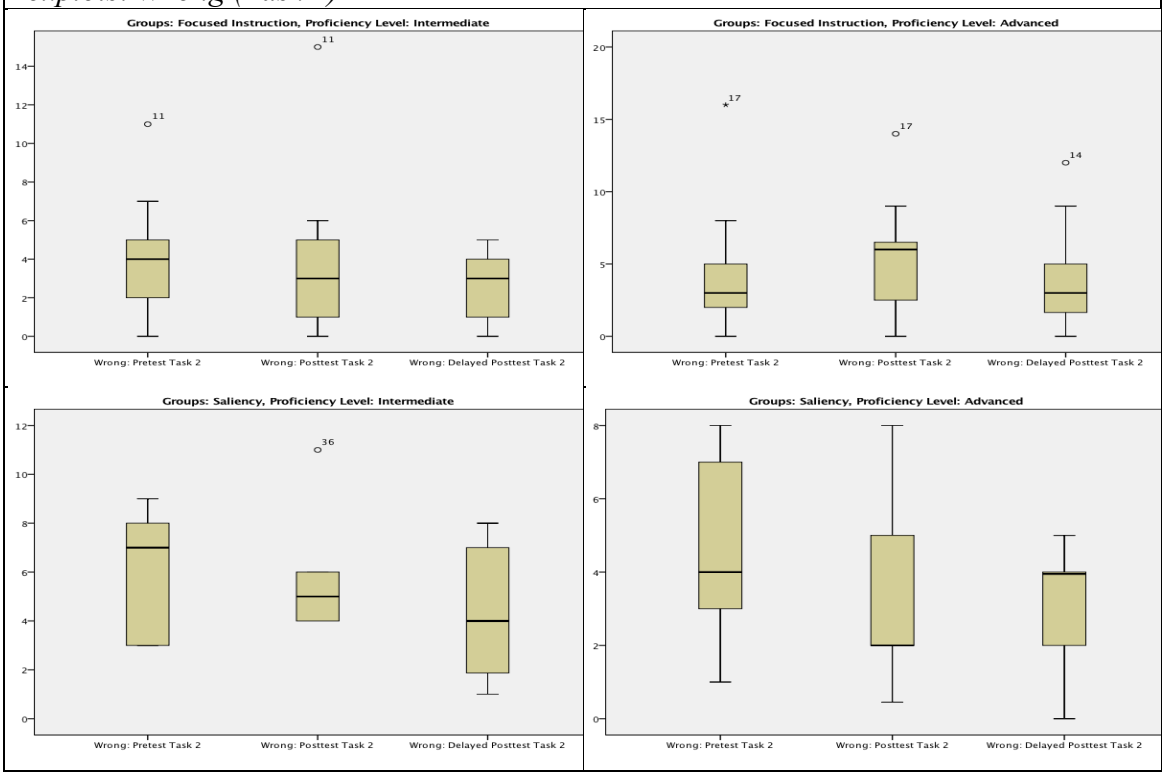
The total number of correct formulaic sequences for case#29 posttest (i.e., 29) was, therefore, changed to 15, the second highest number of correct formulaic sequences

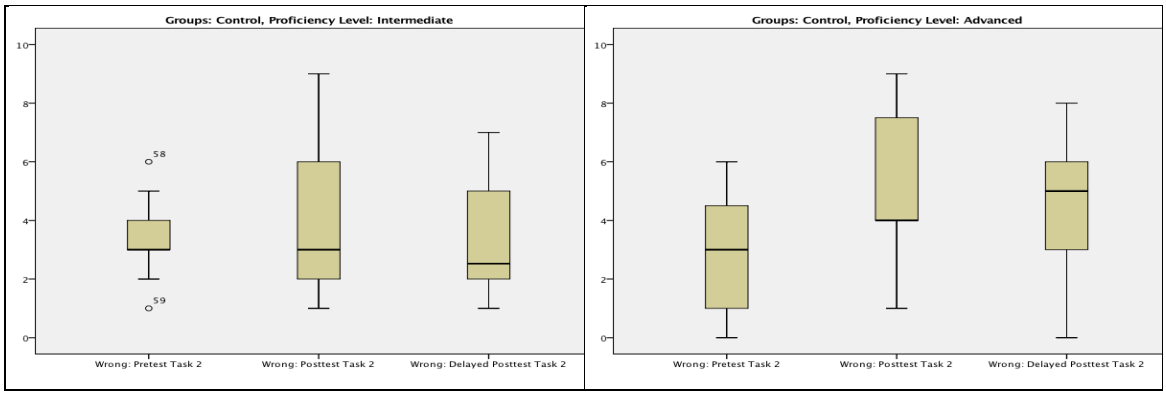
within the advanced focused instruction group's posttest. The *z scores* for all other cases fell within +/-2.58.

Wrong (task 2). Eight potential outlying scores were identified in the variables for the incorrect formulaic sequences that the six groups produced in their essay responses, see Table E17 below. Five of these potential outlying scores were identified in the upper-intermediate and advanced focused instruction groups, and the remaining three outlying scores were detected in the upper-intermediate saliency and control groups.

Table E17

Boxplots: Wrong (Task 2)





Of the eight potential outliers, the *z scores* for three cases (i.e., case#17 pretest and posttest and case#11 posttest) exceeded +2.58 (see Table E18 below). The raw score for case#11 posttest (i.e., 15) was replaced with 6, the second highest number of incorrect formulaic sequences within the posttest variable of the upper-intermediate focused instruction group. The pretest score for case#17 (i.e., 16) was also changed to 8, the second highest scores within the advanced focused instruction group's pretest variable. Case#17 posttest, on the other hand, was not altered because it only exceeded the *z score* threshold by .01 (see Table E18 below), and, hence, it was not expected to affect the results.

Table E18

Z Scores: Wrong (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.42	-.89	-1.69	-1.06	-.69	-1.10	-1.62*(59)	-1.14	-1.11
	Max.	2.41*(11)	2.89*(11)	1.36	1.06	1.71*(36)	1.18	1.90*(58)	1.91	1.85
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.19	-1.54	-1.24	-1.49	-1.33	-1.84	-1.22	-1.52	-1.64
	Max.	3.27*(17)	2.59*(17)	2.49*(14)	1.46	1.88	1.23	1.34	1.32	1.32

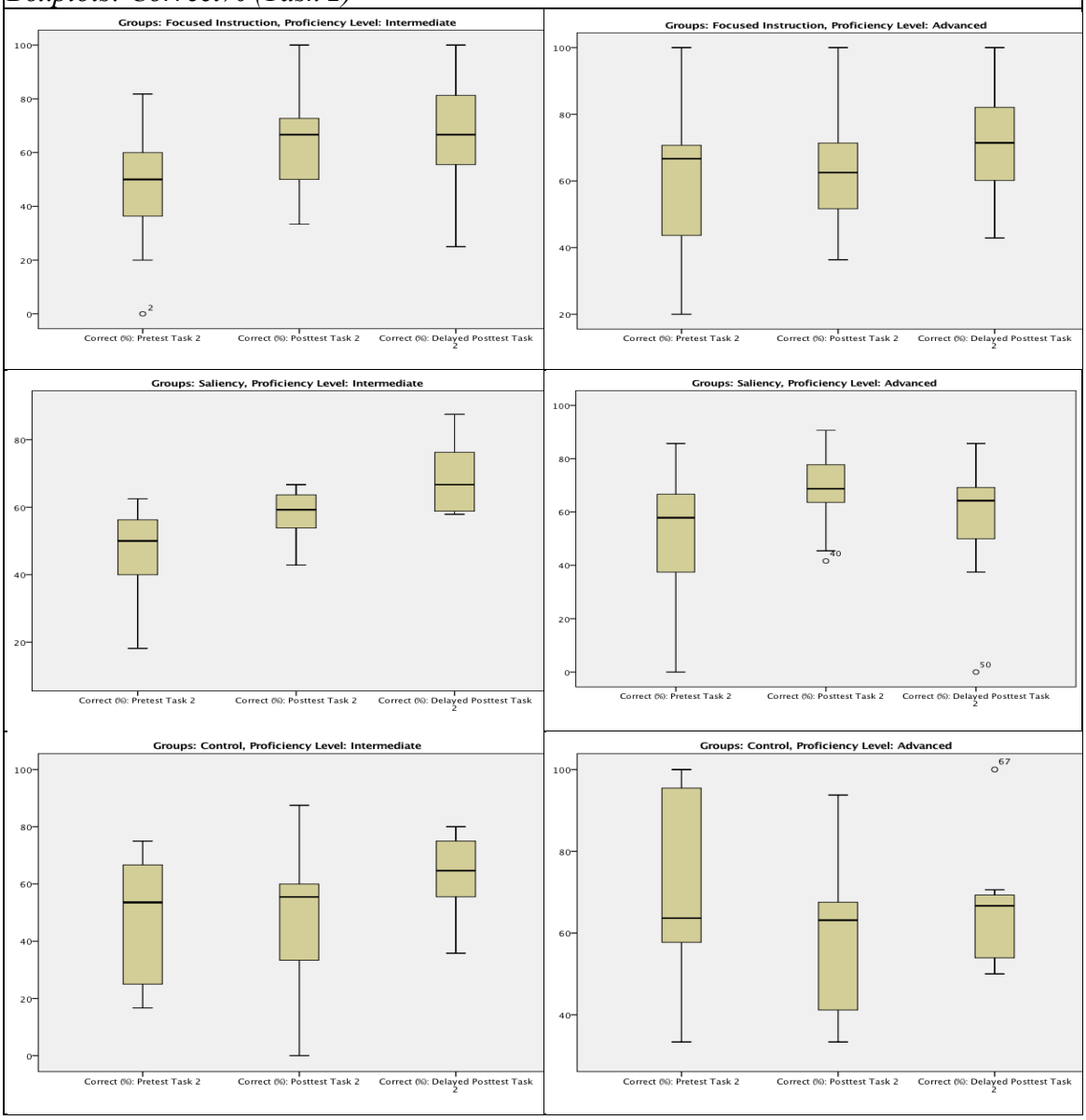
*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

None of the other *z scores* listed in Table E18 above or the *z scores* for all other cases exceeded +/-2.58.

Correct% (task 2). The three variables representing the raw scores for the percentage of the correct formulaic sequences included four potential outliers in the upper-intermediate focused instruction group (one score), the advanced saliency group (two scores), and the advanced control group (one score), as illustrated in Table E19 below.

Table E19

Boxplots: Correct% (Task 2)



The *z scores* presented in Table E20 below demonstrated that the *z score* for case#50 delayed posttest was greater than -2.58. The percentage of the correct formulaic sequences for case#50 delayed posttest (0%) was, thus, replaced with the second lowest percentage within the advanced saliency group's delayed posttest variable, that is, 37.5%. All the other extreme scores and cases remained unaltered because their *z scores* were within +/-2.58.

Table E20

Z Scores: Correct% (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-2.15*(2)	-1.31	-1.96	-1.57	-1.53	-.92	-1.53	-1.97	-1.79
	Max.	1.76	1.49	1.65	.99	1.00	1.45	1.25	1.43	1.22
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.96	-1.66	-1.67	-2.23	-1.97*(40)	-2.61*(50)	-1.51	-1.16	-.95
	Max.	2.17	2.24	1.74	1.40	1.68	1.25	1.10	1.66	1.99*(67)

*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

Words (task 1). Five potential outlying scores were identified in the upper-intermediate focused instruction group (two scores), the advanced focused instruction group (one score), the upper-intermediate control group (one score), and the advanced control group (one score), as shown in Table E21 below. Of the five scores, the *z score* for case#30 was greater than -2.58 (see Table E22 below). The number of words produced by case#30 in the posttest (i.e., 29) was, therefore, replaced with the second lowest number of words generated by the advanced focused instruction group at the posttest stage (i.e., 82).

