Université de Montréal

The effectiveness of written corrective feedback on French as a second language accuracy

par Fatma Bouhlal

Département de didactique

Faculté des sciences de l'éducation

Thèse présentée
en vue de l'obtention du grade de Philosophiæ doctor (Ph. D.)

en éducation

option didactique

Juin 2019

Université de Montréal

Faculté des sciences de l'éducation

Thèse intitulée:

The effectiveness of written corrective feedback on French as a second language accuracy

Présentée par

Fatma Bouhlal

A été évaluée par un jury composé des personnes suivantes

Mme Marie-Claude Boivin

Présidente-rapporteuse

Mme Ahlem Ammar

Directrice de recherche

Mme Antonella Valeo

Membre du jury

Mme Eva Kartchava

Examinatrice externe

Résumé

Cette étude quasi-expérimentale a pour objectifs de (1) comparer les effets de trois techniques de rétroaction corrective à l'écrit (RCÉ), à savoir la RCÉ directe (reformulation) et deux RCÉs indirectes (soulignement et soulignement plus indices métalinguistiques) sur la précision langagière à l'écrit, (2) d'examiner les effets modérateurs de la catégorie d'erreur et (3) d'examiner les effets modérateurs du niveau de compétence langagière des apprenants ainsi que leur capacité d'analyse verbale.

Quatre groupes de français langue seconde (n = 65) inscrits dans des programmes de français enrichi dans des écoles secondaires anglophones à Montréal ont participé à cette étude. L'intervention expérimentale a comporté trois cycles d'écriture, de rétroaction et de révision qui ont été complétés sur des intervalles de deux semaines, et a ciblé l'accord dans le groupe nominal (GN), l'accord dans le prédicat, la structure du GN, la structure du groupe verbal (GV) ainsi que les homophones grammaticaux. L'intervention a été réalisée à travers trois conditions expérimentales (RCÉ directe, RCÉ indirecte sous forme de soulignement, RCÉ indirecte sous forme de soulignement plus indices métalinguistiques) et une condition de comparaison (pas de RCÉ). Chaque groupe expérimental a complété six tâches d'écriture (rappel de texte) : une tâche durant le pré-test et le post-test immédiat, une autre tâche pour le post-test différé et trois tâches différentes durant l'intervention expérimentale. Les quatre groupes ont révisé les textes complétés durant l'intervention expérimentale. Tous les participants ont aussi complété un test d'analyse verbale.

Les résultats indiquent que (1) la provision de la RCÉ est plus bénéfique que son absence, (2) la RCÉ indirecte est meilleure que la RCÉ directe, (3) la RCÉ combinée avec des indices

métalinguistiques (i.e., RCÉ indirecte métalinguistique) est plus efficace que le soulignement sans indices métalinguistiques. Concernant la deuxième question, les résultats suggèrent que certaines catégories d'erreurs étaient plus sensibles à la RCÉ que d'autres. Bien qu'aucune amélioration n'ait été constatée pour l'accord dans le GN ou pour l'accord dans le prédicat, l'utilisation précise des homophones grammaticaux s'est améliorée pour tous les groupes. Une telle amélioration était plus durable pour le groupe ayant reçu la RCÉ indirecte métalinguistique. Quant à la troisième question, les résultats indiquent que le niveau de compétence langagière n'a joué un rôle modérateur que pour le groupe recevant la RCÉ indirecte métalinguistique, mais aucun effet de ce type n'a été observé pour la capacité d'analyse verbale, et ce, quelle que soit la condition expérimentale.

Mots clés: rétroaction corrective à l'écrit, français langue seconde, révision, RCÉ directe, RCÉ indirecte, indice métalinguistique, précision langagière, apprenants au secondaire, catégorie d'erreur, niveau de compétence langagière, capacité d'analyse verbale

Abstract

This quasi-experimental study aims (1) to compare the effects of three written corrective feedback (WCF) techniques on linguistic accuracy in writing. The three techniques consist of direct WCF, a technique in which the teacher provides the correct form, and two indirect WCF types, underlining only versus underlining plus metalinguistic clues, techniques in which the teacher elicits the correct form from the student. This study also aims (2) to examine the moderating effects of error category, and (3) to explore the moderating effects of individual variables such as language proficiency and language analytical ability.

Four groups of French as a second language students (n = 65) enrolled in enriched French language programs in English secondary schools in Montreal participated in this study. They were divided into three experimental conditions, namely 1) direct WCF, 2) underlining only and 3) underlining plus metalinguistic clues, and a comparison condition - no WCF. The experimental intervention, which consisted of three cycles of writing, feedback and revision, was completed over two-week intervals. It targeted five grammatical categories: agreement in the nominal phrase (NP), agreement in the verb phrase (VP), the structure of the NP, the structure of the VP as well as grammatical homophones. Each group completed six writing tasks (text-reconstruction tasks): one task during the pre-test and the immediate post-test, another task for the delayed post-test and three different tasks during the experimental intervention. All four groups revised the texts they completed during the intervention. All participants also completed a test of language analytical ability.

Results indicate that (1) providing WCF was more beneficial than withholding it, (2) indirect WCF was better than direct WCF, (3) that indirect WCF plus metalinguistic clues was more effective than indirect WCF only. Regarding the second question, results suggest that some

error categories were more responsive to WCF, regardless of its type, than others. While no improvement was found for agreement in the NP or for agreement in the VP, the accurate use of grammatical homophones increased for all groups. Such an increase was more lasting for the group that received indirect WCF that is combined with metalinguistic clues. As for the third question, results indicate that proficiency played a moderating role only for the group receiving indirect WCF that is combined with metalinguistic clues, but no such effects were observed for the language analytical abilities, irrespective of the experimental condition.

Keywords : written corrective feedback, direct WCF, indirect WCF, metalinguistic clues, French as a second language, secondary classrooms, accuracy, proficiency, language analytical abilities.

Table of contents

List of tables

List of figures

List of abbreviations

ANP: agreement in the noun phrase

AVP: agreement in the verb phrase

CF: corrective feedback

EFL: English as a foreign language

ESL: English as a second language

FSL: French as a second language

FL: foreign language

ISLA: instructed second language acquisition

LAA: language analytical ability

L1: first language

L2: second language

ME: metalinguistic explanations

MEES: Ministère de l'Éducation et de l'Enseignement Supérieur

MLAT: modern language aptitude test

NP: noun phrase

SLA: second language acquisition

VP: verb phrase

WCF: written corrective feedback

To my parents, Ismail & Rawdha

Acknowledgments

"*No one starts out on top. You have to work your way up. Some mountains are higher than others, some roads steeper than the next. There are hardships and setbacks, but you can't let them stop you. Even on the steepest road, you must not turn back. You must keep going up. In order to reach the top of the mountain, you have to climb every rock*"— Muhammed Ali.

This doctoral journey has been one of the most challenging endeavors I have ever undertaken, but it has taught me perseverance, resilience and gratitude. Its completion would not have been possible without the support and love of my family, friends and professors.

I would like to start by thanking my thesis supervisor, Dr. Ahlem Ammar for her valuable support, guidance, and mentorship. Thank you, Dr. Ammar, for giving me the opportunity to learn and work with you. You have contributed academically and financially to this achievement, and I am thankful for your expertise, dedication and patience as I navigated the different stages of this doctoral adventure.

Sincere thanks go to my committee members, Dr. Marie-Claude Boivin and Dr. Antonella Valeo, for their valuable time, insightful feedback and thought-provoking questions. I appreciate the stimulating exchanges we had during my comprehensive exam and the useful comments they provided on earlier versions of this work. Many thanks are also due to Dr. Eva Kartchava for her thorough reading and critical review.

I am particularly indebted to my former professors and mentors in the United States. Heartfelt gratitude to Dr. Steve Chandler and Dr. Stephen Flores for encouraging me to pursue my masters in TESL. Thank you for making my graduate studies at the University of Idaho an enjoyable, rewarding and memorable experience. Sincere thanks go to Dr. Marie-Anne Judge and

Dr. Jodie Nictora for helping me gain first-hand teaching experience and for contributing to my professional growth. I would also like to send special thanks to Dr. Kelly Puzio from Washington State University for constantly providing me with valuable advice and for generously sharing helpful resources with me.

I will always be in debt to my former professors and teachers in Tunisia, many of whom have marked my academic and personal life. Sincere thanks go to my English teachers, notably Ms. Chedhlia Mejri, for her dedication and her encouragement. Thank you for nurturing my love for the English language!

This dissertation would not have been possible without the help of my colleagues and friends. My sincere thanks are due to Sana Khlaifa, Hana Al Ahmadi and Syrine Benmassaoud for their time and help during the processes of instruments validation and data collection. Special thanks to Katrine Roussel for her efficient work at double-checking and commenting on data coding and to Simon Houle for his extremely helpful guidance with the statistical analyses.

My journey at Université de Montréal would not have been rewarding had it not been for the wonderful team of the Vice-Dean Office of Graduate Studies and Research (Vice-décanat aux études supérieures et à la recherche). Heartfelt thanks go to Mme Claudine Jomphe. Je vous remercie Mme Jomphe pour votre temps, votre patience, vos précieuses suggestions, vos relectures minutieuses de mes demandes de bourse et vos rétroactions efficaces. Je vous en suis reconnaissante.

I am grateful for the all the funding opportunities I have received since I started my studies at Université de Montréal. As an international student, the different scholarships I was awarded made this doctoral program a possible reality. I am particularly honored for being among the very first international students to receive a doctoral grant from the Fonds de Recherche du Québec

Société et Culture. Thank you to the anonymous advocates who made such prestigious scholarships accessible to international students; thank you for believing in the strength of inclusivity and in the beauty of pluralism.

Last but not least, I am so grateful for the love and care of my husband and my parents. Thank you Zied for your continuous patience and tremendous support. Thank you for driving me to schools, for preparing us meals and for putting up with my stress and complaints. Finally, I would like to extend my undying gratitude to my parents, Rawdha and Ismail, both of whom have never saved money, energy or time to take care of my brothers and me. I can never thank my mother enough for her love, dedication and encouragement. She has always believed in me and was the first one to encourage me to pursue my education and excel academically and professionally. Despite the long distances and the time difference, she has always been there for me. Mom, I owe it all to you, to your unconditional love and to your unwithering support.

إلى أبي و أمي، حفظكما الله تاجا فوق رؤوسنا و أدام عليكما نعم الصحة و العافية

Introduction

The place and role of error treatment in general, and written corrective feedback (WCF) in particular, have generated considerable interest in second language acquisition (SLA) theorizing and L2 teaching (Ferris, 2010; Sheen & Ellis, 2011). From a theoretical perspective, WCF can trigger the cognitive processes of attention, noticing and comparison (Schmidt, 2001; Sharwood-Smith, 1981; Swain, 1985, 1995). It has also the potential of enhancing the transformation of declarative knowledge into procedural and eventually automatized knowledge (DeKeyser, 2007a, 2007b), and of preventing the formulation of erroneous hypotheses about the target language (Swain, 1995). Accordingly, WCF can contribute to the acquisition of new knowledge and the consolidation and restructuring of partially acquired knowledge (Polio, 2012; Williams, 2012). Empirically, descriptive and quasi-experimental studies indicate that second language (L2) teachers provide their learners with various types of WCF (e.g., Ammar, Daigle, & Lefrançois, 2016; Lee, 2007), that L2 learners expect and want to receive their instructors' feedback (e.g., Chandler, 2003; Lee, 2004) and that WCF leads to increased levels of accuracy (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001).

However, when it comes to more specific questions such as comparing the differential impacts of various WCF techniques (e.g., direct versus indirect WCF) and investigating the potential mediating effects of error types and learner individual differences, the extant WCF research is plagued with conflicting findings (see Guénette, 2007; Storch, 2010). The inconclusive results about the relative merits of different WCF operationalizations, as well as the need to account for mediating variables, motivate the present study. Such an endeavor is also justified by the different calls for conducting more WCF classroom-based studies (e.g., Ammar *et al.*, 2016; Bitchener & Storch, 2016; Kang & Han, 2015; Ortega, 2012).

The study examines the differential impacts of three WCF techniques, one direct technique (i.e., reformulation without metalinguistic clues) and two indirect techniques (i.e., underlining only versus underlining with metalinguistic clues). It also explores the potential mediating effects of error type (i.e., morphological grammar and grammatical homophones) and learner individual differences (i.e., proficiency and language analytic abilities).

In Chapter One, an overview of the theoretical framework for the general research question that is investigated in the present study is presented. The chapter begins by underscoring the social relevance of writing and written corrective feedback in Quebec's French L2 programs. Then, it outlines the various theoretical foundations of corrective feedback. Finally, it provides a brief overview of WCF empirical research, which culminates into the general research question of the study.

Chapter Two provides a more detailed discussion of the conceptual and empirical frameworks of the present study. In this chapter, the different key concepts, such as L2 acquisition, written corrective feedback, and their different operationalizations are delineated. The various theoretical arguments underlying the different WCF techniques are also presented, and a detailed overview of the extant WCF research is provided. Against this backdrop, the specific research questions are presented.

Chapter Three presents the different methodological choices that were made to carry out the current study. In this chapter, the research context, the participants, the different experimental conditions, the data collection tools, the experimental procedure and the data analysis are described.

In Chapter Four, the descriptive and statistical analyses that were carried out to answer the research questions are reported. In this chapter, the analyses performed to examine between and

within group differences in terms of the overall accuracy and accuracy for specific error category are described. The regression analyses carried out to explore the moderating effect of learners' proficiency and language analytical ability are reported.

In Chapter five, the results of the descriptive and statistical analyses pertaining to each one of the research questions are discussed. Pedagogical and theoretical implications are also presented. Methodological limitations are acknowledged and directions for future inquiries are proposed.

Chapter 1

Statement of the Research Problem

1 Statement of the Research Problem

The aim of the first chapter is to lay the groundwork for the general research question that motivates the present study. It consists of three sections that put forward different arguments in favor of written corrective feedback. In the first section, a general overview of the social and educational relevance of writing and written corrective feedback in Quebec's French as a second language (FSL) settings is offered. In the second section, different theoretical underpinnings of corrective feedback are presented and discussed. In light of these theoretical arguments, empirical research on WCF is briefly presented and its major limitations are highlighted in a third section. This chapter, then, culminates into the general research question that the present study seeks to address.

1.1 Socio-educational Relevance of Writing and Written Corrective Feedback

A growing number of researchers and practitioners have underscored the importance of developing literacy skills (Goldenberg & Coleman, 2010; Guthrie, Schafer, & Hutchinson, 1991; Kennedy, 2013). Goldenberg and Coleman (2010), for example, argue that "students need to read and write at sufficiently high levels if they are to be successful throughout school and beyond" (p. 39). Not only are literacy skills correlated with academic success, but they are also closely related to socio-economic well-being, fair health status, informed financial decisions and professional achievement (Guthrie, Schafer, & Hutchinson, 1991; Kennedy, 2013; Murnane & Levy, 1996). In other words, limited literacy skills jeopardize academic success and limit social and economic opportunities (Goldenberg & Coleman, 2010). Not surprisingly, and in light of the alarming statistics showing that "approximately 53 per cent of the population [in Quebec] do not reach the necessary threshold to function properly in … society" (Lévesque, 2015, n.p.), different stakeholders, including educators and policy makers, have voiced their concern about the

negative repercussions of low literacy rates on Quebec's social economy. It is against this backdrop that promoting literacy skills, be they in reading and/or in writing, as well as enhancing the quality of written French have become major preoccupations for policymakers and language researchers alike (Chartrand & Lord, 2013; Ministère de l'Éducation et de l'Enseignement Supérieur, 2017).

More specifically, different researchers have emphasized the importance of developing writing competency in first and second language learning contexts alike. The necessity to develop this competency is accounted for by the crucial societal and acquisitional roles that writing plays (Fillmore & Snow, 2000; Leki, Cumming & Silva, 2008; Ortega, 2012). As emphasized by Chartrand (2013), written language is closely linked to the development of high-level intellectual skills, cultural appropriation and personal and social development. Writing serves academic and professional purposes: it is a transversal skill that students call upon at different levels of their schooling, in language and content classes alike (Fillmore & Snow, 2000). Their academic success is, thus, tightly related to the development of their writing competency (Bussière, Catwright, Crocker, *et al.*, 2001; Le Ferrec, 2008). Writing is also a major medium of communication in the workplace: when applying for a job, candidates need to submit clear and compelling curriculum vitae and motivation letters, and in an increasingly digitalized and globalized world, different professional communications are carried out through the written modality (Grégoire, 2012). In addition to its communicative functionalities, writing has a potential for first and second language acquisition (Leki, *et al.*, 2008; Manchòn, 2011). Engaging learners in writing activities would not only enable them to use their developing linguistic skills in contextualized tasks, but it would also facilitate and enhance the process of acquiring the target language (Manchòn, 2011; Williams, 2012). In Quebec, for example, where French is the official

language, writing clearly and accurately in French would enable learners to better integrate to Quebec's socio-economic landscape (MEES, 2008).

Against this backdrop, the *MEES*'s description of the enriched French program emphasizes the importance of fostering French learners' composition skills. It clarifies that secondary French as a second language (FSL) students are expected to develop fluency and produce various written texts that are marked by both clarity and linguistic precision. The *MEES* confirms that clear and precise written production would enable learners to play an effective role in Quebec's French-speaking community (p. 131). For this reason, the enriched French program aims at equipping learners with the tools that would assist them in pursuing their academic studies in French and/ or in working in a francophone environment. Accordingly, writing in French is described as one of the three "inter-related" competencies, along with oral interaction and reading comprehension. In other words, developing the writing competence is assumed to bring about improvement in reading and in oral communication.

Not only does writing in a second language (L2) help learners integrate to their host community, but it can also enhance and facilitate their linguistic development. Such a perspective, anchored within the *writing to learn* framework, marks a departure from the traditional view of writing (and any output activity) as a simple reflection or a secondary by-product of the acquisition process (Ferris, 2010; Manchòn, 2011). In fact, thanks to the growing interest in the interface between second language acquisition (SLA) and second language (L2) writing, writing in an L2 has been reconceptualised as a vehicle of second language development (Williams, 2012). This reconceptualization has brought about a shift from the *learning to write* perspective, i.e., using the written modality for self-expression, to the *writing to learn language* view, i.e., using writing tasks for language development (Cumming, 1990; Manchòn, 2011). As a

consequence, it has motivated theoretical and empirical investigations that set out to examine how engaging learners in writing activities can be conducive to language learning. It is argued that the language development potential of writing can be explained by two major characteristics that distinguish written output activities from their oral counterparts. Compared to speaking, writing is marked by (1) its slow pace and (2) its permanent record (Williams, 2012). These two inherent features allow learners more control over the comprehension and production of their messages (and their peers' texts). As such, writing creates increased chances for learners to attend to both meaning and form and to cater to content and to linguistic accuracy (Ortega, 2012; Williams, 2012).

In light of the multifaceted relevance of writing, practitioners and researchers have put forward different instructional and interventional techniques that are meant to help learners enhance their writing competency and maximize its potential for language learning. Examples of those teaching strategies include text-reconstruction activities (e.g., dictogloss), collaborative composition tasks and written corrective feedback. Written corrective feedback is "a written response to a linguistic error that has been made in the writing of a text by an L2 learner. It seeks to either correct the inaccurate usage or provide information about where the error has occurred and/or about the cause of the error and how it may be corrected" (Bitchener & Storch, 2016, p. 1). Written corrective feedback is also described as a reactive option of form-focused instruction (FFI), with the latter defined as "any pedagogical effort which is used to draw the learners' attention to form either implicitly or explicitly . . . within meaning-based approaches to L2 instruction [and] in which a focus on language is provided in either spontaneous or predetermined ways" (Spada, 1997, p. 73). Underlining the advantages of reactive instructional techniques (i.e., corrective feedback), Doughty and Williams (1998) explain that CF techniques are contingent on

learners' needs and are thus more consistent with the tenets of communicative language teaching approaches, which are widely used in Quebec's FSL classrooms. Furthermore, within the reactive teaching methods, WCF is distinguished from oral CF by its enduring trace, accentuating thus its presence and potential noticeability by learners (Ortega, 2012).

Fittingly, some of the above-mentioned theoretical arguments in favor of WCF are echoed in the *MEES*'s description of its Enriched French program, according to which learners need to engage in producing varied types and genres of texts, for the latter constitute a fertile ground for contextualized grammar teaching and use. The *MEES*'s description states that FSL learners are expected to produce coherent texts in which linguistic and formal conventions are respected. Put differently, learners need to attend to rhetorical requirements, as well as to morpho-syntactic and orthographic rules. The importance of grammatical accuracy in written texts is reflected in the evaluation grid that the *MEES* developed for FSL teachers grading the French exit exams completed by secondary five L2 students. While 55 points are allocated to macrostructure issues such as content and coherence, 45 points are allotted to linguistic conventions. In case the number of errors exceeds 15% of the total number of written words, a student's written text receives an F.

To better assist learners in the complex task of producing varied, coherent and accurate texts, FSL teachers are encouraged to help their students develop their linguistic competency and to perceive their errors as a source of learning. They are also expected to provide frequent feedback that would draw their students' attention to their erroneous use of the target language (MEES, 2016).

In a nutshell, developing the writing competency is a major concern for the *MEES* because of its academic and social roles. Particularly, helping FSL learners produce accurate texts and foster their language acquisition is a central goal of the Enriched French program. Along these

lines, practitioners and researchers propose that engaging learners in written production activities and providing them with WCF, which would draw their attention to their erroneous output, can facilitate their writing and language development (Ferris, 2010; Ortega, 2012; Williams, 2012). These premises are in large informed by different theoretical accounts of corrective feedback in writing research in general and in second language acquisition (SLA) research in particular. Accordingly, the following section highlights the role of written corrective feedback in writing models and in second language learning theories.

1.2 Theoretical underpinnings of WCF

Arguments about the usefulness of written corrective feedback have been put forward in two distinct, yet overlapping, bodies of literature (Ferris, 2010; Ortega, 2012). These consist of L2 writing research and second language acquisition (SLA) accounts. Consecutively, the section below will first highlight the place of corrective feedback in writing models and will second delineate its debated role in key SLA theories.

1.2.1 Corrective Feedback in Writing Models

Research into L2 writing theory and practice has been largely informed by L1 writing theoretical models. These models have witnessed a paradigm shift during the 1980s as the focus has gradually shifted from describing the final product to analyzing the cognitive processes involved in the writing activity (Flower & Hayes, 1981; Silva, 1993). L1 cognitive models of writing as propounded by Flower and Hayes (1981) reconceptualised writing as a recursive, hierarchically-organized, problem-solving process. The resultant cognitively-oriented conceptualization consists of three major components that influence the strategies writers call upon during the writing process (see Figure 1).

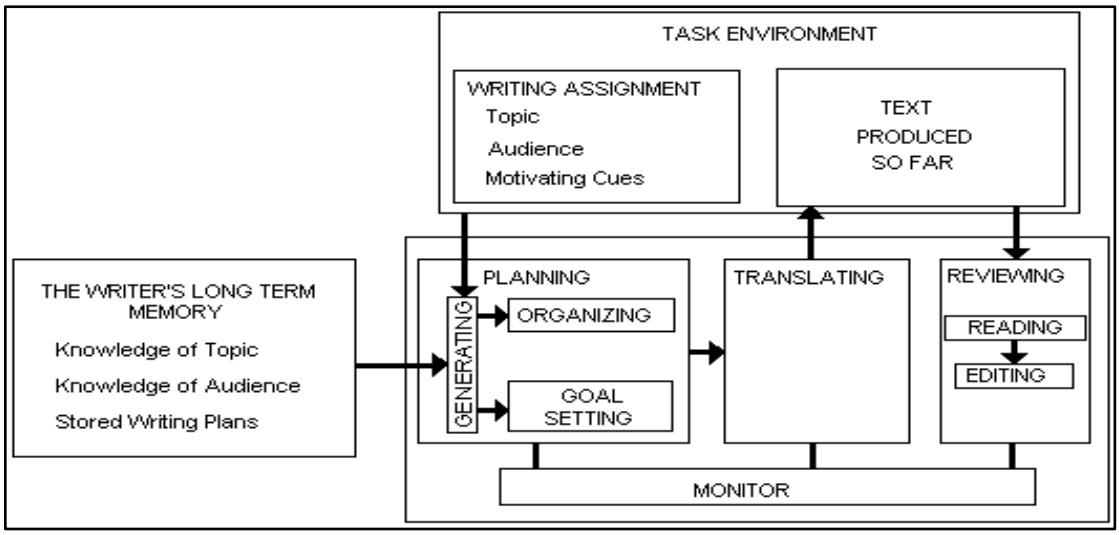


Figure 1. Structure of the writing model by Flower and Hayes (1981)

The first component of this model is the *task environment* which refers to the writing assignment, the audience and the text produced, etc. The second element is *long-term memory knowledge* about the topic, the audience, the writing plan and the rules of grammar. Informed by these two elements are the three sub-processes of composition: planning, translating and reviewing. Planning refers to generating ideas and organizing them, translating (also known as transcribing) is about putting the conceptual and ideational content into text, while reviewing consists of reading and editing. The interplay between these three recursive sub-processes is coordinated by the "Monitor", which regulates the switching between the different processes and the allocation of attentional resources (Becker, 2006; Fitzgerald, 1987).

Within the cognitive models, revision was repositioned as an important sub-process, which resulted into a proliferation of theoretical and empirical investigations into the nature of revision (Fitzgerald, 1987; Flower, Hayes & Carey, 1986; Hayes, Flower & Shriver, 1987). According to Flower and Hayes (1981), reviewing is the thinking process of evaluating what has been written that may (or may not) culminate into making changes (i.e., into revision). In other words, revision

is the manifestation of the reviewing process (Fitzgerald, 1987). As they are evaluating their texts, writers need to detect and recognize "incongruities between goals and instantiated text" to make, consecutively, the necessary changes (Fitzgerald, 1987, p. 484). In their revised models of the reviewing process, Flower, *et al.* (1986) outline three mechanisms that operate recursively during the revising process: (1) evaluating the written text to detect and diagnose incongruities, (2) selecting a revision strategy, and (3) executing the strategy. Proposing a more complex conceptualization, Scardamalia and Bereiter (1987) redefined revision as a self-regulated process consisting of three recursive mechanisms: the first mechanism refers to *comparing* the representations of the actual and the intended texts. A mismatch between the two representations would trigger the second mechanism, i.e., *diagnosing*, whereby writers try to identify the nature and the origin of the observed inadequacy and think of possible solutions. The third mechanism, i.e., *operating*, refers to choosing a solution and making changes to the instantiated text.