Table E21

Boxplots: Words (Task 1)

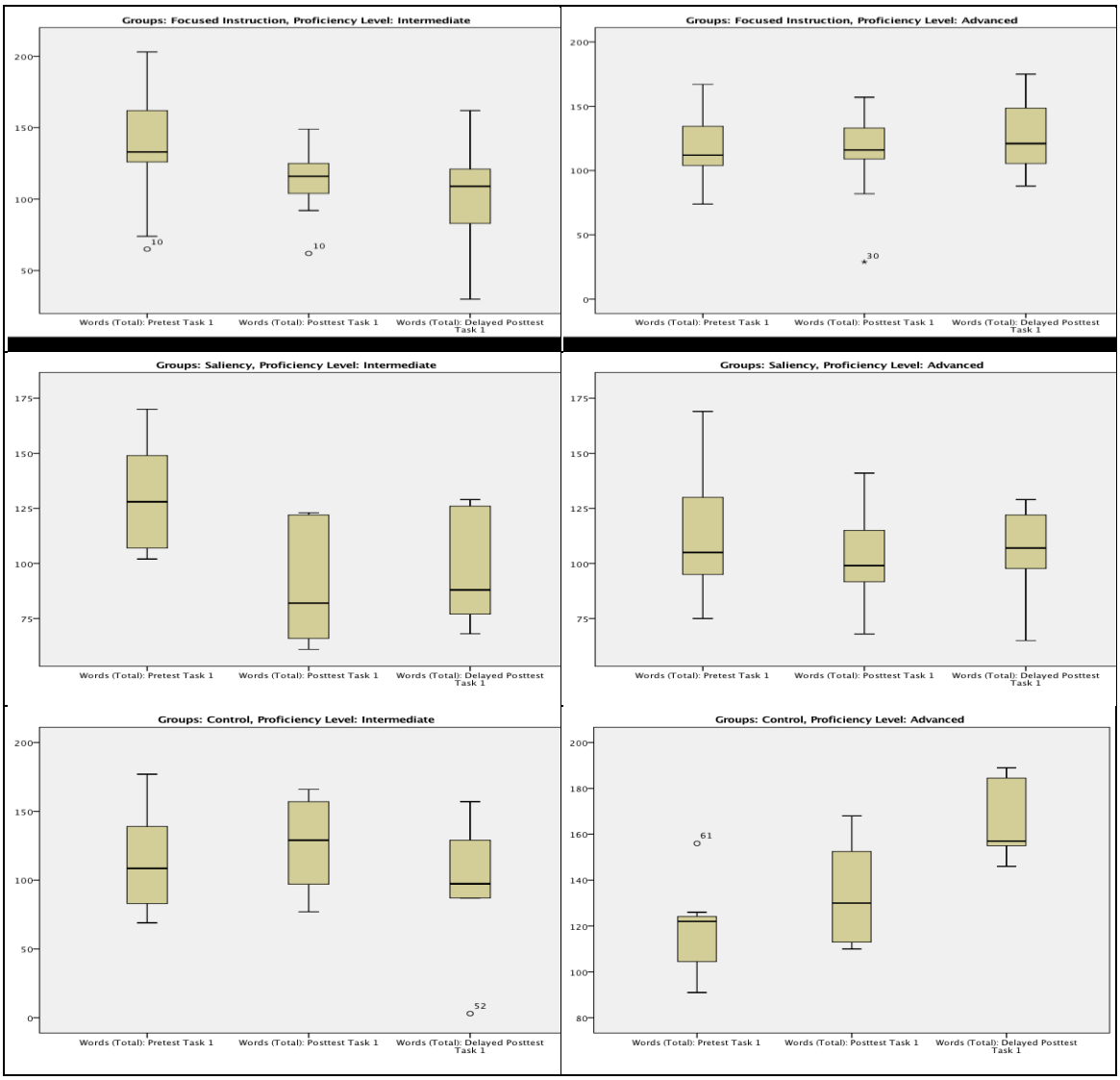


Table E22

Z Scores: Words (Task 1)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.82*(10)	-2.28*(10)	-1.93	-1.02	-.99	-1.05	-1.31	-1.50	-2.31*(52)
	Max.	1.62	1.47	1.67	1.36	1.07	1.11	1.96	1.28	1.37
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.62	-2.95*(30)	-1.55	-1.26	-1.66	-2.03	-1.23	-1.03	-1.18
	Max.	1.82	1.41	1.87	1.82	1.69	1.13	1.73*(61)	1.44	1.20

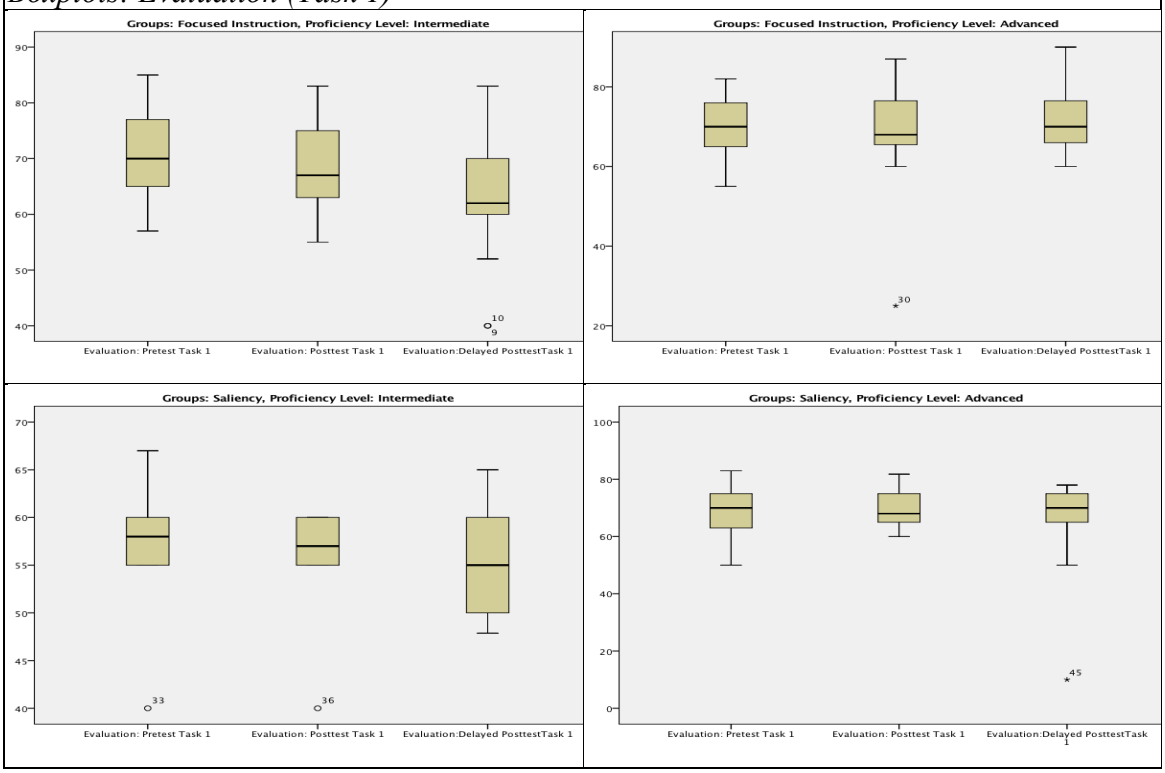
*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

Since the *z scores* for the other potential outliers (see Table E22 above) and all other cases were not greater than +/-2.58, none of them was altered.

Evaluation (task 1). Nine potential outliers were identified in the variables that included the raw scores for the evaluation that the six groups received for their graphical summaries (see Table E23 below). Based on the *z scores* presented in Table E24 below, case#30 posttest and case#45 delayed posttest were outliers. Therefore, the evaluation scores that case#30 received for the posttest (i.e., 25) and case#45 received for the delayed posttest (i.e., 10) were respectively replaced with 60 and 50, the second lowest scores within the posttest and the delayed posttest variables of the advanced focused instruction and saliency groups. All other potential outlying scores and cases had *z scores* that fell within +/-2.58.

Table E23

Boxplots: Evaluation (Task 1)



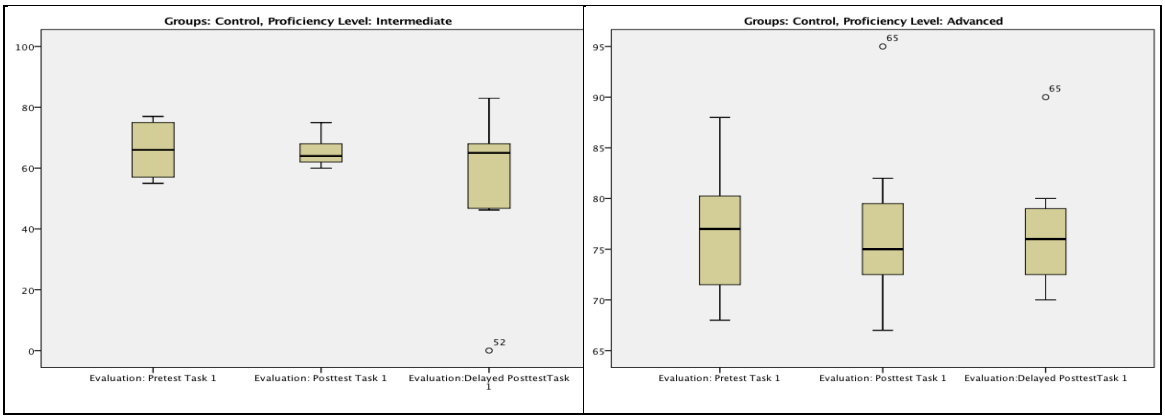


Table E24

Z Scores: Evaluation (Task 1)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.77	-1.67	-1.75*(9&10)	-1.60*(33)	-1.73*(36)	-1.09	-1.23	-1.08	-2.49*(52)
	Max.	1.98	1.74	1.70	1.10	.67	1.34	1.23	1.82	1.06
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-2.23	-3.31*(30)	-1.42	-2.17	-1.33	-3.02*(45)	-1.25	-1.13	-1.02
	Max.	1.76	1.33	2.09	1.77	1.74	.76	1.64	1.95*(65)	1.90*(65)

*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

Words (task 2). According to Table E25, there were 12 potential outliers that were scattered in the variables for the number of words the six groups generated when responding to essay prompts at the three production stages. The *z scores* for almost all the cases, as shown in Table E26 below, were within the range of +/-2.58. The only case that exceeded the +/-2.58 threshold was case#29 posttest. The number of words generated by case#29 in the posttest (i.e., 422 words) was, therefore, replaced with the second highest number of words generated by the advanced focused instruction group in the posttest (i.e., 319 words). None of the other cases was altered because their *z scores* fell within the +/-2.58 range.

Table E25

Boxplots: Words (Task 2)

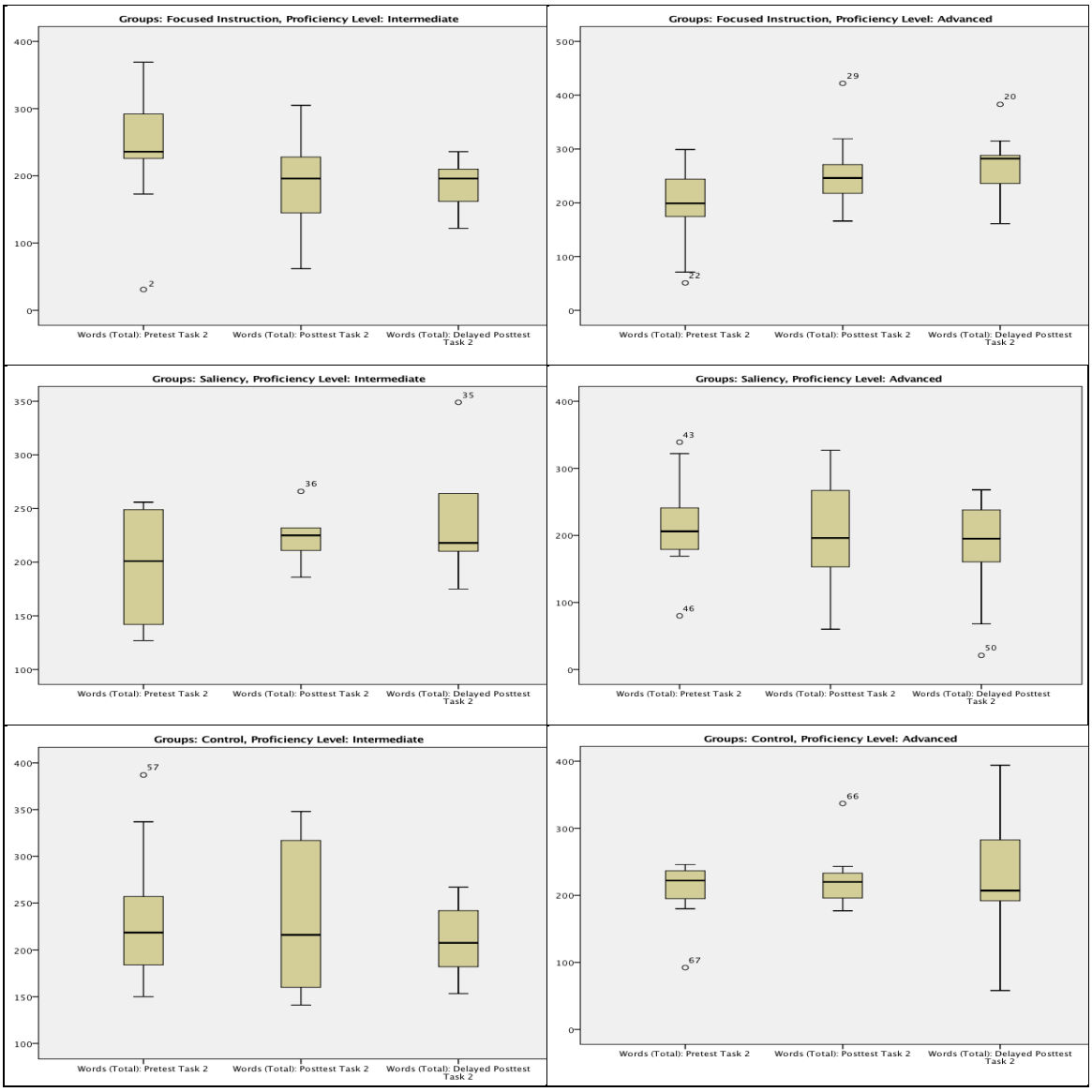


Table E26

Z Scores: Words (Task 2)