Regardless of the different categorizations, both models suggest that activating the above-mentioned sub-processes depends on the linguistic knowledge that is stored in long- term memory (Alamargot & Chanquoy, 2001). They also presuppose that writers have enough linguistic knowledge to detect, diagnose and solve discrepancies (Alamargot & Chanquoy, 2001), which is not necessarily the case for all learners and all age groups. More importantly, these cognitively- anchored models imply that revision is internally initiated, rather than being externally triggered. In other words, they sideline revision which may occur in reaction to teacher or peer feedback.

Informed by the above-mentioned cognitive accounts, different descriptive and experimental studies have been carried out in L1 and L2 writing contexts. Research on L2 writing examines the extent to which L1 cognitive models are equally applicable to L2 composing

processes and whether writing in an L2 results in more demands for attentional resources. Studies also investigate if the additional constraints of L2 writing (such as the developing linguistic competence) affect the problems learners would attend to and how they would go about solving them (De Larios, Murphy & Marin, 2002; Silva, 1993). These studies reveal that while L1 and L2 writing involve comparable processes, i.e., planning, transcribing and reviewing, they exhibit "salient differences" (Silva, 1993. p. 661; Silva & Leki, 2004). In her critical overview of 72 studies comparing composition in English L1 and English as a second language (ESL), Silva (1993) argues that "L2 composing [is] clearly more difficult and less effective" than L1 writing (p. 661). She explains that ESL learners do not only produce less accurate texts, but they also struggle more with reviewing their compositions. L2 writers are reluctant to reread (Uzuwa, 1996), and when asked to revise, they are often unable to improve the accuracy of their texts (Truscott & Hsu, 2008; Van Beuningen, Jong & Kuiken, 2012).

The lack of effective revisions is due to learners' inability to detect errors, to diagnose their nature and to reflect on possible solutions (Becker, 2006). As argued by Becker (2006), "detection of a problem becomes the key determiner for the direction any revision work might take" (p. 39). It follows that very little revision is done unless teacher or peer feedback is provided (Becker, 2006; Fitzgerald, 1987; Van Beuningen, Jong & Kuiken, 2008). When provided properly, written corrective feedback can signal errors, which addresses the problem of detection. It can also help learners diagnose the nature of their erroneous language use and prompt them to consider other alternatives. Studies that compare L2 learners who received WCF to those who did not show that the former were more successful at revising their texts, underscoring thus the effectiveness of feedback as a revising tool (Ashwell, 2000; Ferris & Roberts, 2001; Truscott & Hsu, 2008; Van Beuningen *et al.*, 2008; 2012).

While conceding that WCF can help learners revise their texts, some SLA researchers have argued that leading to effective revisions would not necessarily result in promoting second language development (Truscott, 1996, 1999), downplaying in turn the potential usefulness of WCF. Other researchers, however, have reaffirmed the facilitative role of WCF in L2 learning and have sought to examine its long-term impacts on sustained accuracy in new texts. The debate about the short- and long-term effectiveness of WCF is best captured in Manchòn's (2011) distinction between feedback as a revising tool, "feedback for accuracy," which is generally the focus of L2 writing investigations, and feedback as a learning tool, "feedback for acquisition," which is the locus of SLA inquiries (p. 162). Manchòn's (2011) distinction is anchored within the broader perspectives of "learning to write" (i.e, feedback to improve learners' drafts) versus the "writing to learn" (i.e., written feedback to promote second language development)[1].

Against this background, SLA oriented studies on WCF draw on various SLA theoretical accounts (Bitchener & Ferris, 2012; Ortega, 2012; Williams, 2012), which are outlined and discussed in the following section.

1.2.2 Corrective Feedback in L2 Learning Theories

Out of the many SLA theories, there are four broad models that are particularly relevant to inquiries into corrective feedback[2]. These consist of (1) behaviourism, (2) nativism/ innatism, (3) interactionism, and (4) psycho-cognitivism (Bitchener & Ferris, 2012; Kartchava, 2013).

---

[1] These classifications of feedback as revising tool and feedback as a learning tool are not mutually exclusive.

[2] Depending on their theoretical paradigms, researchers have used different terms to refer to corrective feedback. For clarity purposes, corrective feedback will be used all along in this section, and the differences between the various constructs will be explained in chapter 2.

1.2.2.1 Behaviorism

Behaviorist accounts view language learning in terms of habit-formation, which is independent of mental or internal processes and which is rather contingent on the surrounding environment (e.g., Skinner, 1957). Learners are subject to environmental stimuli for which they are trained to develop accurate responses. Those responses will either become habits if learners receive reinforcement (i.e., rewards or verbal appraisals), or will be discarded in the case of punishment (i.e., corrective feedback). Against this framework, L2 learning consists of having learners repeatedly imitate correct models. Following the modeling and imitation, learners are to be provided with either reinforcement if their renditions are accurate or corrective feedback if they are inaccurate (Lightbown & Spada, 1999; VanPatten & Williams, 2007). Within the behaviorist model, overt and explicit corrective feedback is instrumental in establishing the desired new L2 habits. This is the case because behaviorists hold negative views of errors: as highlighted by Brooks (1960), "error, like sin, is to be avoided and its influence overcome" (p. 58, cited in Bitchener & Ferris, 2012, p. 4). Errors, thus, should be addressed through immediate and explicit corrective feedback. In a nutshell, behaviorists contend that "repetitions of correct models, as well as immediate and consistent … error correction, [are] the best way to eradicate errors before learners develop bad habits" (Van Patten & Williams, 2007, p. 21). It should be noted that although behaviorists have prioritized oral communication over writing, their arguments concerning the necessity and usefulness of corrective feedback have informed research into both oral and written corrective feedback (e.g., Bitchener & Ferris, 2012).

Nonetheless, the behaviorist interventionist approaches, and their emphasis on the role of feedback, were soon disregarded by innatist/generativist theories dominating the fields of linguistics and first language acquisition (VanPatten & Williams, 2007).

1.2.2.2 Generativist Accounts

In his demolishing review of Skinner's book, Chomsky (1959) explains, "it is simply not true that children can learn language only through 'meticulous care' on the part of adults who shape their verbal repertoire through careful differential reinforcement" (p. 42). Chomsky (1965, 1981) argues that language acquisition is driven by an innate mental capacity, known as language acquisition device (LAD). LAD consists of built-in universal linguistic principles, i.e., Universal Grammar, which operate unconsciously and guide learners to internalize a complex grammar despite an impoverished input (White, 2003). For UG to be activated, all that L1 acquirers need is access to primary language data in their input, also known as positive evidence. Furthermore, given that caregivers usually prioritize the veracity of the message over its form, seldom do they provide children with negative evidence, i.e., information about what is unacceptable in their language (Chomsky, 1986). As such, negative evidence including corrective feedback has no bearing on the process of L1 acquisition, which rather occurs unconsciously thanks to the LAD.

Although Chomsky did not make any claims about L2 learning, different SLA researchers (e.g., Schwartz, 1993; Schwartz & Gubala-Ryzak, 1992; White, 1996, 2003) have espoused his arguments and maintained that the underlying L2 linguistic competence is governed by the same UG principles as L1 acquisition. In other words, developing L2 grammar is also an unconscious process that results from UG interacting with linguistic input. These researchers disagree, however, on the potential role of negative evidence, including corrective feedback, in developing L2 grammar (e.g., Schwartz vs White). On the one hand, it has been argued that L2 learners might need to receive direct negative evidence, i.e., information about what is not allowed in the target language (e.g., L1 French speakers learning adverb placement in L2 English), which they cannot induce based on positive evidence from primary linguistic data alone (White, 1991, 2003).

On the other hand, it has been claimed that "the grammar-building process cannot make use of negative evidence to restructure (Interlanguage) grammars" (Schwartz & Gubala-Ryzak, 1992, p. 1), implying that corrective feedback is of little (no) relevance to L2 acquisition (Krashen, 1982; Schwartz, 1993). Along the same lines, Schwartz (1993) explains that negative evidence does not trigger Universal Grammar, arguing, however, that the latter can be activated only by positive evidence (i.e., input). While she concedes that negative evidence, including corrective feedback, may lead to explicit learning, she emphasizes that the resultant explicit knowledge would not necessarily change L2 competence. Similar arguments against the utility of corrective feedback are articulated more clearly in Krashen's Monitor Model, which draws heavily on Chomsky's accounts of L1 acquisition. This model is discussed in more detail next.

Krashen's Monitor Model. Considered as the first foundational theory in SLA (Van Patten & Williams, 2007), Krashen's model consists of five major hypotheses, which are the acquisition/learning hypothesis, the comprehensible input hypothesis, the monitor hypothesis, the affective filter hypothesis, and the natural order hypothesis.

The Monitor Model is based on Krashen's distinction between language acquisition and learning. In his first hypothesis, Krashen upholds that acquisition is a subconscious process which, similarly to children's L1 acquisition, occurs implicitly as a result of L2 learners' exposure to input, which he described as "the primary causative variable in second language acquisition" (Krashen, 1982; p. 32). In contrast, learning is a "conscious process that results in 'knowing about' language" (Krashen, 1985, p. 1). He argues that acquisition is the default mechanism in L2 development whereas "learning is more peripheral" (Krashen, 1982; p. 20). The peripheral role of learning consists of monitoring (i.e., editing) the output produced by the acquired system. However, this role is quite constrained because it can take place only if three

conditions are fulfilled. First, learners should have sufficient time to call upon their learned knowledge; second, they should be interested in the linguistic accuracy of their message, and third, they should have access to the necessary grammatical knowledge (Krashen, 1985).

More controversially, he claims that these two modes, i.e., acquisition and learning, are separate and remain totally independent and non-interfacing. Acquisition is not impacted by learning given that rules of language are acquired in a "predictable order" that cannot be altered by the "order in which rules are taught in language classes" (Krashen, 1985, p. 1). An obvious corollary of this is that grammar teaching, and by extension CF- both of which are forms of metalinguistic input- would feed into learning and would not bring about any acquisitional benefits; they are thus unnecessary. Rather than dissipating learners' time and effort on form-focused instructional interventions, teachers need to expose learners to comprehensible input, which should be slightly beyond learners' current interlanguage (i+1). This comprehensible input should also be provided with a low affective filter, i.e., in low anxiety situations. Krashen (1982) adds that pushed output and corrective feedback can indeed bring about high levels of anxiety, which in turn prevent learners from "seek[ing] and obtaining more input" (p. 31). As such, "a safe procedure is simply to eliminate error correction entirely" (Krashen, 1982, p. 76).

Overall, Krashen's Monitor Model advocates for a naturalistic approach to language teaching that prioritizes "meaningful interaction in the target language - natural communication - in which speakers are concerned *not with the form* of their utterances but with the messages they are conveying and understanding" (Krashen, 1981, p. 1; emphasis added). There should be little effort devoted to explicit learning given that the accruing learned knowledge remains separate from acquired knowledge, and that the latter is viewed as the driving force in developing L2 competence.

Krashen's theoretical arguments have been embraced by some L2 writing practitioners who have put forward pedagogical arguments undermining the role of written corrective feedback in developing L2 acquisition. These arguments are described next.

Truscott's criticism of WCF. In an ongoing debate about the merits of corrective feedback (Truscott, 1996, 1999, 2004, 2007; Truscott & Hsu, 2008), Truscott articulates two arguments that cast doubt on the utility of WCF. First, in accordance with Krashen's natural order hypothesis, according to which, rules of language are acquired in a "predictable order" that cannot be altered by the "order in which rules are taught in language classes" (Krashen, 1985, p. 1), Truscott claims that WCF is incongruent with the L2 acquisition process for it ignores learners' developmental readiness. He points out that teachers tend to provide WCF on grammar points for which learners are not ready yet, and by doing so they ignore that "the acquisition of a grammatical structure is a gradual process, not a sudden discovery as the intuitive view of correction would imply" (Truscott, 1996, p. 342). Second, and in line with Schwartz's (1993) claims about feedback, Truscott argues that the possible outcomes of WCF pertain more to "pseudo-learning", which he defines as "superficial and possibly transient form of knowledge" (p. 345). Drawing on Krashen's distinction between learning and acquisition, Truscott reasons that learners' possible modifications as a response to WCF do not necessarily mean that their underlying processes of L2 acquisition will be affected (also see Schwartz, 1993). In other words, and consistent with Krashen's skepticism about the role of grammar teaching, WCF may lead to

learning, but it would not feed into acquisition[3]. Truscott (1996) concludes by stating that, "grammar correction has no place in writing courses and should be abandoned" (p. 328).

It should be noted here that Truscott's arguments were rebutted by Ferris (1999, 2003, 2004), and that their heated debates have motivated an increasingly-growing number of descriptive and (quasi)-experimental studies on WCF and its role. Most of these inquiries are underpinned by various theoretical arguments that were developed as a reaction to Krashen's Model and its underestimation of grammar teaching and error treatment. These positions are discussed next.

1.2.2.3 The Insufficiency of Comprehensible Input

Krashen's Model, influential as it was, was challenged on both theoretical and empirical grounds. Theoretically, different researchers have questioned the validity of his dichotomous conceptualization of learning and acquisition (e.g., McLaughlin, 1978, 1987); others have pointed out that acquisition and comprehension are distinct processes (e.g., Sharwood Smith, 1986). Sharwood Smith (1986) explains that acquisition, conceptualized as the "creation (or restructuring) of grammatical competence" is different from comprehension, which he defines as the extraction of meaning (p. 239). It follows that comprehending input for its meaning would not automatically result into acquisition as speculated by Krashen. Along those lines, Skehan (1998) maintains that as learners attempt to comprehend a text, be it visual or auditory, they "draw on a wider range of resources, including both schematic and contextual knowledge" and rely on top-down strategies (e.g., world knowledge) (p. 15). He claims that the ensuing

[3] The acquisition/learning distinction applies to Krashen's model only. The two constructs are henceforth used interchangeably throughout this thesis.

comprehension process can be "detached from the underlying syntactic system and from production" (p. 15), thus undermining Krashen's comprehension-driven language acquisition. In comprehension-focused contexts, learners might develop effective strategies to process meaning, which is often achieved by "bypassing the form of language" (p. 11).