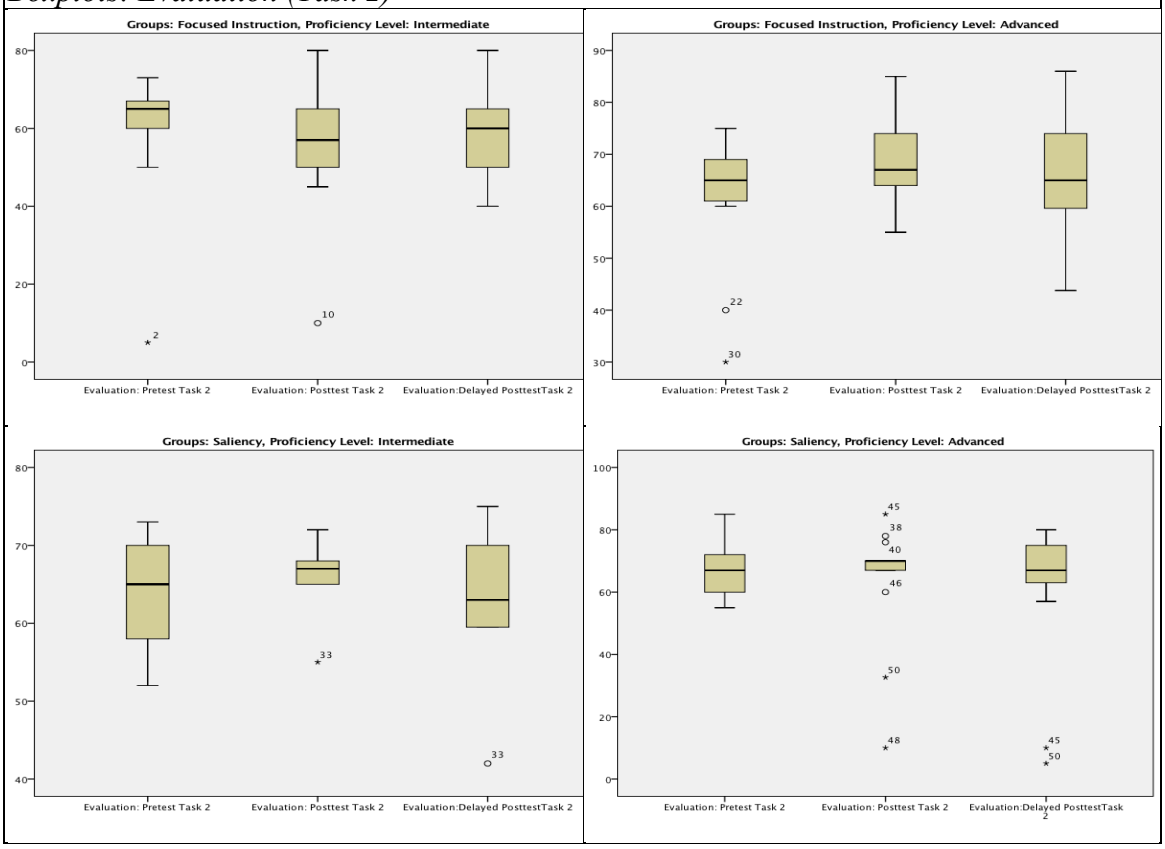
		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-2.51*(2)	-2.13	-2.02	-1.15	-1.30	-1.02	-1.17	-1.16	-1.46
	Max.	1.40	1.82	1.37	1.03	1.43*(36)	1.58*(35)	2.03*(57)	1.55	1.44
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-2.31*(22)	-1.41	-2.03	-2.01*(46)	-1.74	-2.19*(50)	-2.06*(67)	-.94	-1.64
	Max.	1.49	2.93*(29)	2.20*(20)	1.78*(43)	1.51	1.22	.79	2.05*(66)	1.57

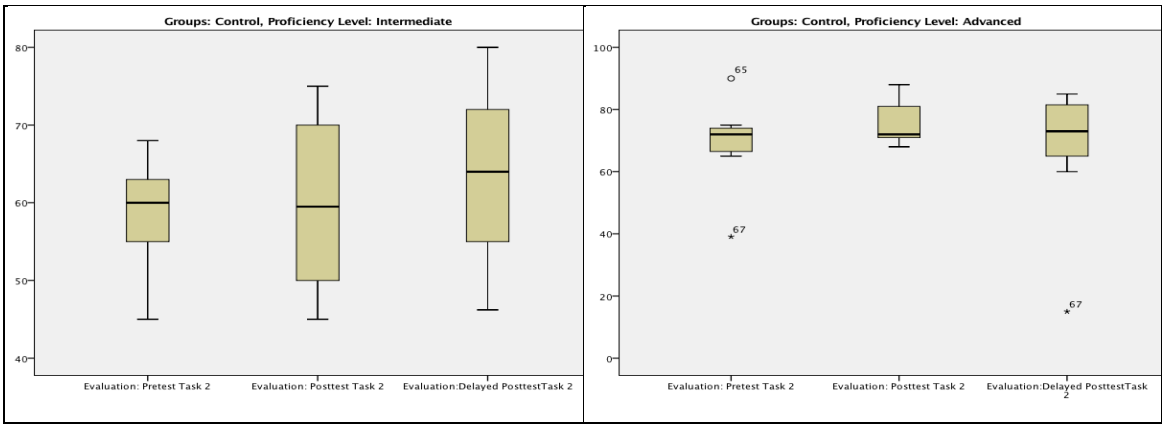
*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

Evaluation (task 2). The highest number of potential outlying scores (i.e., 17) was detected in the variables representing the evaluation that the six groups received for their essays. These scores, as shown Table E27 below, were distributed in all the groups except for the upper-intermediate control group. According to Table E28, four of these scores were actual outliers as their *z score* values exceeded -2.58. They were, therefore, replaced with the second lowest scores within the variable of each group.

Table E27

Boxplots: Evaluation (Task 2)





The evaluation score for case#2 pretest was changed from 5 to 50, the second lowest score in the pretest of the upper-intermediate focused instruction group. The evaluation for case#30 pretest (i.e., 30), in turn, was replaced with 40, the second lowest evaluation score in the advanced focused instruction group pretest. Case#10 and case#48 posttest evaluation scores (10 each) were respectively changed to 45 and 32.63, the second lowest scores within the upper-intermediate focused instruction group's and the advanced saliency group's posttest variables.

Table E28

Z Scores: Evaluation (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-3.10*(2)	-2.75*(10)	-1.59	-1.35	-1.64*(33)	-1.57*(33)	-1.88	-1.48	-1.59
	Max.	.78	1.46	1.82	1.09	1.04	1.04	1.42	1.46	1.61
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-3.03*(30)	-1.69	-2.26	-1.25	-2.65*(48)	-2.28*(50)	-1.94*(67)	-1.04	-2.12*(67)
	Max.	1.12	2.05	1.92	1.95	1.07*(45)	.85	1.38*(65)	1.55	.76

*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

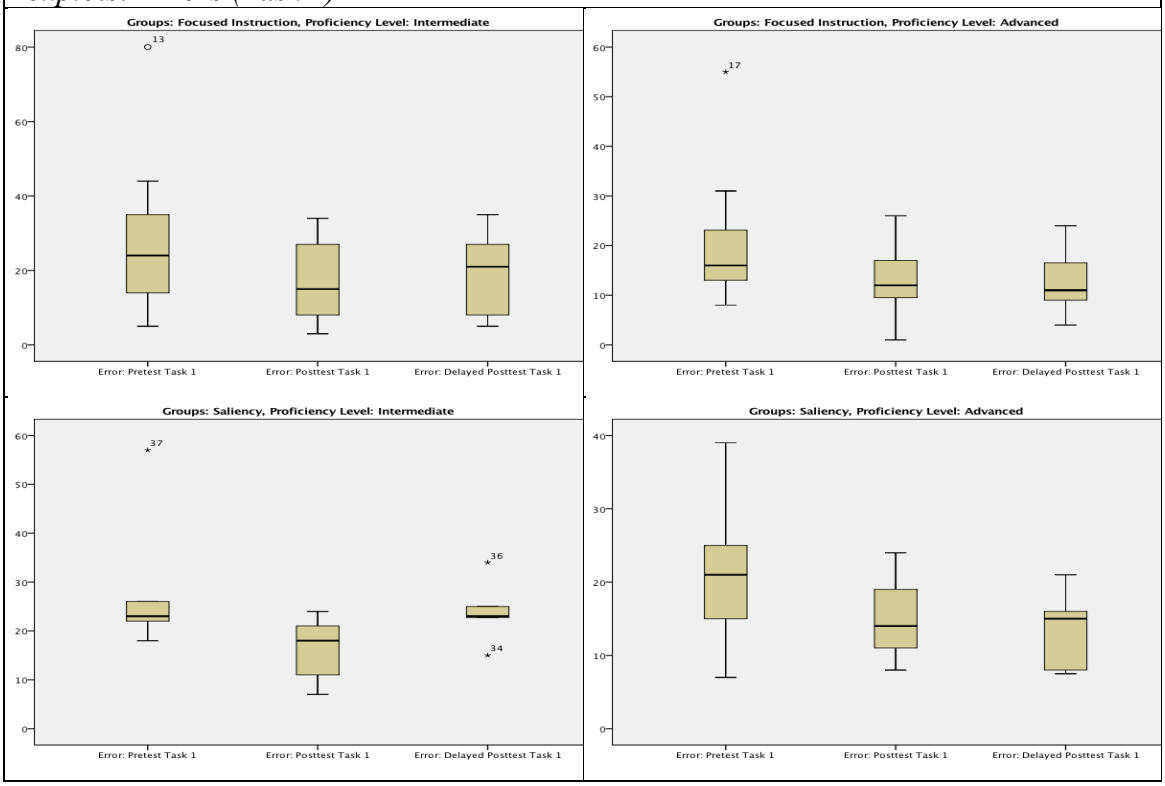
The *z scores* for the other 13 cases and all other cases, on the other hand, did not exceed the maximum cut-off set for this study, including the pretest *z score* for case#22 (*z score*=-2.11); the posttest *z scores* for case#38 (*z score*=.73), case#40 (*z score*=.63),

case#46 (*z score*=-.17), case#50 (*z score*=-1.53); and the delayed posttest *z score for* case#45 (*z score*=-2.07). Therefore, they were not altered.

Errors (task 1). Eight potential outliers were present in the variables that included the number of errors the six groups made in their graphical summaries at the three production stages (see Table E29 below). The upper-intermediate focused instruction group, the advanced focused instruction group, and the upper-intermediate control group had one potential outlying score each. The upper-intermediate saliency and the advanced control groups had three and two potential outliers, respectively.

Table E29

Boxplots: Errors (Task 1)



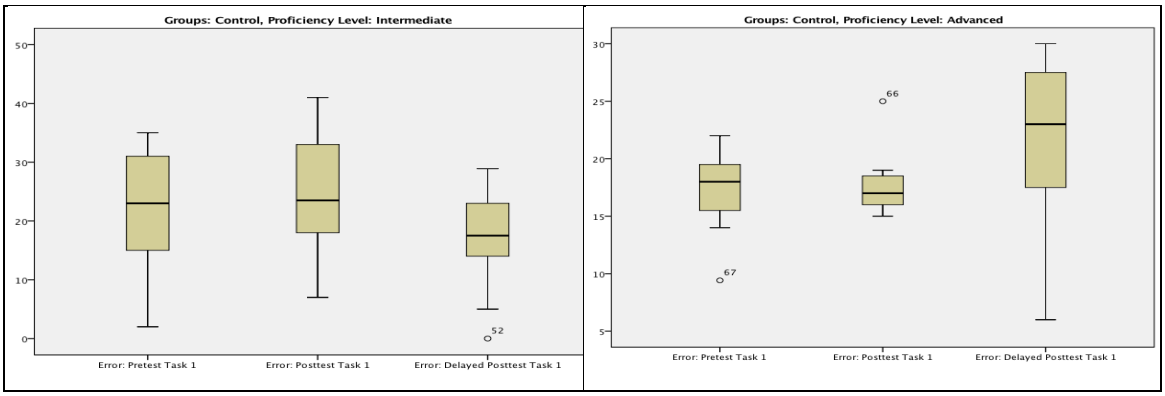


Table E30 below shows that only two of these scores were outliers that needed to be adjusted, namely, the pretest scores for case#13 and case#17. The outlying score for case#13 pretest (i.e., 80 errors) was replaced with the second highest number of errors made by the upper-intermediate focused instruction group in the pretest (i.e., 44), and the raw score for case#17 pretest (i.e., 55 errors) was replaced with 31, the second highest number of errors within the advanced focused instruction group's pretest variable.

Table E30

Z Scores: Errors (Task 1)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.14	-1.31	-1.25	-.71	-1.31	-1.32*(34)	-1.87	-1.67	-1.87*(52)
	Max.	2.66*(13)	1.42	1.50	1.76*(37)	1.11	1.48*(36)	1.20	1.40	1.34
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.02	-1.98	-1.48	-1.45	-1.41	-1.37	-1.82*(67)	-.88	-1.74
	Max.	3.35*(17)	1.95	1.86	1.92	1.70	1.64	1.18	2.05*(66)	.99

*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

The other six scores were not altered because their *z scores*, as shown in Table E30 above, were within the +/-2.58 range, so were the *z scores* for all other cases within these variables.

Errors% (task 1). According to Table E31 below, there were nine potential outliers in the variables representing the percentage of errors, with the highest number in the advanced focused instruction group (a total of four). Two of these potential outliers

(i.e., case#17 pretest and case#18 posttest) had *z scores* greater than +2.58, as shown in Table E32. Therefore, the percentages of errors for case#17 pretest (i.e., 36.91) and case#18 posttest (i.e., 27.65) were respectively changed to 31.31 and 18.54, the next highest percentage of errors within the pretest and posttest variables of the advanced focused instruction group. The *z score* for case#18 (*z score*=2.05) and all other cases did not exceed +/-2.58.

Table E31

Boxplots: Errors% (Task 1)

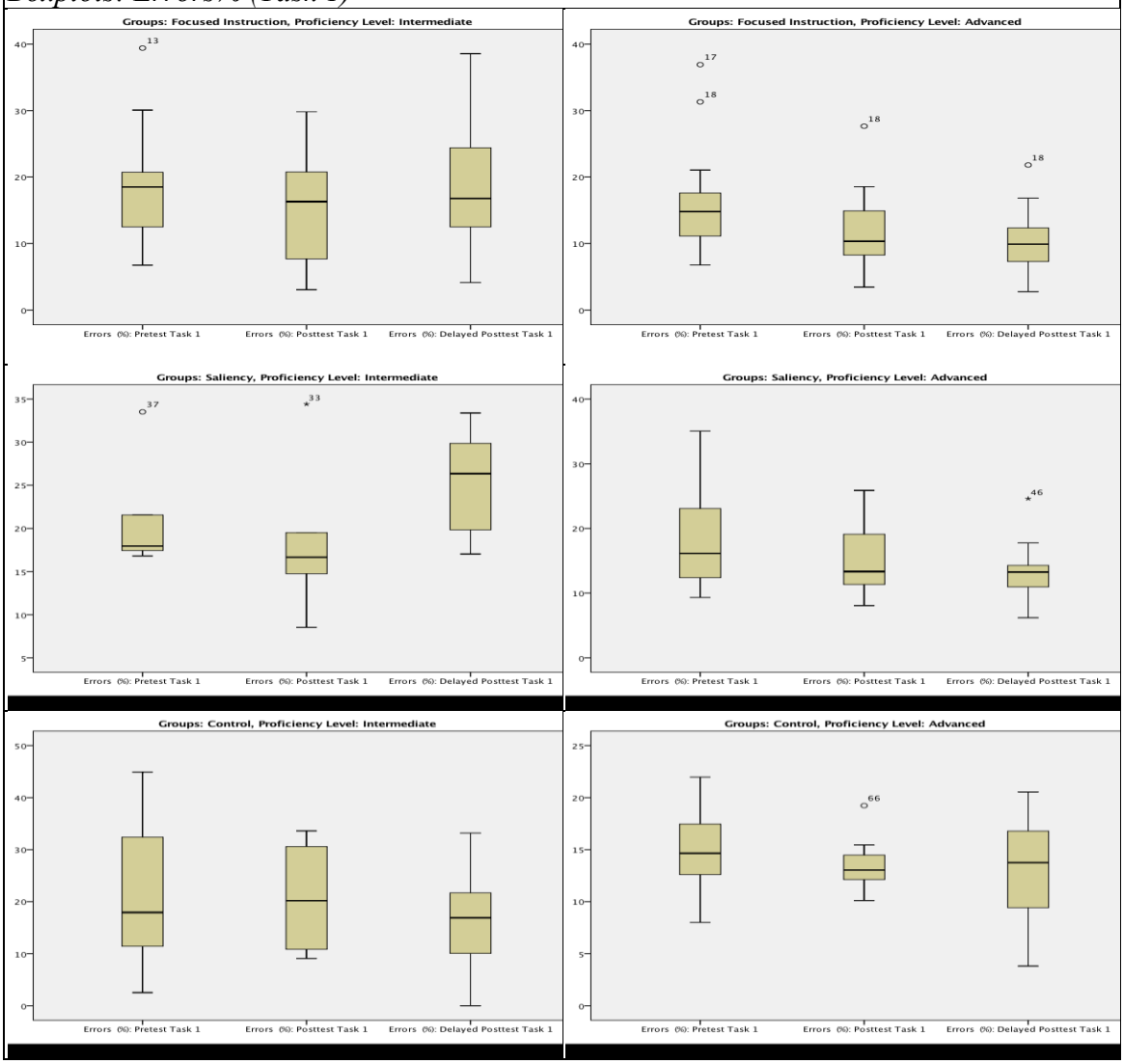


Table E32

Z Scores: Errors% (Task 1)

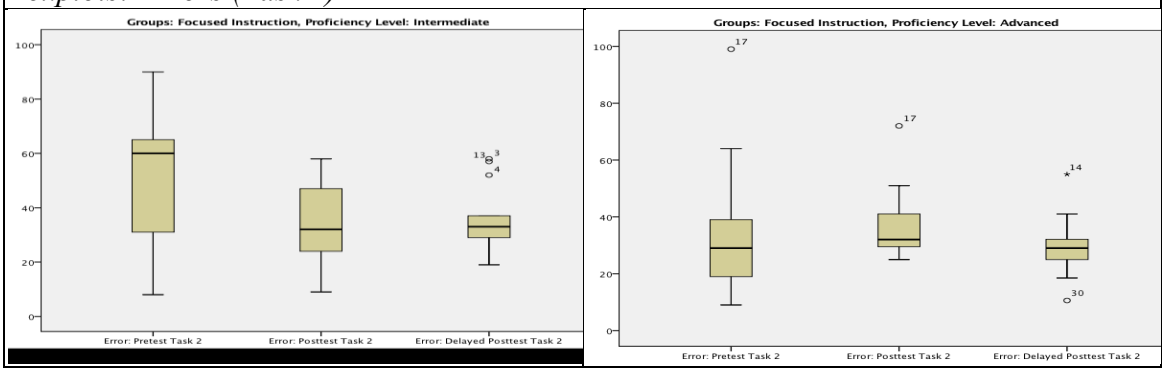
		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.28	-1.45	-1.54	-.66	-1.06	-1.21	-1.38	-1.20	-1.62
	Max.	2.17*(13)	1.84	2.12	1.73*(37)	1.62*(33)	1.19	1.70	1.22	1.73
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.25	-1.47	-1.60	-1.17	-1.19	-1.46	-1.51	-1.20	-1.53
	Max.	2.80*(17)	2.84*(18)	2.46*(18)	2.19	1.83	2.39*(46)	1.47	1.89*(66)	1.27

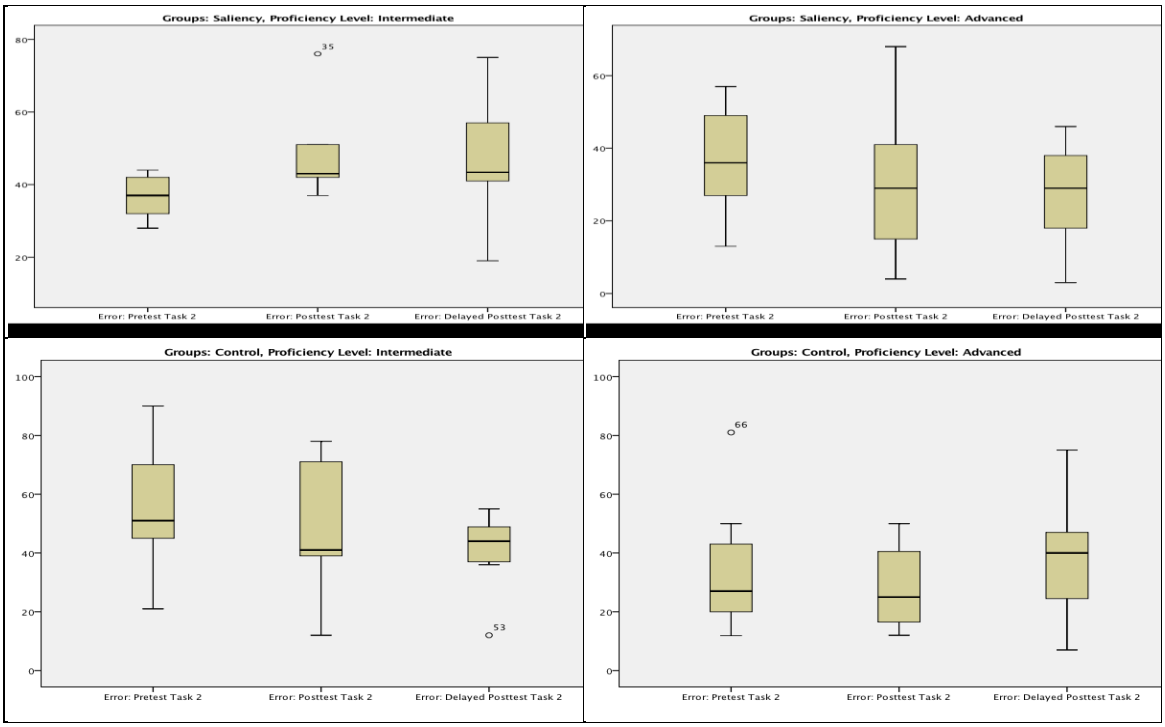
*(n) indicates the number and *z score* of an extreme case; green represents an actual outlier

Errors (task 2). Ten potential outlying scores were identified in Table E33 below, seven of which were in the upper-intermediate and advanced focused instruction groups. The other three scores were in the upper-intermediate saliency group, the upper-intermediate control group, and the advanced control group, one each. Of the 10 potential outlying scores, three were actual outliers, as shown in Table E34 below, and were, thus, replaced with the second highest number of errors within each target variable.

Table E33

Boxplots: Errors (Task 2)





The raw scores for case#17 pretest (i.e., 99) and posttest (i.e., 72) were respectively substituted with 64 and 51, the second highest number of errors in the pretest and posttest of the advanced focused instruction group. The raw score for case#14 delayed posttest (i.e., 55) was replaced with 41, the second highest number of errors within that variable. The *z scores* listed in Table E34 below for the remaining four cases did not exceed +/-2.58; nor did the *z scores* for case#4 (*z score*=1.30), case#13 (*z score*=1.69), or any other cases. Therefore, none of them was adjusted.

Table E34

Z Scores: Errors (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.68	-1.70	-1.28	-1.28	-.83	-1.36	-1.54	-1.74	-2.43*(53)
	Max.	1.46	1.52	1.77*(3)	1.11	1.69*(35)	1.35	1.76	1.42	1.16
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.16	-1.04	-1.96*(30)	-1.82	-1.40	-1.70	-.97	-1.11	-1.36
	Max.	3.15*(17)	3.08*(17)	2.73*(14)	1.49	2.14	1.44	1.92*(66)	1.42	1.64

*(n)indicates the number and *z score* of an extreme case; green represents an actual outlier

Errors% (task 2). According to table E35, there were seven potential outlying scores in the variables that included the percentage of errors in Task 2. These scores were distributed in all groups except for the upper-intermediate focused instruction group.

Table E35

Boxplots: Errors% (Task 2)

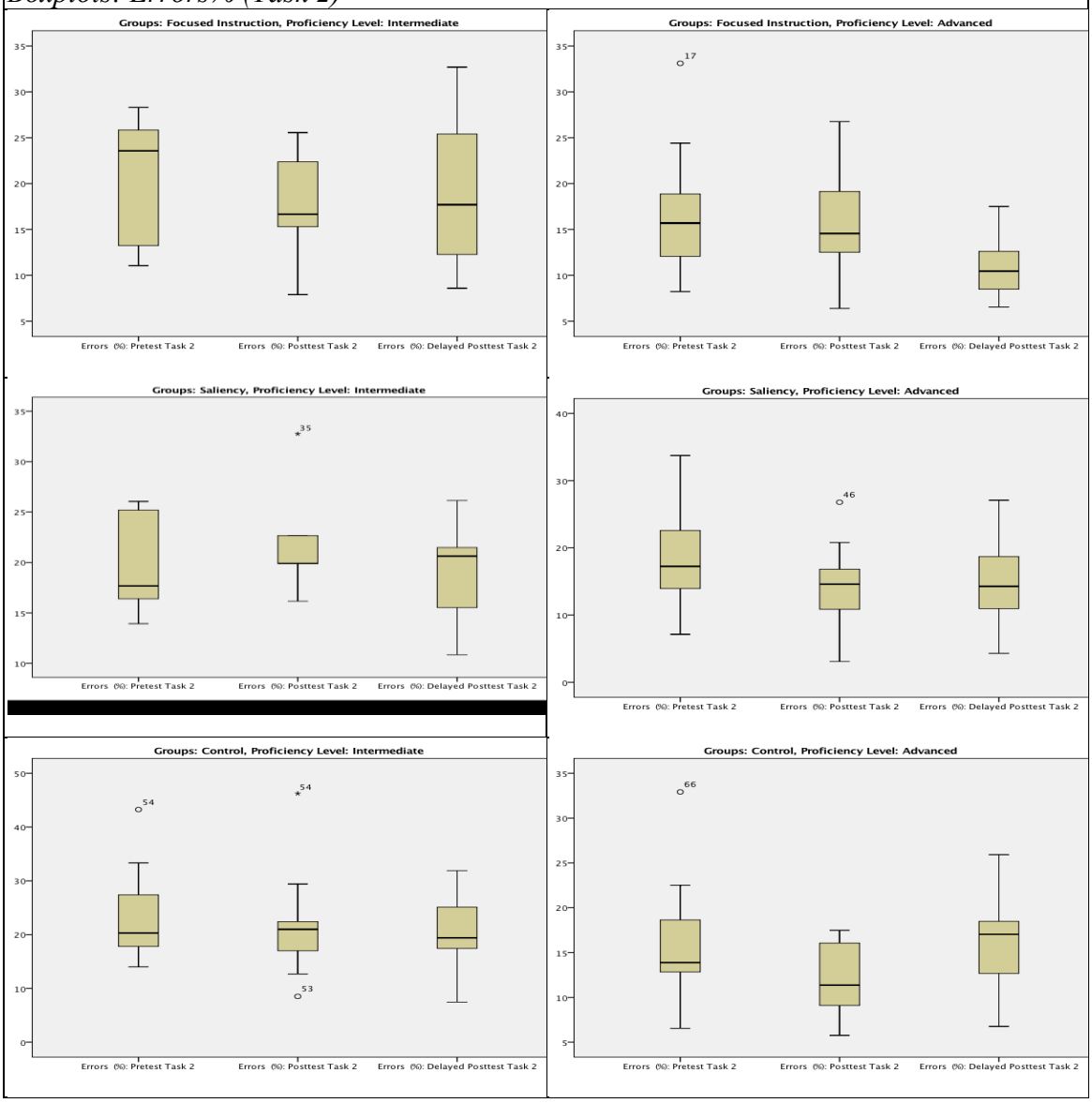


Table E36

Z Scores: Errors% (Task 2)