The criticism levelled against Krashen's Model was also based on empirical findings from studies on naturalistic learning contexts (e.g., Schmidt, 1983; Schmidt & Frota, 1986), as well as inquiries on French immersion (e.g., Swain, 1985) and intensive English programs (Spada & Lightbown, 1989). Those enquiries revealed that even after extensive exposure to ample L2 input, L2 learners lag behind native speakers especially in terms of the well-formedness and accuracy of their production (Lightbown & Spada 1990; Swain, 1985, 1993). These results led SLA theorists and researchers to contest the sufficiency of comprehensible input, which is a central hypothesis in Krashen's model, and to argue for the need to draw learners' attention to the formal properties of the target language, via form-focused instruction and corrective feedback, be it oral or written (Lightbown & Spada, 1990, 1994; Swain, 1993).

Support for (written) corrective feedback can be found in three major cognitively-oriented hypotheses and theories: the Noticing Hypothesis (Schmidt, 1990; 1994), the Output Hypothesis (Swain, 1985, 1993, 1995), Skill-Learning Theory (Anderson, 1993, 2015; DeKeyser, 2007a, 2007b). Although they agree on the importance of comprehensible input, these theories have pointed to its insufficiency in ensuring L2 acquisition. Besides concurring on the insufficiency of input in L2 acquisition, these theories share a common interest in the mental representations and processes involved in creating, automatizing and restructuring L2 knowledge (Ortega, 2009). Furthermore, they are predicated on the grounds that attention and memory, two key cognitive resources, are limited (Ortega, 2009).

One of the first arguments to contest the sufficiency of Krashen's comprehension-based model was put forward by Schmidt (1990, 1994, 2001), whose hypothesis is discussed next.

Schmidt's Noticing Hypothesis. While agreeing with Krashen about the necessity of input, Schmidt (1990, 1994, 2001) disagrees that it is the only sufficient condition for L2 acquisition. He emphasizes that noticing, which involves attention and awareness, plays a central role in driving L2 development forward. Schmidt (1994a) defines noticing as "the registration of the occurrence of a stimulus event in conscious awareness" (p. 166). Such a registration might involve noticing a form, noticing a form-meaning mapping, noticing gaps between one's developing language and the target language, or noticing the holes in one's knowledge (Izumi, 2013). Learners can, in fact, attend to "a mismatch or gap between what they can produce and what they need to produce, as well as between what they produce and what target language speakers produce" (Schmidt, 2001, p. 6).

In its original version, the Noticing Hypothesis postulates that noticing is "the necessary and sufficient condition for the conversion of input into intake," which subsequently leads to acquisition (Schmidt, 1993, p. 209). Conceding that unattended learning (i.e., that which takes place without awareness) is possible, albeit "limited in scope and relevance for SLA" (2001, p. 3), Schmidt upholds that "more noticing leads to more learning" (1994a, p. 18). As a result, the adjusted version of the Noticing Hypothesis emphasizes that noticing plays a facilitative, rather than a causative, role in propelling L2 development.

More importantly, Schmidt argues that noticing can be ensured via instruction and that CF can play a crucial role in helping learners notice formal properties in the L2 input. He contends that "[c]orrective feedback … juxtaposes the learner's form i with a target language form i+1 [which] puts [the learner] in an ideal position to notice the gap'' between her utterance and the L2

(Schmidt, 1990, p. 313). As such, corrective feedback is viewed as an attention getting device which enables learners to notice mismatches or discrepancies between their developing linguistic system and the target language. This eventually would help them restructure and fine-tune their partially- acquired knowledge.

The potential role of CF in facilitating the noticing of discrepancies and gaps in learners' L2 knowledge is further emphasized in Swain's Pushed Output Hypothesis, which is discussed next.

Swain's Pushed Output Hypothesis. Of the different SLA critiques of Krashen's sufficiency of input, it is Swain's Pushed Output hypothesis that lends the strongest support in favor of output activities and the provision of corrective feedback. Comparing the written and oral performance of learners from immersion programs to those of French native speakers, Swain (1985) shows that immersion students fail to reach a target-like level of accuracy despite their exposure to plentiful, comprehensible input. She explains that learners falling short of well-formedness in the target language is due to their limited engagement in L2 production and the lack of form-focused instructional interventions (Swain, 1985, 1993, 1995). As she notes, immersion students, "are simply not given adequate opportunities to use the target language in the context of classroom [and they] are not being pushed in their output" (1985, p. 249). Her findings and observations were interpreted as evidence for the insufficiency of input and for the potential role that output plays in facilitating L2 acquisition, which is encapsulated in her Pushed Output Hypothesis.

To account for how output facilitates and drives L2 acquisition, Swain (1995) explains that output plays three overlapping functions. First, it enhances noticing; in fact, in their attempts to encode their intended meanings, learners come to recognize the gaps and 'notice the holes' in

their interlanguage (Swain & Lapkin, 1998), becoming hence more aware of the limitations of their linguistic capabilities. A second function of output consists of allowing learners to test hypotheses about the target language. In other words, when producing language, learners are evaluating assumptions about how well-formed their messages are. More important, their non-target like output would call for feedback, which may prompt them to "modify or reprocess their output" (Swain, 1995, p. 126). Third, output has a metalinguistic function, which is particularly evident when learners engage in negotiating formal features of their output. Taken together, these three functions of output underscore its potential in restructuring and consolidating learners' linguistic knowledge.

Along those lines, Izumi (2003) highlights the psycholinguistic mechanisms by which pushed output sets in motion acquisition-related processes. Drawing on Levelt's model of speech production, he explains that it is possible for learners to rely on their top-down strategies (i.e., world knowledge) in comprehension activities and consequently to circumvent syntactic processing (i.e., noticing and understanding formal properties of language features). In this respect, Izumi is echoing Skehan (1998)'s claim that "processing language to extract meaning [i.e., semantic processing] does not guarantee automatic sensitivity to form [i.e., syntactic processing]" (p. 41). Conversely, when engaged in production, learners find themselves "responsible for message generation and formulation that requires grammatical encoding" (Izumi, 2003; p. 183); they are thus engaged in syntactic processing. Subsequently, when producing language, L2 learners are "forced to move from the semantic processing prevalent in comprehension to the syntactic processing needed for production" (Swain & Lapkin, 1995, p. 375).

While some have suggested that learners may rely on discourse competence and compensatory strategies to overcome their lacking grammatical competence (e.g., Skehan & Foster, 2001), one can argue that circumventing syntactic processing is more likely to occur in oral communication, since learners are pressured to speak online and to make themselves understood. To this end, they might rely on a shared situational knowledge, speak in incomplete sentences or pause hoping for their interlocutor to cue them more. However, such compensatory strategies are less likely to be called upon in writing, given that in comparison to speaking, writing is distinguished by its recursive nature, slow pace and permanent record (Flower & Hayes, 1981; Manchòn & Roca de Larios, 2007; Williams, 2012). It is recursive because it involves a continuous back and forth between the intended meaning (and its mental representation) and the instantiated text (in terms of form and meaning). This is particularly possible thanks to the offline character of writing and its slow pace, which allows learners more control over the comprehension and production of their written texts. It follows that learners are afforded more opportunities to attend to both meaning and form and to cater to content as well as, linguistic accuracy (Williams, 2012). The advantages created by writing are also underscored by Manchòn and Roca de Larios (2007) who explain that "it is perfectly possible for the communicative function (i.e., actual writing) to be called to a halt while one pays attention to the form of the language" while this might engender a communication breakdown during speaking (p. 108). In other words, writing pushes and allows learners to go beyond semantic (ideational) coding and to carry out deeper syntactic and morpho-syntactic processing, so as to produce their intended messages in a clear and accurate way (Manchòn, 2011).

The resultant syntactic processing would, in turn, elicit feedback that would help learners "confirm, reject or modify their hypothesis" (Izumi, 2003, p. 186). This is in line with Swain's

(1991) cautionary note that the facilitative roles of pushed output are constrained "if students are given insufficient feedback or no feedback regarding the extent to which their messages have successfully (accurately, appropriately, and coherently) been conveyed" (p. 98). By the same token, the potential learning roles of writing are maximized when learners receive adequate WCF that pushes them to test hypotheses about the target language and to restructure and fine-tune their L2 knowledge (Bitchener, 2012; Polio, 2012). What clearly appears from Swain's Output Hypothesis and its proponents is that pushed output and corrective feedback — be they oral or written — have clearly evolved from being irrelevant, and even undesirable, practices to becoming effective teaching practices with the potential of promoting second language development (Ferris, 2010; Manchòn, 2011; Williams, 2012).

Support for corrective feedback can also be drawn from skill acquisition theories (Anderson, 1983, 2015; DeKeyser, 2007a), which are described next.

Skill Acquisition Theories. Unlike Krashen who opposes CF on the grounds that it results in explicit learning, which is categorically independent from acquisition, proponents of skill acquisition theories posit a strong-interface position between explicit and implicit knowledge and uphold that developing explicit knowledge is a necessary, initial building block in acquiring any complex cognitive skill, including language learning and possibly, more specifically, writing. They also posit an instrumental role of practice in the process of skill (language) acquisition (Anderson, 2015; DeKeyser, 2007b).

Predicated on the assumption that "language … is similar in character to other cognitive activities" (Anderson, 1983, p. 267), skill acquisition theories uphold that first and second language learning are not different from the acquisition of any other cognitive skill (e.g., solving an algebra problem). As such, language learning follows a gradual progression from "initial

representation of knowledge" to an automatic, effortless "highly skilled behavior" (DeKeyser, 2007a, p. 97). This progression consists of three major stages, i.e., the cognitive, associative and autonomous stages (Anderson, 2015), which differ on (a) the type of knowledge developed, i.e., declarative, procedural or automatized, and on (b) the nature of the activated processing, i.e., slow and controlled or smooth an automatic (Anderson, 1983, 2015; DeKeyser, 2007a; DeKeyser & Criado, 2013).

During the cognitive stage, learners "develop a declarative encoding of the skill" (Anderson, 2015, p. 211). In the case of language learning, declarative knowledge (knowing that) refers to the explicit knowledge of morphosyntax and word meanings (DeKeyser & Criado, 2013). Defined as conscious and metalingual knowledge about the target language (R. Ellis, 1994), explicit knowledge can be acquired either through "perceptive observation" and/or explicit explanations (e.g., explicit feedback) from a more expert user (DeKeyser, 2007a, p. 98). Given that learners would acquire new linguistic knowledge during the cognitive stage, they would possibly benefit more from direct and/or metalinguistic feedback. More specifically, because learners' use and access of their knowledge during this stage "require a great deal of attention—or what cognitive psychologists call 'controlled processing'" (McLaughlin, 1990, p. 114), one can possibly argue that direct corrective feedback is better suited at lightening the "heavy burden" that "knowledge compilation" is already placing on learners' working memory (Anderson, 1981, p. 26).

To pass to the associative stage, learners need to repeatedly activate and draw upon the available declarative knowledge (Anderson, 2015). The associative stage is characterized by a gradual decrease of errors and development of procedural knowledge (i.e., knowing how). Procedural knowledge differs from declarative knowledge in terms of representation and use (i.e.,

access). It is no longer represented as "bits and pieces of information" that are retrieved slowly and effortfully; rather, it is available as "ready-made chunks" that can be acted upon smoothly and rapidly (DeKeyser, 2007a, p. 98).

The last stage in the learning trajectory, and the desired outcome, is the autonomous stage, whereby procedural knowledge is automatized and routinized (Anderson 2015; McLaughlin, 1990). DeKeyser (2007a, 2007b) advances that while proceduralization is not particularly time-consuming, the final consolidation and fine-tuning of procedural knowledge into automatized knowledge are relatively long. More importantly, they require extensive and repeated practice so that learners' performance becomes faster and more accurate. Arguably, it is during the last two stages (i.e., the associative and autonomous stages) that learners develop greater control over their partially acquired knowledge.

More importantly, the gradual transition from one stage to another is facilitated by practice in tandem with extensive and repeated corrective feedback (Lyster, Saito & Sato, 2013). As highlighted by DeKeyser (2007b), "a large amount of practice is required to decrease time required to execute the task (reaction time), the percentage of errors (error rate), and the amount of attention required … This practice leads to gradual automatization of knowledge" (p. 98- 99). Corrective feedback, be it input-providing or output-prompting, has the potential of assisting learners acquire declarative knowledge and eventually solidify it and restructure it.