		Focused Instruction			Saliency			Control		
		T1	T2	T3	T1	T2	T3	T1	T2	T3

Inter.	N	13	13	13	5	5	5	10	10	10
	Min.	-1.52	-1.88	-1.38	-1.09	-.97	-1.37	-.97	-1.28*(53)	-1.84
	Max.	1.14	1.37	1.69	1.14	1.66*(35)	1.23	2.20*(54)	2.36*(54)	1.80
Adv.	N	19	19	19	13	13	13	7	7	7
	Min.	-1.27	-1.83	-1.44	-1.53	-1.78	-1.86	-1.17	-1.38	-1.52
	Max.	2.74*(17)	2.34	2.18	2.16	2.01*(46)	2.15	1.90*(66)	1.16	1.63
*(n)indicates the number and *z score* of an extreme case; green represents an actual outlier										

Almost all of these potential outlying scores, as demonstrated in Table E36 above, had *z scores* between -2.58 and +2.58 and, therefore, were not altered. The only score that exceeded the *z score* threshold was the percentage of errors made by case#17 in the pretest. The raw score for this case (i.e., 33.11) was, thus, replaced with the second highest percentage of errors within the pretest by the advanced focused instruction group (i.e., 24.42).

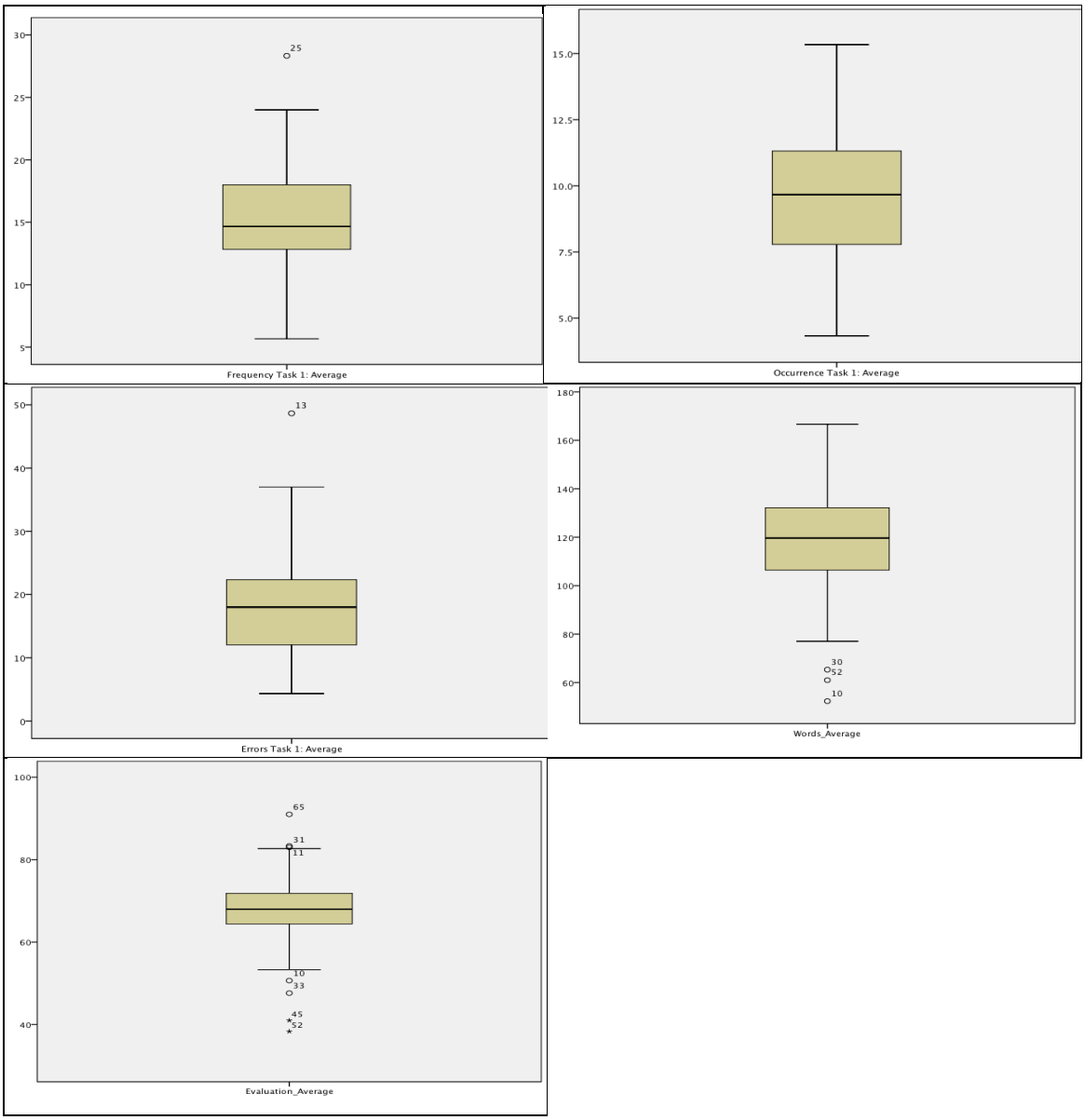
Outliers: Correlation Coefficients

Boxplots and *z scores* were generated for the five variables that represented the pretest, posttest, and delayed posttest average of the frequency and occurrence of the target formulaic sequences, the number of words and errors, and the EAP judge's evaluation in both writing tasks in order to identify any outlying scores. All cases whose *z scores* were greater than +/-2.58 were replaced with the next highest/lowest score that had a *z score* within +/-2.58 (Fidell & Tabachnick, 2003; Tabachnick & Fidell, 2007a).

According to Table E37, there were 12 potential outlying scores distributed in all variables except for the occurrence variable. An inspection of the *z scores* for these 12 cases and all other cases confirmed that five cases exceeded the +/-2.58 threshold.

Table E37

Boxplots: Correlation Variables (Task 1)



The *z scores* for case#10, case#25, case#13, and case#52, as shown in Table E38 below, were greater than +/-2.58. *The z score* for case#45 evaluation also exceeded -2.58 (*z score*=-2.78). These cases were, therefore, replaced with the next highest/lowest score within each variable. The scores for case#25 frequency (i.e., 28.33) and case#13 errors (i.e., 48.67) were respectively replaced with 24 and 37, the second highest scores within each variable. The evaluation scores for both case#45 (i.e., 41) and case#52 (i.e., 38.33) were replaced with the second lowest evaluation score within the evaluation variable (i.e.,

47.63). The average number of words generated by case#10 (i.e., 52.33) was replaced with the second lowest average score within the words variable (i.e., 61).

Table E38

Z Scores: Correlation Variables (Task 1)

	Frequency	Occurrence	Errors	Words	Evaluation
N	67	67	67	67	67
Min.	-2.06	-2.08	-1.71	-2.66*(10)	-3.06*(52)
Max.	2.90*(25)	2.35	3.62*(13)	2.03	2.44*(65)

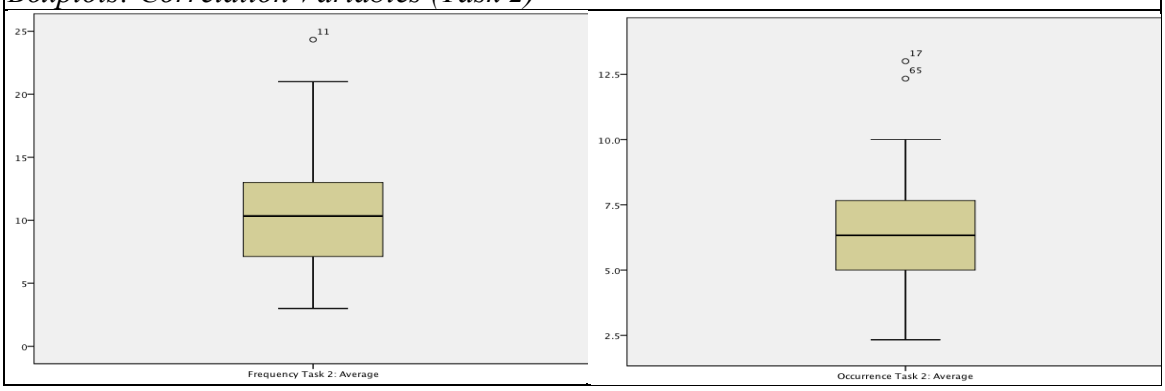
*(n)indicates the number and *z score* of an extreme case; green represents an actual outlier

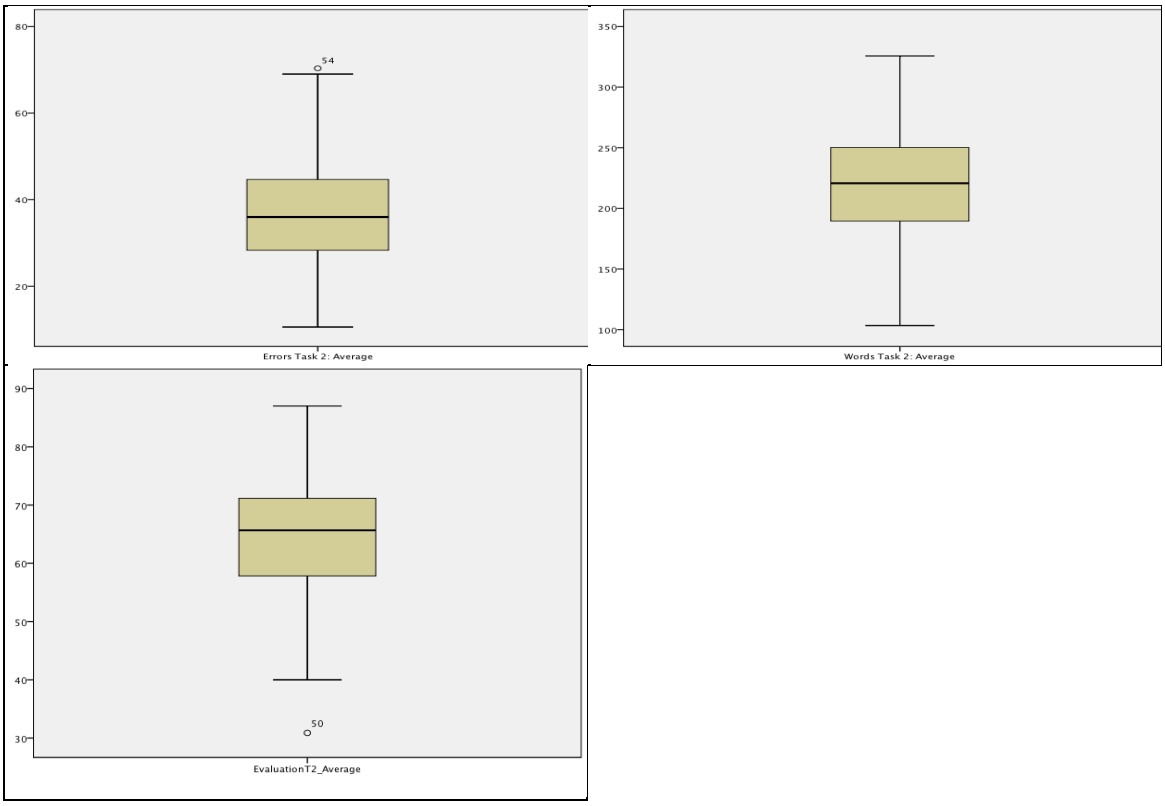
None of the other *z scores* in Table E38 above exceeded the threshold for this study (i.e., +/-2.58); nor did the *z scores* for the words generated by case#30 (*z score*=-2.13) and case#52 (*z score*=-2.31) or the evaluation scores for case#10 (*z score*=-1.77), case#11 (*z score*=-1.64), case#31 (*z score*=1.60), or case#33 (*z score*=-2.09).

Boxplots and *z scores* were also generated for the average scores of the frequency, occurrence, words, evaluation and errors in the second task. As can be seen in Table E39, five potential outliers were distributed in all variables except for words. An inspection of the *z scores* for these cases and all other cases revealed the presence of four outliers, three of which are shown in Table E40 below (i.e., case#11 frequency, case#17 occurrence, and case#50 evaluation).

Table E39

Boxplots: Correlation Variables (Task 2)





The fourth outlier was the occurrence of case#65 whose z score was above +2.58 (*z score*=2.71). The frequency of case#11 (i.e., 24.33) was replaced with 21, the second highest score within the frequency variable; the occurrence scores of both case#17 (i.e., 13) and case#65 (i.e., 12.33) were replaced with 10, the next highest score within the occurrence variable. The evaluation of case#50 (i.e., 30.88), in turn, was replaced with the second lowest evaluation score (i.e., 40). None of other scores had a *z score* that exceeded the threshold for this study, and, hence, none of them was adjusted.

Table E40

Z Scores: Correlation Variables (Task 2)

	Frequency	Occurrence	Errors	Words	Evaluation
N	67	67	67	67	67
Min.	-1.67	-1.87	-1.81	-2.41	-3.01*(50)
Max.	3.19*(11)	3.01*(17)	2.25*(54)	2.16	2.19

*(n)indicates the number and *z score* of an extreme case; green represents an actual outlier

Appendix F: Normality Tests

Normality Tests for Mixed Factorial ANOVA

Shapiro-Wilk normality tests and skewness and kurtosis *z scores* were calculated for the sum of the frequency, occurrence, correct, wrong, correct%, words, evaluation, errors, and error% scores within each of the six groups in both writing tasks in order to statistically assess their distribution. As explicated in the fourth chapter, approximately normally distributed data had a nonsignificant *Shapiro-Wilk* normality test results (p>.01) and skewness and kurtosis *z scores* that fell within +/-1.96 (Fidell & Tabachnick, 2003; Kim, 2013; Meyers et al., 2006; Tabachnick & Fidell, 2007a).

Frequency (task 1). The results of the normality tests computed on the sum of the frequency scores utilized by the upper-intermediate focused instruction group (M=41, SD=10.66, n=13), the advanced focused instruction group (M=53.93, SD=13.91, n=19), the upper-intermediate saliency group (M=39.40, SD=14.50, n=5), the advanced saliency group (M=45.43, SD=14.08, n=13), the upper-intermediate control group (M=34.02, SD=9.11, n=10), and the advanced control group (M=49.33, SD=8.22, n=7) confirmed that the distribution of the sum of these variables was approximately normal.

Table F1

Normality Tests: Frequency (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	.161	.278	-.172	.099	-.674	1.154
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.261	.531	-.188	.161	-.981	1.453
Kurtosis	-1.115	.158	-1.948	.788	-1.310	1.756
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.936	.156	-.974	.662	-.982	1.106
Shapiro-Wilk statistic	.959	.981	.922	.951	.859	.920
Shapiro-Wilk sig.	.743	.956	.542	.621	.075	.471

The approximately normal distribution of each dependent variable was inferred from the nonsignificant results of the *Shapiro-Wilk* normality tests (p>.01) and the skewness and kurtosis *z scores* that were within the +/-1.96 range for each summative variable (see Table F1 above).

Occurrence (task 1). The distribution of the sum of scores for the occurrence of the target formulaic sequences within each of the six groups was approximately normal. This was confirmed by the nonsignificant results of the *Shapiro-Wilk* normality test (p>.01) summarized in Table F2 below. It was also demonstrated by the skewness and kurtosis *z scores*, which were within the range of +/-1.96 for the upper-intermediate focused instruction group (M=28.31, SD=6.45, n=13), the advanced focused instruction group (M= 31.85, SD= 6.32, n=19), the upper-intermediate saliency group (M=26.40, SD=7.99, n=5), the advanced saliency group (M=29.66, SD=8.70, n=13), the upper-intermediate control group (M=21.29, SD=4.85, n=10), and the advanced control group (M=29.39, SD=7.09, n=7).

Table F2

Normality Tests: Occurrence (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	-.144	-.231	1.179	.177	-.822	1.195
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	-.234	-.441	1.291	.287	-1.197	1.505
Kurtosis	-.298	-.568	.946	.639	-.661	1.839
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.250	-.560	.473	.537	-.496	1.159
Shapiro-Wilk statistic	.984	.964	.908	.965	.889	.893
Shapiro-Wilk sig.	.994	.645	.457	.830	.166	.293

Correct (task 1). The distribution of the summative scores of correct formulaic sequences that were produced by the upper-intermediate focused instruction group (M=26, SD=12.10, n=13) and the advanced focused instruction group (M=38.60,

SD=11.87, *n*=19) was approximately normal, which was inferred from the nonsignificant *Shapiro-Wilk* test results (*p*>.01) and the skewness and kurtosis *z scores* for both groups (see Table F3 below). Table F3 also confirmed the approximately normal distribution of the sum of scores within the upper-intermediate saliency group (*M*=18.10, *SD*=9.02, *n*=5), the advanced saliency group (*M*=29.91, *SD*=9.35, *n*=13), the upper-intermediate control group (*M*=21.18, *SD*=10.02, *n*=10), and the advanced control group (*M*=33.83, *SD*=11.77, *n*=7) because of the nonsignificant *Shapiro-Wilk* test results (*p*>.01) as well as the skewness and kurtosis *z scores* that were all within the +/-1.96 range.

Table F3

Normality Tests: Correct (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	.317	.090	-.715	-.659	-.411	.733
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.515	.172	-.783	-1.070	-.598	.923
Kurtosis	-1.376	.168	-1.512	.234	-.968	-1.163
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-1.155	.166	-.756	.196	-.726	-.733
Shapiro-Wilk statistic	.923	.980	.908	.937	.931	.879
Shapiro-Wilk sig.	.277	.941	.457	.418	.461	.223

Wrong (task 1). The distribution of the combined numbers of incorrect formulaic sequences produced by the upper-intermediate focused instruction group (*M*=14.46, *SD*=5.44, *n*=13), the advanced focused instruction group (*M*=14.75, *SD*=7.28, *n*=19), the upper-intermediate saliency group (*M*=21.30, *SD*=6.04, *n*=5), the advanced saliency group (*M*=15.52, *SD*=6.89, *n*=13), the upper-intermediate control group (*M*=12.85, *SD*=6.12, *n*=10), and the advanced control group (*M*=15.50, *SD*=7.58, *n*=7) was approximately normal because of the skewness and kurtosis *z scores* that were within +/-1.96 and the nonsignificant *Shapiro-Wilk* test results (*p*>.01), as shown in Table F4 below.

Table F4

Normality Tests: Wrong (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	-.173	.985	.601	1.103	-.958	-1.333
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	-.281	1.880	.658	1.791	-1.394	-1.679
Kurtosis	-.741	1.279	-.578	.146	1.205	2.759
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.622	1.261	-.289	.123	.903	1.739
Shapiro-Wilk statistic	.951	.934	.952	.845	.906	.874
Shapiro-Wilk sig.	.617	.209	.750	.025	.255	.200

Correct% (task 1). The skewness and kurtosis *z scores* included in Table F5 below for the sum of the percentage of the correct formulaic sequences utilized by the upper-intermediate focused instruction group (*M*=183.39, *SD*=51.87, *n*=13), the advanced focused instruction group (*M*=214.49, *SD*= 41.32, *n*=19), the advanced saliency group (*M*=199.57, *SD*=25.52, *n*=13), the upper-intermediate control group (*M*=188.02, *SD*=71.72, *n*=10), and the advanced control group (*M*=204.53, *SD*=47.82, *n*=7) were all within the range of +/-1.96. Table F5 also shows that the *Shapiro-Wilk* test results for these variables were not significant ($p>.01$).

Table F5

Normality Tests: Correct% (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	-.028	-.851	2.086	-.307	-.573	1.273
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	-.045	-1.624	2.285	-.498	-.834	1.603
Kurtosis	-.193	-.210	4.457	-1.466	.637	1.866
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.162	-.207	2.229	-1.231	.478	1.176
Shapiro-Wilk statistic	.933	.892	.699	.916	.912	.905
Shapiro-Wilk sig.	.372	.035	.009*	.220	.365	.362

*$p<.01$

The distribution of the sum of scores within the upper-intermediate saliency group (*M*=142.30, *SD*=11.92, *n*=5), on the other hand, was slightly skewed and kurtotic, and it had a significant *Shapiro-Wilk* result. However, since the difference between the skewness and kurtosis *z scores* threshold used in this study (i.e., +/-1.96) and the skewness and kurtosis *z scores* for the upper-intermediate saliency group was very small (i.e., less than 0.5) and *mixed ANOVA* is robust to slight violations of normality (Tabachnick & Fidell, 2001, 2007a), the deviation was considered very minor and was not expected to affect the analysis.

Frequency (task 2). As shown in Table F6 below, the skewness and kurtosis *z scores* for the total number of the target formulaic sequences produced by the three upper-intermediate groups, that is, the focused instruction group (*M*=23.85, *SD*=10.23, *n*=13), the saliency group (*M*=38.58, *SD*=10.06, *n*=5), and the control group (*M*=22.91, *SD*=7.83, *n*=10) were within the +/-1.96 range; the same applies to the advanced focused instruction group (*M*=34.95, *SD*=11.61, *n*=19), the advanced saliency group (*M*=29.76, *SD*= 9.72, *n*=13), and the advanced control group (*M*=35.36, *SD*=14.36, *n*=7).

Table F6

Normality Tests: Frequency (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	.338	-.323	-.496	.035	.609	.307
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.549	-.616	-.543	.057	.886	.387
Kurtosis	-.470	-.431	.672	.379	-.390	-2.109
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	.395	-.425	.336	.318	-.292	-1.329
Shapiro-Wilk statistic	.967	.944	.984	.985	.919	.886
Shapiro-Wilk sig.	.862	.308	.955	.995	.352	.253

The skewness and kurtosis *z scores* for the six groups along with the results of the *Shapiro-Wilk* normality test ($p > .01$), presented in Table F6 above, confirmed that these summative variables were approximately normally distributed.

Occurrence (task 2). Similar results to those of the frequency of the target formulaic sequences were obtained for the sum of the distinct types the six groups utilized. That is, the *Shapiro-Wilk* normality test results for all the six groups were not significant ($p>.01$), as can be seen in Table F7 below. Moreover, the skewness and kurtosis *z scores* for the upper-intermediate focused instruction group (M=15.54, SD=6.04, n=13), the advanced focused instruction group (M=21.37, SD=5.98, n=19), the upper-intermediate saliency group (M=23.98, SD= 5.52, n=5), the advanced saliency group (M=19.15, SD=5.49, n=13), the upper-intermediate control group (M=16.32, SD=4.58, n=10), and the advanced control group (M=20.70, SD=8.13, n=7) were within the +/-1.96 range, which can be seen as evidence of the approximately normal distribution of these six summative variables.

Table F7

Normality Tests: Occurrence (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	.683	-.090	-.630	-.294	.999	1.451
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	1.109	-.172	-.690	-.477	1.454	1.827
Kurtosis	.675	-.100	-.076	-.863	.999	2.978
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	.567	-.099	-.038	-.725	.749	1.876
Shapiro-Wilk statistic	.950	.974	.962	.939	.917	.860
Shapiro-Wilk sig.	.601	.860	.825	.443	.336	.152

Correct (task 2). The distribution of the total number of correct formulaic sequences that were used by five of the six groups at the three production stages was approximately normal. This was inferred from the skewness and kurtosis *z scores*

summarized in Table F8 below that were within +/-1.96 for the advanced focused instruction group (*M*=21.69, *SD*=7.27, *n*=19), the upper-intermediate saliency group (*M*=22.20, *SD*=6.06, *n*=5), the advanced saliency group (*M*=18.65, *SD*=6.80, *n*=13), the upper-intermediate control group (*M*=12.37, *SD*=5.79, *n*=10), and the advanced control group (*M*=22.79, *SD*=12.23, *n*=7).

Table F8

Normality Tests: Correct (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	1.548	-.093	.579	.030	.540	1.519
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	2.513	-.177	.634	.049	.786	1.913
Kurtosis	2.697	-.633	.160	-.613	-.463	2.238
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	2.264	-.624	.08	-.515	-.347	1.410
Shapiro-Wilk statistic	.851	.959	.955	.971	.937	.816
Shapiro-Wilk sig.	.029	.559	.774	.911	.520	.059

The *Shapiro-Wilk* test results also confirmed the approximately normal distribution of these variables ($p>.01$). The sum of correct formulaic sequences within the upper-intermediate focused instruction group (*M*=15.69, *SD*=10.99, *n*=13), on the other hand, was slightly skewed and kurtotic. However, since the deviation from the threshold set for this study was very slight (i.e., less than 1) and the *Shapiro-Wilk* test result for the upper-intermediate focused instruction group was not significant ($p>.01$), the distribution of the sum of correct formulaic sequences within that group was not considered a major one and, thus, was not expected to affect the analysis.

Wrong (task 2). The results presented in Table F9 below for the sum of the number of incorrect formulaic sequences that the six groups produced indicated that the distribution of these summative variables was approximately normal. This was inferred from the nonsignificant *Shapiro-Wilk* test results ($p>.01$). The skewness and kurtosis *z*

scores for the advanced focused instruction group (*M*=13.05, *SD*=6.91, *n*=19), the upper-intermediate saliency group (*M*= 16.38, *SD*=5.41, *n*=5), the advanced saliency group (*M*=11.11, *SD*=4.16, *n*=13), the upper-intermediate control group (*M*=10.55, *SD*= 4.59, *n*=10), and the advanced control group (*M*=12.57, *SD*= 6.68, *n*=7) were also within the +/-1.96 range, as shown in Table F9 below.