Besides underscoring the importance of explicit knowledge and extensive practice in L2 learning, proponents of skill acquisition learning theories have also pointed out that automatization is skill-specific (DeKeyser, 2007a, 2007b; Segalowitz & Lightbown, 1999). Skill-specificity is borne out in studies by DeKeyser (1996, 1997; DeKeyser & Sokalski, 1996) which show that extensive practice of morphosyntactic rules in either comprehension or production

activities result into considerable gains in the skill practiced but less so on the reverse skill. These findings run against Krashen's predictions that once knowledge is acquired via comprehension, it becomes available in both comprehension and production (DeKeyser, 1997; DeKeyser & Sokalski, 1996). They are, however, more in line with Transfer-Appropriate Processing, according to which knowledge is easier to retrieve when "the cognitive processes that are active during learning are similar to those that are active during retrieval" (Lightbown, 2008, p. 27; Segalowitz & Lightbown, 1999). In other words, if learners help practice grammatical rules and receive WCF in drill-type exercises, they would be more successful at retrieving and using their learned knowledge in similar activities and less so in more communicatively oriented tasks, for the latter engage different cognitive processes.

Insights from Skill Acquisition theories (particularly skill-specificity) and from Transfer-Appropriate Processing can be used to explain French L1 learners' attested "difficulty in using their grammatical knowledge in writing context," in spite of the importance accorded to grammar teaching and learning in French classes (Boivin & Pinsonneault, 2016, p. 109, personal translation from French). Boivin and Pinsonneault criticize the "partitioning" ("cloisonnement") of grammar and writing instruction and propose a more comprehensive, theoretical model that "gives grammar a more central position" in the writing classroom (p. 113). Although it is not explicitly advanced in their report (Boivin & Pinsonneault, 2014), it is possible to argue that written corrective feedback affords opportunities for contextualized practice, whereby focus on grammar occurs as learners attempt to make their texts clearer and more accurate, particularly when they are required to revise their drafts. WCF has the potential of leading learners to restructure and consolidate their grammatical knowledge as they are engaged in meaningful and contextualized tasks.

In light of the different theoretical arguments outlined so far, WCF can be considered as a form-focusing device that is likely to trigger the cognitive processes of attention, noticing and comparison (Schmidt, 2001; Sharwood-Smith, 1981; Swain, 1985, 1995). WCF has, thus, the potential of enhancing the transformation of declarative knowledge into procedural and eventually automatized knowledge (DeKeyser, 2007a). By preventing the formulation of erroneous hypotheses (Swain, 1995), it contributes to the acquisition of new knowledge and the consolidation and restructuring of partially acquired knowledge (Polio, 2012; Williams, 2012).

The theoretical claims, whether they originated in L2 writing literature or they were propounded by SLA researchers, have motivated a growing number of empirical and descriptive studies that seek to describe WCF practices and to examine the extent to which WCF facilitates second language acquisition and writing. A brief overview of WCF research is presented next.

1.3 Research on WCF

Studies on WCF are informed by L2 writing research and Second Language Acquisition (SLA) research. Each one of these fields of inquiry has its own theoretical underpinnings, pedagogical motivations and methodological tools (Ferris, 2010; Ortega, 2012). As noted by Ferris (2010), "although L2 writing and SLA researchers look at similar phenomena, often (but not always) in similar ways … they do not necessarily ask the same questions" (p. 188). Differences in research foci and by extension in methodological choices account for the lack of consensus that characterizes the extant research on WCF (Guénette, 2007; Storch, 2010). Proponents and advocates of WCF have argued about its overall effectiveness and about the relative merits of different WCF types, yet they have not reached conclusive answers about the usefulness of WCF and the differential impacts of its types.

L2 writing research focuses primarily on identifying the conditions and tools that would help L2 learners improve the overall quality of their texts. L2 writing researchers are, thus, more concerned with global (i.e., macrostructure) issues such as content, organization and rhetorical devices than local (i.e., microstructure) issues such as grammar or mechanics (Bitchener & Ferris, 2012; Hyland & Hyland, 2006). It follows that in many studies conducted in L2 writing classrooms, written corrective feedback, i.e., feedback on form, is either compared to feedback on content or combined with it (e.g., Fathman & Whalley, 1990; Kepner, 1991; Semke, 1984). Furthermore, L2 writing researchers are more interested in examining the usefulness of WCF as a 'revising' tool; consequently, they focus more on learners' revision following the provided WCF (Manchòn, 2011).

In contrast, SLA research is more interested in investigating the factors and instructional interventions (be they reactive or proactive) that are likely to facilitate L2 development. SLA-framed studies on WCF are usually conducted in L2 language classes and are mostly concerned with investigating the effectiveness of corrective feedback on linguistic errors of morphology, syntax and grammar (e.g., Ellis, Sheen, Murakami, & Takashima, 2008; Stefanou & Révész, 2015; Van Beuningen, *et al.*, 2008; 2012). Whereas L2 writing researchers examine the efficacy of WCF as a revising tool, those working from an SLA perspective are more concerned with the '*language learning potential*' of WCF, i.e., to its facilitative role in the L2 acquisition process (Manchòn, 2011). Subsequently, their investigations go beyond the revised texts, though these might be considered, and pay more attention to learners' accuracy on new texts, i.e., in immediate and/or delayed post-tests (e.g., Sheen, Wright, & Moldawa, 2009; Truscott & Hsu, 2008).

Regardless of the research focus, empirical investigations on WCF are also motivated by two major interconnected questions. The first one aims at identifying whether WCF brings about

L2 development as reflected in short- and long-term accuracy. The second research question seeks to determine the differential impacts of specific types of WCF. In terms of research design, WCF studies fall into one of two broad strands: (1) early, longitudinal studies, in which the WCF treatment is provided over a sustained period of time (e.g., Lalande, 1982; Semke, 1984) and (2) recent examinations which operationalized WCF mostly as one-shot treatment (e.g., Storch, Bitchener & Knoch, 2005; Sheen, *et al.*, 2009).

The first group of empirical studies, i.e., the longitudinal studies, have provided conflicting findings about the overall effectiveness of WCF (regardless of its form). Their inconclusive results may be attributed to different methodological limitations such as lack of control groups, dissimilar time-on-task and incomparable accuracy measurements (Li & Brown, 2015; Storch, 2010). However, the second group, i.e., the more recent one, has yielded mounting evidence for the usefulness of WCF in improving learners' short-term and long-term accuracy. Notwithstanding, the second research focus (i.e., which WCF type is more effective) is still under debate. A more zoomed-in scrutiny at WCF research reveals an unbalanced focus on certain operationalizations of WCF. In fact, direct WCF, i.e., WCF that provides learners with the correct target form, has received the lion's share of empirical examinations (Bitchener & Knoch, 2008, 2009, 2010a, Bitchener, Young & Cameron, 2005; Ellis, *et al.*, 2008). On the other hand, indirect WCF, i.e., WCF that prompts learners to self-correct, remains relatively unexplored. The few studies that examined indirect WCF have resulted into contradictory results about the efficacy of its different operationalizations (Chandler, 2003; versus Ferris & Roberts, 2001). Likewise, studies that compared direct to indirect WCF have yielded conflicting findings (Lalande, 1982; Robb, Ross, & Shortreed, 1986; Van Beuningen, *et al.*, 2008; 2012).

The conflicting findings regarding both questions can be attributable to issues in research design and methodological choices (Ferris, 2003, 2004, Guénette, 2007; Storch, 2010). In fact, however tightly controlled, those investigations are marked by 1) the limited focus on specific linguistic features (mostly determiners and past verb forms), 2) the problematic short duration of WCF (in many studies, WCF is provided as a one-shot treatment) and 3) the absence of revision after the provision of feedback. A thorough discussion of those methodological limitations will be provided in Chapter 2.

1.4 General research objective

The role and place of written corrective feedback in L2 classrooms are underscored by different researchers who have questioned the insufficiency of Krashen's comprehensible input. Drawing on cognitive-interactionism and skill acquisition theories, these researchers argued that WCF can facilitate noticing, activate the processes of hypothesis testing and aid learners in proceduralizing, restructuring and automatizing their knowledge (e.g., DeKeyser, 2007a; Swain, 1995). Against this backdrop, different empirical investigations have set out to examine the facilitative role of WCF in developing L2 acquisition (Bitchener & Ferris, 2012; Chandler, 2003; Van Beuningen *et al.*, 2012). Their aggregate findings point to its effectiveness in facilitating L2 acquisition, at least as far as particular language features are concerned, thus underlining its potential for learning (e.g., Bitchener, *et al.*, 2005; Ellis *et al.*, 2008; Sheen, 2007; Shintani, Ellis & Suzuki; 2014).

Notwithstanding their agreement on the instrumental roles played by corrective feedback, there is no consensus on which WCF technique works best. Findings from the extant WCF research, while concurring on its overall effectiveness, have not yet provided conclusive answers to some pending questions. This lack of straightforward answers about the relative merits of

WCF techniques motivates the increasingly growing WCF research. As an attempt to contribute to this ongoing research, the current study sets out to explore the relationship between L2 accuracy and WCF.

Chapter 2

Conceptual and Theoretical Frameworks

2 Conceptual and Theoretical Frameworks

The purpose of this chapter is to review the theoretical and empirical frameworks for the current study. It is divided into four major sections. In the first section, the two key terms of L2 acquisition and written corrective feedback (WCF) are defined and the different operationalizations of WCF are delineated. In the second section, an overview of the various theoretical arguments underpinning the WCF types is presented. Some of the major variables that moderate the efficacy of WCF such as errors' amenability to WCF and learners' individual differences are also discussed. In the third section, a detailed overview of the extant empirical research on WCF is provided; this overview is divided into three distinct parts: (1) teachers' WCF practices; (2) learners' WCF preferences and (3) differential impacts of WCF techniques. In the latter, both early and recent studies are scrutinized and their major findings as well as their methodological limitations are highlighted. This empirical overview sets the ground for the specific research questions that are presented in the fourth section of this chapter.

2.1 Defining key terms

Given that the current study seeks to explore the relationship between WCF and second language acquisition, it is necessary to delineate the two constructs of interest. These are: (1) L2 acquisition and (2) written corrective feedback. An overview of the different conceptualizations associated with these constructs is presented next.

2.1.1 Second language (L2) acquisition

As a concept, L2 acquisition has been defined in at least two distinct ways (de Bot, 1996, R. Ellis, 1994; Lyster, 2004). Acquisition refers to (1) learning and internalization of new forms and to (2) increased control and mastery of partially acquired knowledge, which is evidenced in

effortless retrieval and accurate use of the target language. The second conceptualization of L2 acquisition puts more emphasis on increased and systematic accuracy. Along those lines, Housen and Pierrard (2005) propose that increased control of partially acquired knowledge can be achieved via two sequential, yet overlapping, learning processes. The first one is knowledge modification whereby learners restructure, extend and fine-tune their L2 knowledge, "particularly the deviant, non-target like aspects of their knowledge and performance," which in turn would lead to increased accuracy (p. 6). The second process is knowledge consolidation, whereby learners solidify their knowledge as a result of repeated retrieval and deeper processing. Knowledge consolidation is often reflected in learners' accruing ability to "use the L2 with greater ease and for a wider range of tasks and functions" (p. 6). In general, acquisition refers to either the learning of new forms and/or the mastery of previously learned knowledge as illustrated in increased accuracy and automatized, fluent access to the L2 knowledge.

Irrespective of how L2 acquisition is conceptualized, it seems safe to argue that until recently, the extant SLA research has been influenced by two biases. The first one, and actually the first to be challenged, refers to the role attributed to output, or language production, in the process of L2 acquisition (Skehan, 2002). Traditionally, output was seen as a by-product of the acquisition process and as a reflection of learners' developing linguistic knowledge, as such it "ha[d] little significance for theory" (Skehan, 2002, p. 85). This lack of interest is exemplified in Krashen's (1981) assertion that "speaking and writing are not essential to acquisition" (p. 181). L2 acquisition depends largely on input, i.e., linguistic data to which learners are exposed (e.g., Krashen, 1981, 1985). Accordingly, teachers were discouraged from "forcing early production", i.e., before learners have developed enough L2 competence through "comprehensible input," because pushed production can be "the single most anxiety-provoking thing about language

classes!" (Krashen, 1985, p. 72). This view was contested by interactionist researchers (Long, 1996; Pica, 1994; Swain, 1985, 1993, 1995) who highlighted the insufficiency of input in furthering L2 acquisition and who underscored the importance of interaction, negotiation (of both meaning and form) and of language production in promoting L2 development.

The second bias consists of the almost-exclusive interest in online oral performance in comparison to the minimal attention given to L2 writing (Byrnes & Manchòn, 2014; Ortega, 2012). Learners' performance on writing tasks was often ignored on the assumption that their written output is monitored and unspontaneous, failing thus to reflect learners' "tacit linguistic knowledge" or their "online ability for use"- both of which were hypothesized to provide clear and strong evidence for real L2 acquisition (Ortega, 2012, p. 405). Consequently, and due to its monitored quality, "written evidence takes a back seat compared to oral evidence in SLA research programs" (Ortega 2012, p. 405). However, the thriving interface between the fields of L2 writing and SLA has triggered interest in the "instrumental role that writing can play in the acquisition of a second language" (Harklau, 2002, p. 345; also Ferris, 2010; Leki, Cumming & Silva, 2008; Manchòn, 2009; 2011). Along those lines, it has been argued that writing provides rich learning sites "wherein learners analyze and consolidate second language knowledge that they have previously (but not yet fully) acquired" (Cumming, 1990, p. 483). Pointing to the offline character of writing, to its slow pace and permanent record, Manchòn and Roca de Larios (2007), as well as Williams (2012) contend that engaging learners in writing activities has the potential to facilitate their language learning. The reconceptualization of writing as a vehicle of L2 development has fueled an increasingly growing interest in the learning potential of written corrective feedback (Ortega, 2012; Williams, 2012). The next section defines WCF and describes its various operationalizations.