Table F9

Normality Tests: Wrong (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	1.348	.991	-.232	-.742	.114	.192
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	2.188	1.891	-.254	-1.205	.166	.242
Kurtosis	2.325	1.189	-.310	1.437	-1.268	-1.521
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	1.952	1.173	-.155	1.207	-.951	-.958
Shapiro-Wilk statistic	.890	.934	.959	.953	.942	.954
Shapiro-Wilk sig.	.097	.209	.800	.648	.574	.766

The distribution of the sum of incorrect formulaic sequences that the upper-intermediate focused instruction group produced (*M*=9.69, *SD*=4.15, *n*=13) was slightly positively skewed (Skewness=1.348, *SE*=.616, *z score*=2.188). Since the *Shapiro-Wilk* test result for that group was not significant (*p*>.01) and this variable exceeded the skewness threshold only slightly (i.e., less than one 0.5), the slight departed from normality was not expected to affect the analysis.

Correct% (task 2). According to Table F10 below, the distribution of the sum of the percentage of correct formulaic sequences utilized by the six groups at the three production stages was approximately normal. This was first and foremost inferred from the nonsignificant *Shapiro-Wilk* test results (*p*>.01). The skewness and kurtosis *z scores* for the summative variables of the upper-intermediate focused instruction group (*M*=175.25, *SD*=32.04, *n*=13), the advanced focused instruction group (*M*=192.31,

SD=37.13, *n*=19), the upper-intermediate saliency group (*M*=172.03, *SD*=26.17, *n*=5), the advanced saliency group (*M*=181.65, *SD*= 33.56, *n*=13), the upper-intermediate control group (*M*=161.56, *SD*= 44.12, *n*=10), and the advanced control group (*M*= 196.30, *SD*=54.62, *n*=7) also confirmed the approximately normal distribution as they were all within the +/-1.96 range.

Table F10

Normality Tests: Correct% (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	-.086	.080	-1.288	-1.128	-1.176	.439
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	-.140	.153	-1.411	-1.831	-1.712	.553
Kurtosis	-.709	-.008	1.576	.982	.389	-2.114
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.595	-.008	.788	.825	.292	-1.332
Shapiro-Wilk statistic	.981	.959	.905	.907	.802	.841
Shapiro-Wilk sig.	.983	.552	.439	.167	.015	.101

Words (task 1). According to Table F11, the distribution of the sum of words that were generated at the three production stages by the advanced focused instruction group (*M*=363.55, *SD*=60.86, *n*=19), the advanced saliency group (*M*=323.80, *SD*= 51.99, *n*=13), the upper-intermediate control group (*M*= 336.67, *SD*= 97.99, *n*=10), and the advanced control group (*M*=419.48, *SD*=42.12, *n*=7) was approximately normal, which was confirmed by the skewness and kurtosis *z scores* which were within +/-1.96 as well as the nonsignificant results of the *Shapiro-Wilk* normality test (*p*>.01).

Table F11

Normality Tests: Words (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	-1.442	-.249	-2.025	-.248	.040	-.158
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	-2.341	-.475	-2.218	-.403	.058	-.199

Kurtosis	1.763	-.285	4.280	-1.508	-.615	-1.218
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	1.480	-.281	2.14	-1.266	-.461	.767
Shapiro-Wilk statistic	.856	.969	.737	.911	.981	.945
Shapiro-Wilk sig.	.034	.759	.023	.187	.970	.683

The distribution of the sum of words for the upper-intermediate focused instruction group and the upper-intermediate saliency group slightly departed from normality; that is, while the sum of words generated by the focused instruction group (*M*=353.54, *SD*=81.58, *n*=13) was negatively skewed, the sum of words produced by the saliency group (*M*=319.63, *SD*=50.43, *n*=5) was negatively skewed and kurtotic (see Table F11 above). However, they were still considered approximately normally distributed because of the slightly elevated *z scores* (less than 0.5) and the nonsignificant *Shapiro-Wilk* test results (*p*>.01).

Evaluation (task 1). Based on the results summarized in Table F12 below, the skewness and kurtosis *z scores* for the sum of the evaluation of the upper-intermediate focused instruction group (*M*=200.69, *SD*=22.64, *n*=13), the advanced focused instruction group (*M*=213.30, *SD*=20.26, *n*=19), the upper-intermediate saliency group (*M*=165.98, *SD*=16.53, *n*=5), the advanced saliency group (*M*= 205, *SD*= 21.44, *n*=13), the upper-intermediate control group (*M*= 189.80, *SD*= 33.23, *n*=10), and the advanced control group (*M*=230.93, *SD*=21.71, *n*=7) were all within the +/-1.96 range, with the exception of the upper-intermediate control group whose skewness *z score* was slightly elevated (i.e., less than 0.5). The *Shapiro-Wilk* results for the six groups were also not significant (*p*>.01).

Table F12

Normality Tests: Evaluation (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.

Skewness	.037	.395	-.231	-.612	-1.402	1.367
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.060	.754	-.253	-.994	-2.040	1.722
Kurtosis	2.308	-.458	.159	.130	1.866	1.820
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	1.938	-.452	.080	.109	1.399	1.147
Shapiro-Wilk statistic	.940	.929	.992	.939	.860	.867
Shapiro-Wilk sig.	.454	.164	.988	.447	.077	.173

The distribution of these variables was, therefore, considered approximately normal despite the slight negative skewness within the upper-intermediate control group.

Words (task 2). The nonsignificant results of the *Shapiro-Wilk* normality tests that were computed on the sum of words within each of the six groups (see Table F13 below) suggested that these variables were approximately normally distributed. The approximately normal distribution was also confirmed by the skewness and kurtosis *z scores* for the upper-intermediate focused instruction group (*M*=631, *SD*=128.93, *n*=13), the advanced focused instruction group (*M*=713, *SD*= 102.77, *n*=19), the upper-intermediate saliency group (*M*=662.25, *SD*=111.52, *n*=5), the advanced saliency group (*M*=599.68, *SD*=164.42, *n*=13), the upper-intermediate control group (*M*=676.72, *SD*=168.52, *n*=10), and the advanced control group (*M*=660.46, *SD*=194.02, *n*=7), which were all within +/-1.96, see Table F15 below.

Table F13

Normality Tests: Words (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	-.317	-.875	1.082	-.041	.641	-.153
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	-.515	-1.670	1.185	-.067	.933	-.193
Kurtosis	-.409	.305	.904	-.195	-.669	1.994
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.343	.301	.452	-.164	-.501	1.256
Shapiro-Wilk statistic	.926	.921	.894	.979	.927	.940
Shapiro-Wilk sig.	.305	.119	.379	.977	.415	.637

Evaluation (task 2). The results of the normality tests that were run on the sum of the evaluation that the six groups received at the three production stages and summarized in Table F14 below confirmed the normal distribution of these summative variables. This was inferred from the nonsignificant results of the *Shapiro-Wilk* normality tests (p>.01) for the six groups and the skewness and kurtosis *z scores* for the upper-intermediate focused instruction group (M=179.85, SD=24.29, n=13), the advanced focused instruction group (M=198.57, SD= 19.69, n=19), the upper-intermediate saliency group (M=190.90, SD=22.94, n=5), and the upper-intermediate control group (M=181.20, SD=24.48, n=10), which were all within +/-1.96, as shown in Table F16 below.

Table F14

Normality Tests: Evaluation (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	.327	-.340	-1.170	-1.353	-.209	-1.508
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.531	-.649	-1.281	-2.196	-.304	-1.899
Kurtosis	-.774	-.870	.644	2.253	-.987	3.119
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.650	-.858	.322	1.892	-.740	1.965
Shapiro-Wilk statistic	.934	.958	.897	.893	.963	.870
Shapiro-Wilk sig.	.385	.530	.391	.106	.819	.186

The skewness *z score* for the advanced saliency group (M=191.52, SD=39.53, n=13) and the kurtosis *z score* for the advanced control group (M=211.44, SD=44.34, n=7), in turn, minimally exceeded +/-1.96 (see Table F14 above). However, the violation to normality assumption was very minor and was not expected to affect the analysis because of the nonsignificant *Shapiro-Wilk* results (p>.01) and the robustness of *mixed ANOVAs* in case of slight departures from normality (Tabachnick & Fidell, 2001, 2007a).

Errors (task 1). The distribution of the sum of errors made by the six groups was approximately normal. This was inferred from the skewness and kurtosis *z scores* for the

upper-intermediate focused instruction group (*M*=61.23, *SD*=33.37, *n*=13), the advanced focused instruction group (*M*=44.21, *SD*=16.60, *n*=19), the upper-intermediate saliency group (*M*=69.35, *SD*=11.30, *n*=5), the advanced saliency group (*M*=49.68, *SD*=14.90, *n*=13), the upper-intermediate control group (*M*=64.46, *SD*=27.20, *n*=10), and the advanced control group (*M*=56.35, *SD*=12.97, *n*=7), which were all within +/- 1.96 (see Table F15 below). The nonsignificant *Shapiro-Wilk* test results (*p*>.01) also confirmed that the sum of errors within each group was approximately normally distributed.

Table F15

Normality Tests: Errors (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	.076	.894	.884	-.179	-.169	-1.074
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.123	1.706	.968	-.291	-.246	-1.353
Kurtosis	-1.283	.337	1.884	-1.650	.031	.317
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-1.077	.332	.942	-1.385	.023	.200
Shapiro-Wilk statistic	.939	.914	.936	.901	.962	.877
Shapiro-Wilk sig.	.441	.088	.639	.138	.813	.213

Errors% (task 1). The skewness and kurtosis *z scores* for the sum of the percentage of errors within the upper-intermediate focused instruction group (*M*=52.29, *SD*=25.98, *n*=13), the upper-intermediate saliency group (*M*=65.53, *SD*=15.59, *n*=5), the upper-intermediate control group (*M*=58.83, *SD*=28.75, *n*=10), and the advanced control group (*M*=41.51, *SD*= 12.01, *n*=7) were all within +/-1.96, and the *Shapiro-Wilk* test results for these variables were not significant (*p*>.01), as shown in Table F16 below. These results confirmed that the distribution of these summative scores was approximately normal.

Table F16

Normality Tests: Errors% (Task 1)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	.442	1.147	.982	1.249	.086	-.566
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.718	2.189	1.076	2.028	.125	-.713
Kurtosis	.295	.873	.368	1.384	-1.394	-1.179
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	.248	.861	.184	1.162	-1.045	-.743
Shapiro-Wilk statistic	.952	.897	.935	.894	.941	.918
Shapiro-Wilk sig.	.625	.043	.628	.111	.559	.451

The distribution of the sum of errors within advanced saliency group (*M*=46.55, *SD*=13.42, *n*=13) and the advanced focused instruction group (*M*=37.24, *SD*=13.80, *n*=19) was slightly positively skewed. The slight positive skewness was not expected to affect the analysis because of the nonsignificant *Shapiro-Wilk* test results (*p*>.01), as shown in Table F9 above, and the robustness of *Mixed ANOVA* to slight departures from normality (Tabachnick & Fidell, 2001, 2007a).

Errors (task 2). The distribution of the sum of errors made in the essay responses by the upper-intermediate focused instruction group (*M*=122.23, *SD*=45.71, *n*=13), the advanced focused instruction group (*M*=95.46, *SD*=24.52, *n*=19), the upper-intermediate saliency group (*M*=133.48, *SD*=36.12, *n*=5), the advanced saliency group (*M*=92.82, *SD*=33.77, *n*=13), the upper-intermediate control group (*M*=142.73, *SD*= 48.11, *n*=10), and the advanced control group (*M*=101.69, *SD*=60.15, *n*=7) was approximately normal. This was confirmed by both the skewness and kurtosis *z scores* for the six summative variables that were all within +/-1.96 and the nonsignificant *Shapiro-Wilk* normality test results (*p*>.01), as can be seen in Table F17.

Table F17

Normality Tests: Errors (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.

Skewness	.302	.243	1.488	-.484	-.669	.665
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	.490	.464	1.630	-.786	-.974	.838
Kurtosis	-1.182	-.037	2.068	-.926	.637	.226
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.992	-.036	1.034	-.777	.478	.142
Shapiro-Wilk statistic	.937	.967	.835	.902	.957	.946
Shapiro-Wilk sig.	.419	.724	.153	.143	.752	.689

Errors% (task 2). An inspection of the *Shapiro-Wilk* test results and the skewness and kurtosis *z scores* for the sum of the percentage of errors the six groups had at the three production stages indicated that these variables were approximately normally distributed. In other words, the skewness and kurtosis *z scores* presented in Table F18 below were between +1.96 and -1.96 for the upper-intermediate focused instruction group (*M*=58.43, *SD*=17.64, *n*=13), the advanced focused instruction group (*M*=41.90, *SD*=11.80, *n*=19), the upper-intermediate saliency group (*M*=61.05, *SD*=13.18, *n*=5), the advanced saliency group (*M*=47.30, *SD*=16.71, *n*=13), the upper-intermediate control group (*M*=64.56, *SD*=24.52, *n*=10), and the advanced control group (*M*=44.75, *SD*=17.47, *n*=7). Moreover, the *Shapiro-Wilk* test results for these six summative variables were not statistically significant ($p>.01$), which again confirmed that the sum of these scores within each group was approximately normally distributed.

Table F18

Normality Tests: Errors% (Task 2)

	Focused Instruction		Saliency		Control	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Skewness	-.500	.342	-.503	.978	.983	.039
Skewness SE	.616	.524	.913	.616	.687	.794
Skewness z score	-.812	.653	-.551	1.588	1.431	.049
Kurtosis	-1.107	-.232	-2.699	2.267	1.162	-.892
Kurtosis SE	1.191	1.014	2.00	1.191	1.334	1.587
Kurtosis z score	-.929	-.229	-1.350	1.903	.871	-.562
Shapiro-Wilk statistic	.896	.982	.872	.921	.916	.947
Shapiro-Wilk sig.	.119	.960	.276	.258	.327	.703

Normality Tests for Correlation Coefficients

Skewness and kurtosis *z scores* were calculated for and *Shapiro-Wilk* normality tests were computed on the dependent variables that represented Task 1 and Task 2 average scores for the words generated, errors made, and evaluation received by the 67 EAP students. The skewness and kurtosis *z scores* for the words (*M*=117.35, *SD*=24.03, *n*=67), errors (*M*=18.37, *SD*=7.78, *n*=67), and evaluation (*M*=67.88, *SD*=8.96, *n*=67) in Task 1 (i.e., summaries) fell within the +/-1.96 range; the results of the *Shapiro-Wilk* normality tests were also nonsignificant (*p*>.01) as shown in Table F19 below. These results demonstrated that the distribution of scores within each of the above-mentioned dependent variables was approximately normal.

Table F19

Normality Tests for Correlation Coefficients

	Words		Errors		Evaluation	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Skewness	-.457	-.137	.495	.456	-.227	-.516
Skewness SE	.293	.293	.293	.293	.293	.293
Skewness z score	-1.560	-.468	1.689	1.556	-.775	-1.761
Kurtosis	-.128	.022	-.139	-.204	.301	-.065
Kurtosis SE	.578	.578	.578	.578	.578	.578
Kurtosis z score	.221	.038	-.240	-.353	.521	-.112
Shapiro-Wilk statistic	.971	.989	.969	.964	.975	.963
Shapiro-Wilk sig.	.124	.802	.097	.052	.194	.044

Similar to the distribution of the three dependent variables for the first task, the distribution of the average scores for the words (*M*=220.65, *SD*=48.67, *n*=67), errors (*M*=37.24, *SD*=14.74, *n*=67), and evaluation (*M*=63.49, *SD*=10.44, *n*=67) in the second task was also approximately normal, which was confirmed by the skewness and kurtosis *z scores* that were within +/-1.96 and the nonsignificant *Shapiro-Wilk* test results (*p*>.01), as shown in Table F19 above.

Appendix G: Scatterplots for Correlation Coefficients

Several scatterplots were generated to assess the nature of the relationship between different pairs of dependent and independent variables prior to running bivariate correlations. Based on Figure G1, Figure G2, and Figure G3 below, there seemed to be a linear relationship between the frequency of the target formulaic sequences and the number of words generated in, the evaluation received for, and the number of errors made in the summary tasks. Similar patterns were identified between the three dependent variables and the occurrence variable, as shown in Figures G4, G5, and G6.

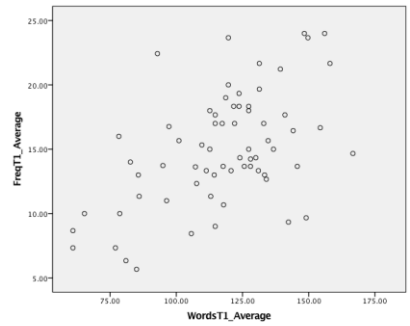


Figure G1. Task 1 (fre&wor)

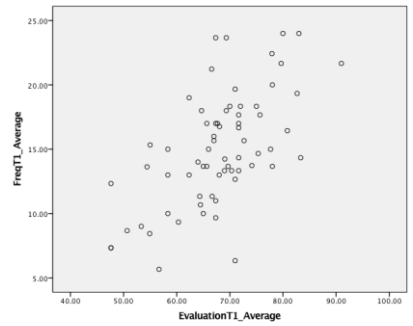


Figure G2. Task 1 (fre&eva)

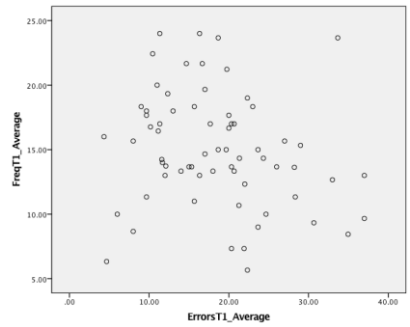


Figure G3. Task 1 (fre& err)

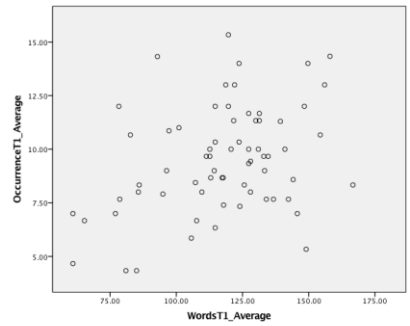


Figure G4. Task 1 (occ&wor)

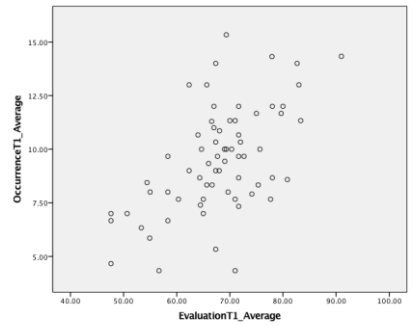


Figure G5. Task 1(occ& eva)

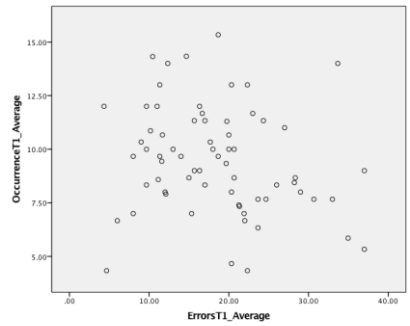


Figure G6. Task 1 (occ&err)

The scatterplots generated to visually assess the relationship between the frequency and occurrence of the target formulaic sequences as independent variables and words and evaluation as dependent variables in the second task suggested linear relationships between each pair, as can be seen in Figures, G7, G8, G10, and G11 below. However,

there seemed to be a very weak, if any, linear relationship between the frequency and error variables (see Figure G9) and the occurrence and errors variables (see Figure G12).

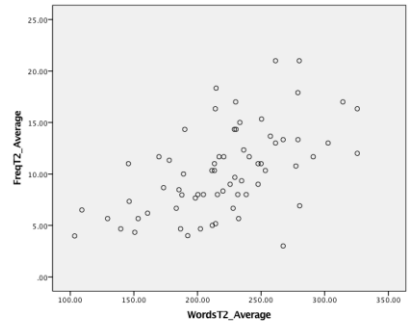


Figure G7. Task 2 (fre&wor)

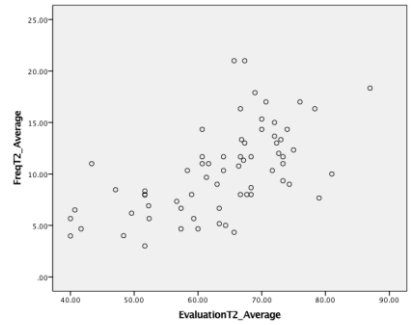


Figure G8. Task 2 (fre&eva)

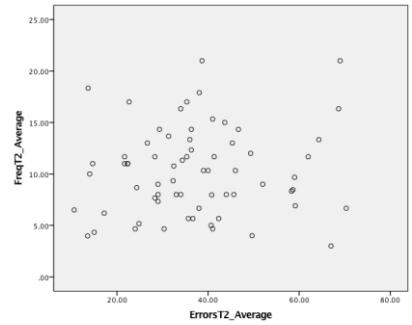


Figure G9. Task 2 (fre&err)

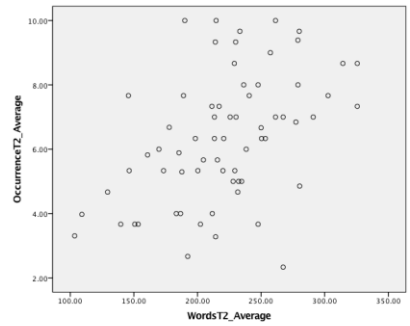


Figure G10. Task 2 (occ&wor)

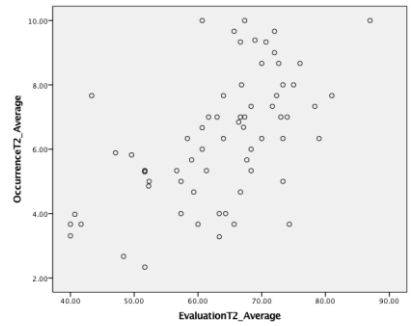


Figure G11. Task 2 (occ&eva)

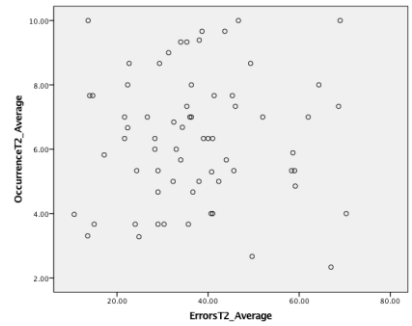


Figure G12. Task 2 (occ&err)

It is worth mentioning here that a few outlying scores were also detected in some scatterplots, but these scores were not expected to affect the results of the *Pearson correlation coefficients* because prior to running the scatterplots all cases whose *z scores* exceeded +/-2.58 were altered (see Tables E37, E38, E39, and E40).

Appendix H: Results Summary

Table H1

Summary of Within-Groups Comparisons Results: Task 1

	Focused Instruction		**Saliency**		**Control**	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Frequency	Ø	T3>T2>T1	Ø	T2>T1	Ø	Ø
Occurrence	T2>T1&T3	T2&T3>T1	T2&T3>T1	T2>T1	Ø	Ø
Correct	T2>T3	T3>T2>T1	Ø	T2>T1	Ø	Ø
Wrong	Ø	Ø	Ø	Ø	Ø	Ø
Correct%	Ø	T3>T1	Ø	Ø	Ø	Ø
Words	T2&T3<T1	Ø	T2<T1	Ø	Ø	T3>T2&T1
Evaluation	Ø	Ø	Ø	Ø	Ø	Ø
Errors	Ø	T2&T3<T1	T2<T1	T2&T3<T1	T3<T2	Ø
Errors%	All groups combined T2<T1					
r: FS & Words	Frequency & Words=27% (+)			Occurrence & Words=12% (+)		
r: FS & Eva.	Frequency & Evaluation=38% (+)			Occurrence & Evaluation=32% (+)		
r: FS & Err.	Frequency & Errors=Ø			Occurrence & Errors=Ø		

Note. T1, T2, T3 respectively refer to pretest, posttest, and delayed posttest
Ø indicates absence of any statistically significant differences
Green indicates prolonged learning gains (i.e., retention); yellow refers to immediate learning gains; red suggests drop in the students' performance

Table H2

Summary of Within-Groups Comparisons Results: Task 2

	Focused Instruction		**Saliency**		**Control**	
	Inter.	Adv.	Inter.	Adv.	Inter.	Adv.
Frequency	Ø	T2>T1	Ø	Ø	Ø	Ø
Occurrence	Ø	T2&T3>T1	Ø	Ø	Ø	Ø
Correct	All groups combined T2 & T3>T1					
Wrong	Ø	Ø	Ø	Ø	Ø	Ø
Correct%	All groups combined T3>T1					
Words	T3<T1	T3>T1	Ø	Ø	Ø	Ø
Evaluation	Ø	Ø	Ø	Ø	Ø	Ø
Errors	T2&T3<T1	Ø	Ø	Ø	Ø	Ø
Errors%	Ø	T3<T1&T2	Ø	T2&T3<T1	Ø	T3>T2
r: FS & Wor.	Frequency & Words=30%			Occurrence & Words=21%		
r: FS & Eva.	Frequency & Evaluation=35%			Occurrence & Evaluation=32%		
r: FS & Err.	Frequency & Errors=Ø			Occurrence & Errors=Ø		

Note. T1, T2, T3 respectively refer to pretest, posttest, and delayed posttest
Ø indicates absence of any statistically significant differences

Green indicates prolonged learning gains (i.e., retention); yellow refers to immediate learning gains; red suggests drop in the students' performance

Table H3

Summary of Between-Groups Comparisons Results: Task 1& Task 2

Group	Variable & Task
Focused Instruction	Frequency Task 1: Adv. FI >Adv. Sa. (T3) Occurrence Task 1: Inter. FI > Inter. Con. (overall & T2) Correct Task 1: Adv. FI >Adv. Sa. (T3) Words Task 1: Adv. FI <Adv. Con. (T3) Words Task 2: Adv. FI > Adv. Sa. (T3) Errors Task 1: Adv. FI <Adv. Con. (T3)
Saliency	Frequency & Occurrence Task 2: Inter Sa>Inter. FI & Inter. Con. (overall) Occurrence Task 2: Inter Sal. T3> Inter. FI T3 Words Task 1: Adv. Sa. < Adv. Con. (overall, T2, & T3) Words Task 1: Adv. Sa. < Adv. FI. (T3) Errors Task 1: Adv. Sa. < Adv. Con. (T3)
Control	Ø
Proficiency	Task 1: Freq., Occur., Corr., Corr. %, Eval., & Error% Task 2: Corr. (%), Eval., Error, & Error%

Note. Green indicates that an experimental group outperformed another experimental group; red indicates that an experimental group outperformed a control group; blue indicates that an experimental group outperformed both experimental and control groups.