2.1.2 Written corrective feedback

Feedback in writing is an umbrella term that refers to any response to student writing. Those responses may vary according to different dimensions: in terms of target, feedback can target macrostructure (e.g., organization; focus) and/or on microstructure issues (e.g., quality of language). In terms of its source, it may be provided by a peer and/ or by a teacher. When it comes to its modality, it can be oral and/ or written; it can also be provided electronically. Partially overlapping with these operationalizations, feedback in SLA literature is defined as "any indication to the learners that their use of the target language is incorrect" (Lightbown & Spada, 1999, p. 171). The SLA-informed definition is thus more geared toward linguistic accuracy. In accordance with this SLA- anchored conceptualization, the current study examines WCF which is provided by the teacher and which focuses on learners' erroneous use of specific linguistic forms.

In addition to their varied conceptualizations, L2 writing and SLA informed studies have used an array of terms to refer to WCF. Examples of those terms include *negative evidence*, *negative feedback (NF), grammar/ error correction* and *error treatment*. As argued by Schachter (1991) and Leeman (2007), these labels reflect different research paradigms in SLA: in fact, L2 pedagogy studies choose the term *corrective feedback*; investigations that are framed by linguistic approaches opt for *negative evidence*/ *data*, while those informed by cognitive psychology use *negative CF*. More nuances are revealed by a brief overview of those definitions and their referents. First, *negative evidence,* as opposed to positive evidence, refers to information or hints about what is not allowed in the target language (White 1991, 2003). *Negative evidence* can be direct (i.e., receiving signs from the interlocutor, be it from a teacher or a peer that the rendition is inacceptable) or indirect (i.e., not hearing the form in the input) (Leeman, 2007; White, 2003). As such *negative evidence* englobes corrective feedback as one of its

manifestations; subsequently, it is a more encompassing concept. Thus, it will not be used. Furthermore, *negative feedback*, which refers to the same concept as corrective feedback, will also be dismissed due to the pejorative connotations which are encoded in "negative". In addition, *grammar correction* is inconvenient because it excludes feedback that targets non-grammatical errors (such as spelling and lexical choices). Finally, *error correction* is also avoided because it is based on the assumption that flagged errors would be automatically corrected. However, feedback would not necessarily lead to the elimination of errors, thus the inappropriateness of the word *correction* (Long, 1996). Taking into consideration that this study is framed within cognitively-oriented L2 pedagogy, *written corrective feedback* and *written error treatment* are found to be neutral and appropriate labels; they are thus used interchangeably in this project.

Exploring the relationship between WCF and L2 development (which is the general objective of this study) requires a delineation of the various WCF types. According to Bitchener and Ferris (2012) and Ellis R. (2009), WCF can be either direct (i.e., input providing) or indirect (i.e., output eliciting or prompting). Direct WCF can refer to the provision of the correct form, as shown in the following example.

Example 1:

She regretS what she said.

Another way of providing direct WCF can be through the crossing out of unnecessary elements, as illustrated in example 2. Direct WCF can also consist of inserting a missing word, as exemplified in example 3.

Example 2:

~~Being~~ belonging to two cultures creates many challenges.

Example 3:

She can no longer meet (*her*) mother's expectations.

In contrast, indirect WCF consists of signaling errors via coding (see example 4), underlining or circling of the non-target form (see example 5) or indicating the occurrence of an error in the margin (see example 6) (e.g., Van Beuningen, *et al.*, 2012).

Example 4:

French is X difficult language. (determiner)

Example 5:

This books is worth-reading.

Example 6:

Throu her journey, she have faced many opstacles. *(spelling; SVA; spelling)*

Both direct and indirect WCF may vary according to their explicitness, in that they may or may not be coupled with written and/ oral metalinguistic information or clues (e.g., Sheen, 2007; see examples 7 and 8).

Example 7: Direct WCF + metalinguistic explanation (ME)

She stealed a loaf of bread. *stole: steal is an irregular verb as such its past form does not take an "ed."*

Example 8: Indirect WCF + metalinguistic explanation (ME)

She <u>stealed</u> a loaf of bread *verb form; steal is an irregular verb; its past form does not take an "ed."*

WCF also varies in terms of its focus: it can be (a) unfocused (comprehensive or untargeted) when targeting all or most errors (e.g., Van Beuningen, *et al.*, 2012), (b) highly focused (targeted) when addressing only one or two linguistic features (e.g., Sheen, 2007) or (c) mid-focused when selectively identifying a limited number of features (R. Ellis, 2009). Although most empirical studies have mostly compared highly focused WCF to comprehensive WCF, their dichotomous classification does not necessarily reflect the pedagogical realities of L2 classrooms, in which teachers would often flag many errors types rather than moving between the two extremes of comprehensive or highly focused WCF (e.g., Ammar *et al.*, 2016; Guénette & Lyster, 2013). It would be more accurate, thus, to conceptualize WCF techniques along a continuum of focus (see Figure 2)

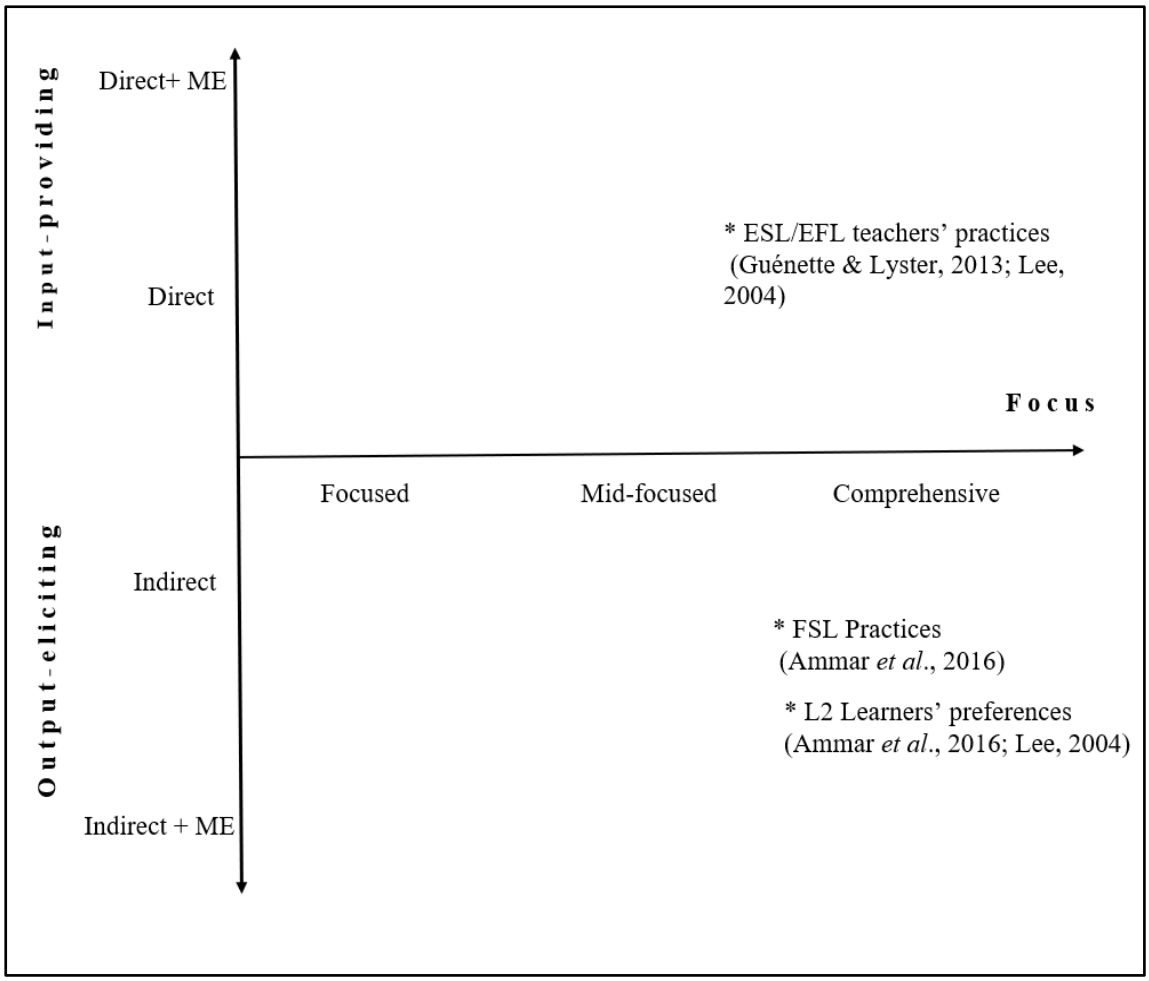


Figure 2. Different types of written corrective feedback

These different operationalizations of WCF have been compared in various studies (e.g., Bitchener & Knoch, 2008, 2009, 2010a; Ellis, *et al.*, 2008; Sheen, 2007; Stefanou & Révész, 2015 Van Beuningen *et al.*, 2008, 2012), yet the extant WCF research reveals a lack of agreement on the ideal WCF technique. Advocates of error treatment, be they SLA theorists or L2 practitioners, debate whether WCF should be input-providing or output-prompting and whether it should be provided comprehensively or selectively. Arguments related to these two dichotomies are discussed next.

2.2 The effectiveness of WCF types: theoretical arguments

Besides discussing the relevance and usefulness of written corrective feedback, SLA and L2 writing researchers have posited different theoretical and pedagogical arguments in favor of the various types of WCF. Their discussions have -albeit to a lesser extent- highlighted the need to take into consideration other moderating variables, including learners' individual differences and error type. Against this backdrop, the theoretical underpinnings of (1) direct versus indirect and (2) focused versus comprehensive WCF are exposed. These will be followed by a discussion of (3) the mediating effects of error types and (4) learner-related factors.

2.2.1 Direct versus indirect WCF

The relative merits of direct versus indirect WCF have been debated by various WCF researchers. Some have argued for the superiority of direct WCF claiming that it is less ambiguous because it clearly juxtaposes the learner's erroneous output and the corresponding target form (Chandler, 2003). Others have contended that indirect WCF allows for "guided learning and problem solving" (Lalande, 1982, p. 140), and pushes learners to think about their errors, and to try out other alternatives (Ferris, 1999, 2002). In the same vein, Bitchener and Knoch (2008) explain that by triggering the processes of hypothesis testing and cognitive comparisons, indirect WCF "promotes the type of reflection that is more likely to foster long-term acquisition" (p. 415).

Standing on middle ground, Ellis, *et al.* (2008) explain that the use and effectiveness of WCF techniques differ depending on the desired outcome of the learning process. They argue that input-providing WCF (i.e., direct WCF) can assist with the internalization of new forms; in contrast, output-eliciting WCF (i.e., indirect WCF) can potentially facilitate control over partially acquired knowledge. This seems to be in line with Swain (1985) and de Bot (1996) who, despite

not advocating for any specific WCF type, hypothesize that engaging learners in output activities results into increased control over partially- acquired knowledge. Consecutively, it can be assumed that output- eliciting techniques, such as indirect WCF, can be more effective than input- providing ones, such as direct WCF, in facilitating the modification and consolidation of previously internalized knowledge.

Support for indirect WCF can also be found in descriptive studies into learners' preferences (e.g., Ammar, *et al.*, 2016; Chandler, 2003; Ferris & Roberts, 2001, Lee, 2004). These studies reveal that students show a clear preference for indirect WCF because it allows them to identify the types of errors they have made. However, it should be noted that learners have either expressed reservations about indirect coded WCF (Lee, 2004) or indicated that they favor indirect WCF that is combined with metalinguistic clues over coded indirect WCF (Ammar, *et al.*, 2016) since the codes used can be ambiguous and quite vague.

Along those lines, the oral CF literature has provided more convincing support for the potential superiority of indirect CF over direct CF (e.g., Ammar, 2008; Ammar & Spada, 2006; Lyster, 2004, see also the metanalysis by Lyster, Saito & Sato, 2013). In contrast, empirical research on WCF has not yielded conclusive answers yet. Likewise, inconclusive arguments are found in the discussions about the differential impacts of focused versus unfocused WCF.

2.2.2 Focused versus unfocused WCF

An equally divisive issue concerns the focus of WCF, i.e., whether it is pedagogically sound to provide comprehensive WCF that targets all errors or focused WCF that addresses a predetermined feature (e.g., the simple past). Advocates for selective or focused WCF argue that it allows teachers to "build students' awareness and knowledge of their most serious and frequent grammar problems," as suggested by Ferris (1999, p.7). Accordingly, some researchers (e.g.,

Ellis *et al.*, 2008; Sheen, 2007a, 2007b; Shintani, Ellis, Suzuki, 2014; Stefanou & Révész, 2015) have adopted an extreme version of such a selective approach by arguing for the usefulness of highly focused WCF, i.e., WCF that targets one language form. In the same vein, Ellis *et al.*, (2008) argue that comprehensive WCF can be overwhelming and that focused WCF is more likely to activate the cognitive processes of attention and noticing, leading thus learners to restructure their linguistic representations. Focused WCF is thus "better equipped to produce positive results" (Ellis *et al.*, 2008, p. 356). Similar arguments are embraced by Sheen (2007) who, drawing on oral CF literature, claims that WCF that intensively targets a single linguistic feature "can have a beneficial effect on interlanguage development" (p. 256). However, other researchers have convincingly pointed out that highly focused WCF can increase the salience of the target feature (Bruton, 2009; Storch, 2010; Xu, 2009). In other words, it is possible that participants become aware of the language form targeted by the experimental intervention and may consciously pay more attention to their use of that form in the post-tests.

A closer look at empirical WCF studies precludes any definitive answer as far as the differential impacts of focused versus comprehensive WCF are concerned. An increasing number of recent inquiries have implemented and investigated highly focused WCF (e.g., Bitchener & Knoch, 2008, 2009, 2010a, 2010b, Sheen, 2007; Stefanou & Révész, 2015), and fewer studies have examined comprehensive WCF (Van Beuningen *et al.*, 2008, 2012; Truscott & Hsu, 2008). Furthermore, only two studies (i.e., Ellis, *et al.*, 2008; Sheen, Wright, & Moldawa, 2009) have systematically compared the two approaches, yielding, however, conflicting findings.

Beside discussing which WCF type is more effective, SLA and L2 writing theorists have been interested in how the effectiveness of WCF may be mediated by linguistic as well as learner variables, which will be discussed in the next sections.

2.2.3 Mediating variables

Most empirical research on WCF has examined the effectiveness of different WCF techniques and has overlooked the mediating effects of error and learner related variables. Yet, in light of findings from research on form-focused instruction and on oral corrective feedback (e.g., Ammar & Spada, 2008; Li, 2014; Trofimovich, Ammar, & Gatbonton, 2007; Yilmaz, 2013a), there is a burgeoning interest in how the effectiveness of a WCF technique hinges upon error type and/or learners' individual differences- both of which are discussed in more detail next.

2.2.3.1 Errors amenability to WCF

Ironically enough, the interest in errors amenability to WCF seems to be triggered by Truscott, who is described as one of the most vocal critics of WCF. To justify his skepticism about the usefulness of error treatment, Truscott (1996) advances that "there is some reason to think that syntactic, morphological, and lexical knowledge are acquired in different manners;" he then adds that "probably no single form of correction can be effective for all three" (p. 343). In one of her counterarguments, Ferris (1999) concedes that not all errors are affected, in the same degree, by all types of WCF. Conversely, she hypothesizes that indirect WCF might be more useful for what she labels as "treatable" errors and that direct WCF is more effective for non-treatable errors. According to Ferris, treatable errors "occur in a patterned, rule-governed way" and include problems such as "subject-verb agreement, run-ons and comma splices, missing articles, verb form errors" (p.6) while non-treatable errors include "lexical errors and problems

with sentence structure, including missing words, unnecessary words, and word order problems" (p.6)[4].

Ferris's hypothesis about possible interactions between WCF techniques and errors was tested in her descriptive study (2006). In this longitudinal examination, ESL students received WCF on different errors that were grouped into five large categories. These included verbs, noun endings, determiners, word choice and sentence structure. Given that participating teachers did not follow the coding schema proposed by Ferris and did not consistently use indirect coded WCF as expected, students received mostly direct WCF for word choice and sentence structure and were provided with indirect coded WCF for verbs, noun endings and determiners. A comparison of students' essays at the beginning and at the end of the study indicated "highly significant" improvement for verb errors only (p. 90) while their performance on determiners and noun-endings was slightly worse. These three errors were classified as treatable (Ferris, 1999, 2006) and were identified by indirect coded WCF. Likewise, students' performance on what was categorized as untreatable errors and for which they received direct WCF was not homogenous either. Their gains in lexical choices approached statistical significance but their scores in sentence structure underwent slight decrease. Overall, these patterns do not lend strong support to the claims that learners' differing levels of improvement were "based on whether errors were 'treatable' or 'untreatable' (…), and this distinction may possibly be attributed to widely disparate teacher feedback strategies" (Ferris, 2006, p. 98).

[4] It is to be noted that some of the errors that are classified as untreatable (such as word order) are more rule-governed than others (such as lexical choices), which makes Ferris's dichotomous classification somehow vague and not straightforwardly applicable.

Albeit interesting, the results from Ferris (2006) should be interpreted with caution for this study did not include a control group; neither did it include clearly defined experimental conditions. Some of these methodological limitations were overcome in Van Beuningen *et al.* (2012), who examined the moderating effects of error types. However, Van Beuningen *et al.* did not use the treatable and untreatable error categorization. Rather, they operationalized errors as being (a) grammatical errors (i.e., determiner errors, inflectional errors, word order errors, omissions of necessary elements, additions of unnecessary elements, pronominal errors) and (b) lexical (word choice) and orthographical errors (spelling and punctuation). Their findings indicate that both techniques resulted in increased accuracy in revised and new texts and that their effectiveness in new texts was mediated by the issues targeted. While direct WCF was superior to the indirect WCF on grammatical issues, indirect WCF was more effective than direct CF for lexical and orthographical errors. Given that Van Beuningen grouped different grammatical features in one large category, it remains unclear how specific error forms are affected differently by WCF. Only three mid-focused studies (i.e., Benson & DeKeyser, 2018; Bitchener *et al.*, 2005, Shintani *et al.*, 2014) have systematically investigated how specific language forms can be affected by WCF techniques.

Bitchener *et al.* (2005) indicate that WCF which was combined with oral conferencing was beneficial for accurate uses of the simple past and the definite determiner, but no such effect was found for the use of prepositions. Drawing on Ferris' distinction, Bitchener *et al.* explain that the simple past and definite determiners "are more readily 'treatable'" and less idiosyncratic than prepositions (p. 201). Examining two rule-patterned forms, which were the simple past and the present perfect, Benson and DeKeyser (2018) report that direct WCF resulted in long-term accuracy gains for the simple past but less so for the present perfect, attributing such differences

to learners' developmental readiness and varying degrees of previous declarative knowledge of both features. Also examining two language forms, which can be classified as treatable according to Ferris's (1999) dichotomy, Shintani *et al.* (2014) show that the past hypothetical conditional was more amenable to direct WCF than the indefinite determiners. In light of their findings, Shintani *et al.* argue that "what constitutes "treatability" is not just a question of whether or not a feature is rule-based but also the complexity of the rule-based structure" (p. 123).

Taken together, findings from these few studies suggest that limiting the discussion of error amenability to WCF to the treatable and untreatable distinction can be, at this stage, premature. More importantly, these inquiries underscore the need to examine how varying degrees of complexity affect the responsiveness of linguistic categories to different instructional techniques (e.g., DeKeyser, 1995; Robinson, 1996; Spada & Tomita, 2010). Although there is a lack of consensus on the conceptualizations and operationalizations of complexity, a comprehensive overview is provided by Spada and Tomita (2010) who explain that grammatical complexity has been interpreted from three overlapping perspectives: psycholinguistic, linguistic and pedagogical dimensions.

The first dimension has to do with the extent to which "a feature is acquired early or late or is more or less difficult to process;" as such complexity, in the psycholinguistic sense, is more related to learners' developmental readiness (Spada & Tomita, 2010, p. 266). From a linguistic perspective, DeKeyser (2005) explains that complexity can be determined by at least three factors: complexity of form, complexity of meaning and complexity of the form-meaning relationship (DeKeyser, 2005). The complexity of form refers to the number of choices in using different morphemes (e.g., adjective endings in French); the complexity of meaning has to do with the level and degrees of abstraction (e.g., determiners in English; grammatical gender in

French), and the complexity of form-meaning mapping corresponds to the transparency between a form and its meaning. According to DeKeyser (2005), transparency can be affected by three different factors, which are "redundancy, optionality, or opacity" (p. 8). For example, the plural marking on nouns and adjectives as well as the gender marking on adjectives in French is redundant (i.e., semantically unessential) as it is usually encoded in determiners. Null subject in Spanish is optional; while the morpheme "s" in English is opaque given that it has different meanings (i.e., it can be used to form the plural, to form the possessive or to inflect the verb with the third personal pronoun). On a pedagogical ground, complexity is related to how problematic a feature can be for learners (e.g., plural noun endings in French are persistent learning problems for French L2 learners as pointed out in Ammar *et al.*, 2016).

It is this pedagogical sense of complexity that is often invoked in many studies on focused WCF, most of which have examined the effects of WCF on two functional uses of English determiners (e.g., Bitchener, 2008; Bitchener & Knoch, 2008, 2009, 2010b, Sheen, 2007; Sheen *et al.*, 2009). As such, it remains unclear to what extent similar uses of other linguistic errors that vary in complexity are amenable to corrective feedback. The few studies that have investigated mid-focused (e.g., Benson & DeKeyser, 2018; Bitchener *et al.*, 2005) and comprehensive WCF (e.g., Ferris, 2006; Van Beuningen, 2012) have also rarely accounted for possible interactions between WCF techniques and the complexity of the targeted errors. It is still unclear whether and to what extent the responsiveness of errors to different WCF types would hinge upon the varying degrees of their complexity. Moreover, given that most studies, except for Van Beuningen *et al.* (2012) and Stefanou and Révész (2015), were conducted in English L2 contexts, it would be useful to explore how errors in other languages are affected by WCF and to what extent their amenability to WCF depends on their complexity.

Finally, and in addition to the question of whether the effectiveness of WCF varies across error categories and their levels of complexity, there is also the issue concerning how learners' individual differences, both cognitive and affective, mediate the effectiveness of WCF.

2.2.3.2 Learners' individual differences.

In many instructed SLA accounts, it is generally agreed upon that learners "differ in how successfully they adapt to, and profit from instruction" (Robinson, 2002a, p. IX; Robinson, 2002b; Skehan, 2002) and that learners' internal factors influence the cognitive processes activated by instructional treatments and the accruing learning outcomes (Ellis, 2010, 2012). Along the same lines, Ferris (1999) anecdotally noted that "there is tremendous variability in students' ability to benefit from grammar instruction and feedback" (p. 7). She and others hypothesized that such a differential success can be better described if WCF research accounts for learners' internal variables such as proficiency, aptitude, motivation and attitude (Ferris, 1999, Ferris & Hedgcock, 2005; Kormos, 2012; Storch, 2010).

To date, very few studies have systematically investigated some learner internal variables. They have either focused on the role of affective factors such as anxiety (Sheen, 2011) or on the impact of cognitive factors, namely language analytic aptitude (Sheen, 2007; Stefanou & Révész, 2015). Their results reveal that learners with lower anxiety levels and higher aptitude for language analysis benefitted significantly more from the feedback provided.

It is important here to highlight that researchers' interest in the role of language analytic aptitude is well-grounded in research and theorizing on aptitude for language learning in general. It is defined as "the capacity to infer rules of language and make linguistic generalizations or extrapolations" (Skehan, 1998, p. 204). In accordance with the theoretical claims that (W)CF engages learners in fine-tuning and restructuring their partially acquired knowledge (de Bot,

1996; Swain, 1990), it seems reasonable to expect that such a role would hinge upon learners' language analytic abilities. Findings from Sheen (2007) and Stefanou and Révész (2015) have provided some empirical support for such a hypothesis. Yet, it is worthwhile to note that these studies have examined direct WCF only and have been carried out in English as a second/ foreign language learning contexts. Subsequently, more studies on other WCF techniques (e.g., indirect WCF) provided for learners of other L2s would expand our understanding of the possible interactions between WCF techniques and learners' language analytic aptitude.

In the same vein, it would be interesting, both theoretically and pedagogically, to examine whether the effectiveness of WCF is impinged by learners' proficiency levels. As a variable, proficiency has not been adequately measured: in fact, in most WCF studies with the exceptions of Stefanou and Révész (2015) and Van Beuningen *et al.*, (2012), proficiency is determined based on learners' institutional level. Neither has it been accounted for as a potentially moderating variable. However, drawing on oral CF research, it is possible to wonder whether low proficiency learners would benefit more from output prompting (i.e., indirect) than input-providing (i.e., direct WCF) and whether high proficiency learners would gain from both techniques equally (as was the case in Ammar & Spada, 2006 for example).

2.3 Empirical research on WCF

The different theoretical debates about the role and usefulness of WCF in writing and language development have motivated several empirical studies which can be categorized in three overlapping strands: the first group of studies examine WCF practices of L2 teachers (Ammar *et al.*, 2016; Furneaux, Paran, & Fairfax, 2007; Guénette & Lyster, 2013; Lee 2004, 2008); the second set investigates learners' preferences vis à vis WCF (Ammar *et al.*, 2016; Chandler, 2003; Ferris, *et al.*, 2013; Ferris & Roberts, 2001; Lee, 2004; Leki,1991; Storch &

Wigglesworth, 2010). The third – and by far the largest – group evaluates the differential impacts of WCF techniques (Bitchener & Knoch, 2008; 2009; Chandler 2003, Ellis *et al.*, 2008; Ferris, 2006; Lalande, 1984; Sheen *et al.*, 2009; Van Beuningen *et al.*, 2008, 2012). It is important to discuss studies on teachers' WCF practices and on learners' preferences to explore how they can guide research examining the differential effectiveness of WCF techniques.

2.3.1 Teachers' WCF practices

Most studies examining L2 teachers' WCF practices were conducted in English as a second or foreign language contexts. Their common finding is that teachers show a preference for direct and comprehensive error correction. For example, Furneaux *et al.*, (2007) examined the WCF patterns of 110 EFL secondary teachers from five different countries. The instructors were asked to comment on the same student essay, which was provided to them by the researchers. Their feedback practices were then classified according to (a) the role assumed by the teachers and (b) the focus of the feedback. There were six roles identified which consist of the following: (1) providers of the correct form, (2) initiators of alerts to problematic areas of the text via codes, (3) supporters pointing out to successful uses of certain features, (4) advisors offering some indirect advice of possible improvements, (5) suggesters by providing alternatives, and (6) mutators who introduce changes to the text altering its original meaning. These roles can be grouped into two larger categories: direct style (roles 1, 5 and 6) and indirect style (roles 2, 3, and 4). In terms of the target of feedback, Furneaux *et al.* examined whether teachers targeted lexical, grammatical, style, semantic, discourse and mechanics errors. Results indicate that of the total 4637 feedback instances, 61.5% took the form of direct WCF while only 37% were indirect WCF. In terms of

focus, grammatical[5] errors (e.g., word order and morphosyntax), including morphosyntax, received the lion's share of teachers' annotations (45.2%), followed by lexical and semantic errors (11.3% and 10.25% respectively). More importantly, Furneaux *et al.*, (2007) reported an interaction effect between the type of corrective feedback and the targeted features. As explained by the researchers, providing direct WCF is "concentrated quite clearly on grammar (61.6% of providing comments fall into this category)" while indirect WCF is "more evenly spread out" (p. 83).

One obvious limitation, which was also acknowledged by the researchers, is that the study does not examine teachers' "real behavior" since the participating instructors responded to a simulated exercise whereby they provided feedback to a text that was not written by their students (p. 89). This limitation was, however, overcome in Lee (2008) who studied 26 secondary EFL instructors' annotations on 174 different texts submitted by their own students. Lee (2008) also analyzed teachers' annotations in terms of their type (i.e., direct or indirect WCF) and their target (i.e., form, including grammatical and lexical errors; content, organization). Similar to those reported in Furneaux *et al.*, his results show that teachers overwhelmingly relied on direct WCF; they provided the correct form 71.5% of the cases. In terms of its target, "94.1% of the teacher feedback focused on form (grammar and vocabulary)" (Lee, 2008; p. 76). Lee (2008) clarifies that teachers' practices are found to contravene their ministry's recommended "selective marking" and balanced use of direct and indirect techniques (p. 78).

[5] Furneaux *et al.* did not provide further details on what counted as grammatical errors in their coding scheme.

Similar patterns emerge from a study by Guénette and Lyster (2013) whose examination of WCF responses by 15 pre-service ESL teachers in Quebec shows a preference for direct feedback. Direct WCF, in fact, accounted for 70% of all errors flagged, followed by indirect WCF which was used only 24% of the time. In addition to the frequencies of each WCF type, Guénette and Lyster examined the relation between the CF technique and its focus (i.e., the error type). They reported that "three types of error (spelling, verbs, and word choice) accounted for 53% of all errors flagged" (p. 140). Of these three error types, errors in verbs (including tense and form) and word choice were mostly marked with direct WCF (71% and 81%, respectively). In comparison, 54% of spelling errors received direct WCF, while 45% of these errors were marked by indirect WCF, which points out to a relative resurgence of indirect feedback.

Of particular relevance to the current examination is a large-scale study conducted in Quebec's French-teaching classrooms by Ammar *et al.* (2016). The study consists of analyzing WCF practices of 26 French teachers in L1 and L2 contexts. The participating teachers provided WCF on 256 texts produced by elementary, secondary and adult French learners. Similar to Guénette and Lyster (2013), Ammar *et al.* examined the distribution of the different WCF techniques across various error categories. They also considered how these distributions vary across learners' proficiency levels. Unlike the patterns observed in ESL studies, Ammar *et al.* indicate that, overall, both French L1 and L2 teachers show a clear preference for indirect WCF, which is more in line with the *MEES'* recommendations. They reported that indirect WCF is the technique of choice for elementary teachers (71.7%), as well as secondary instructors (69.7%) and adult teachers (70.3%) (p. 11-12). More specifically, indirect coded WCF was used more in French as L1 contexts (86.6%) and in "accueil" (93%), while indirect uncoded WCF – underlining and circling – was more prevalent in immersion and enriched French settings

(84.5%). The latter, i.e., secondary enriched French program, is of more interest to the current study.

Zooming in on their findings about secondary L2 enriched contexts, it is clear that teachers varied their WCF practices with regards to the targeted errors. They usually provided direct corrections for lexical errors: 72.2 % of lexical non-target like forms were corrected directly. For syntactic errors, they used direct WCF 46.9% of the time and indirect WCF 51.7% of the cases. Their preference to use indirect WCF was more apparent for morphological and spelling errors, for which they provided indirect WCF 65% and 79.2% of the time (p. 59).

Concerning how teachers' WCF practices varied across learner proficiency levels, Ammar *et al.* (2016) indicated that overall teachers preferred non-coded indirect WCF for both high and low proficiency levels. However, they seemed to differentiate their practices when targeting errors of grammatical morphology. In fact, 56.9% of grammatical morphology errors by low proficiency learners were marked by direct WCF, compared to only 14.5% by high proficiency learners, whose errors were mostly identified by indirect WCF (84.3%).

To sum up, descriptive research into teachers' WCF practices has been mostly conducted in ESL contexts- an only exception is a study by Ammar *et al.* (2016) which examined French L1 and L2 classrooms in Quebec. What these inquiries reveal is that ESL instructors rely on comprehensive direct WCF, while French L1 and L2 teachers prefer indirect techniques.

In light of those findings, it seems fitting to determine whether teachers' WCF practices are congruent with learners' preferences since the effectiveness of WCF seems to depend on learners' perceptions. Put differently, "students' opinions and preferences for certain types and amounts of WCF affect their use of it for learning" (Amrhein & Nassaji, 2010, p. 97).

2.3.2 Learners' WCF preferences

Most studies examining learners' attitudes towards WCF have been conducted for English as a second and/or foreign language contexts. Overall, the extant research indicates that they expect their teachers to provide WCF (Ferris, 2002; Leki, 1991), that they "feel cheated" if their teacher does not give them CF (Schulz, 2001, p. 250) and that they strongly believe that corrective feedback facilitates language learning (Lee, 2004; Schulz, 2001). Furthermore, their perceptions about different types of WCF show minor variations across studies.

In an early study, Leki (1991) surveyed 100 ESL university students in freshman composition classes in an American university. The purpose of her survey was to identify students' attitudes towards their teachers' feedback and their preferences vis à vis the WCF techniques. Her results indicate that students (90%) are usually concerned about the accuracy of their written language and that they "equate good writing in English with error-free writing," which explains why they (70%) want their teachers to provide comprehensive feedback by targeting all errors (p. 203). Leki also shows that most surveyed students look over their teachers' annotations, with 89% of them reporting that they usually look "carefully" at the comments on grammatical and lexical errors (p. 206). When it comes to their preferred WCF technique, 67% of students expressed interest in indirect WCF that locates the error and provides "a clue on how to correct it" (p. 207). In contrast, only 25% indicate that they prefer direct WCF.

Similar findings are reported in Lee (2004), who administered a bilingual questionnaire (in both Chinese and English) to 320 EFL students from eight secondary schools in Hong Kong and who conducted follow-up interviews with 27 students. Students were asked about which WCF technique they prefer most, i.e., whether they like their teacher's feedback to be comprehensive or focused, direct or indirect. Results from the questionnaires show that 83% of them prefer

comprehensive feedback, in that they want their teachers to mark all their errors. These tendencies are further corroborated in the interview data, where students indicate that they want "to know what errors they had made" (p. 294). Like those in Leki (1991), most students (76%) surveyed in Lee (2004) show a preference for indirect feedback. In follow-up interviews, they explain that indirect coded feedback allows them to identify the types of errors they made; however, reservations concerning the intelligibility of codes are highlighted (p. 296). Some students remarked that using codes is "troublesome and time wasting" (p. 297).

More nuanced views were, however, expressed in Chandler (2003) who administered questionnaires to 21 ESL university students, in which they compared four WCF techniques used by their composition teachers. Those techniques consist of (a) direct WCF, (b) underlining plus description, (c) underlining only and (d) description only. More than 66% of the surveyed students contend that direct WCF can be the easiest to implement, and about 50% indicate that indirect WCF is the most helpful for it gives them clues about their errors. Likewise, students in Ferris and Roberts (2001) and in Ferris *et al.* (2013) show preference for indirect coded feedback.

Whether they are students of English as a second (e.g., Chandler, 2003) or a foreign language (e.g., Lee, 2004), in university (e.g., Leki, 1991) or high school (Lee, 2004) settings, most of the surveyed students show a preference for indirect WCF. This pattern of findings was also reported in the French L2 contexts as described in Ammar *et al.* (2016), who indicated that the interviewed French L2 learners (n=154) preferred indirect WCF (in this context underlining) that was combined with metalinguistic clues over coded indirect WCF.

Apart from research on teachers' WCF practices and learners' preferences, different studies have sought to identify whether WCF results into short- and long-term accuracy and to determine the relative merits of specific operationalisations of WCF.

2.3.3 Differential impacts of WCF techniques.

These inquiries can be grouped into two broad categories: (1) early, longitudinal studies, in which WCF is usually provided more than once (Ashwell, 2000; Chandler, 2003; Fazio, 2001; Kepner, 1991; Lalande, 1982; Robb, Ross & Shortreed, 1986; Semke, 1984; Sheppard, 1992), and (2) recent experiments (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2009b, 2010a, 2010b; Bitchener, *et al.*, 2005; Ellis, *et al.*, 2008; Sheen, 2007; Sheen *et al.*, 2009; Shintani *et al.*, 2014; Stefanou & Révész, 2015; Truscott & Hsu, 2008; van Beuingen *et al.*, 2008, 2012), which mostly operationalized WCF mostly as one shot treatment[6]. While the former group of empirical studies yielded contradictory results about the overall effectiveness of WCF, the more recent studies have provided more evidence for its usefulness in promoting L2 learning. These recent inquiries have failed, however, to reach a conclusive finding as far as the differential impacts of different WCF techniques are concerned. The overall conflicting results are due to issues in research design and methodological choices (Ferris, 2003; Guénette, 2007; Storch, 2010). Accordingly, the following sections will review early and more recent studies and will highlight some of their inherent methodological limitations.

2.3.3.1 Early inquiries into WCF.

Early inquiries into WCF share two characteristics: (1) they implemented comprehensive WCF (i.e., targeting most errors) and (2) they adopted longitudinal designs (i.e., treatment provided over more than one single writing task). However, they yielded contradictory results about the usefulness of WCF. They have also provided inconclusive answers about the relative

[6] In Sheen (2007), Sheen et al., (2009) and Stefanou and Révész (2015), feedback was provided twice. In Ellis et al., (2009), it was provided three times.

merits of direct versus indirect WCF. While some studies show that WCF brings about accuracy gains (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001), others have not yielded clear evidence in favor of WCF (e.g., Fazio, 2001; Kepner, 1991; Semke, 1984). These conflicting results can be largely attributed to differences and inconsistencies in the methodological designs of these studies (see Table 1). Early examinations differ in (a) when they measured accuracy, (b) how they controlled for intervening variables, (c) which writing tasks they implemented and (d) whether they had a control group or not. These are discussed next.

When accuracy is measured. Early studies on the usefulness of written corrective feedback fall under two categories. Some studies examined the effectiveness of WCF as an editing tool by assessing its impacts on revised texts only (Ashwell, 2000; Fathman & Whalley, 1990; Ferris & Roberts, 2001). Other studies sought to determine if, and to what extent, WCF brings about durable effects by examining learners' performance on new texts (Fazio, 2001; Lalande, 1982; Semke, 1984, Sheppard, 1992). While the second group i.e., studies that measured accuracy in new texts yielded mixed findings about WCF effectiveness in yielding long-term gains, studies that examined WCF efficacy in students' revisions demonstrated that learners benefited from WCF - regardless of its form (see Table 1). For the purposes of the present discussion, two studies from each group will be described.

A representative study from the first group is Fathman and Whalley (1990), which compared the effectiveness of four feedback treatments on the revised texts of 72 ESL students. Participants were randomly assigned to one of the following conditions: (a) content feedback, which consisted of general comments on organization and ideas (b) feedback on grammar, which took the form of indirect uncoded WCF through the underlining of all grammar errors such as verb forms, tenses, agreement and determiners; (c) content feedback coupled with indirect

uncoded WCF; and (d) no feedback. Students completed a 30- minute picture-cued story narrative and were handed in their annotated texts "few days later." They were then given 30 minutes to rewrite their texts using the feedback provided (p. 182). Two independent raters evaluated the two drafts, i.e. the original and the rewrite, in terms of overall accuracy (by counting the total number of grammar errors) and in terms of content by using a holistic scoring guide – yet, no interrater reliability was reported. Tabulating the mean scores for accuracy and content in both drafts, Fathman and Whalley indicate that all groups, regardless of their experimental condition, have statistically improved the content of their compositions. Interestingly enough, only the groups that received grammar feedback (i.e. groups b and c) have shown score gains in grammatical accuracy. The researchers conclude that "the identification of the location of the errors by the teacher appears to be an effective means in helping student correct their grammar errors" (p. 185).

In another study on the efficacy of WCF in helping learners revise their drafts more accurately, Ferris and Roberts (2001) examined the differential impacts of two indirect WCF techniques, which are coding and underlining. Three different classes (n=67) were randomly assigned to the three experimental conditions: (a) no feedback; (b) indirect coded WCF and (c) underlining. Students first completed in class, 50-minute response essays, and two weeks later revised and edit their drafts according to their treatment group. With the exception of the control group, students receive WCF that targeted five error categories. These consist of errors in verb tense or form, in noun endings, in determiners, in word choice and in sentence structure. Comparing the normalized error scores across drafts, Ferris and Roberts (2001) indicate that both WCF groups outperformed the control group at successfully editing their drafts and that there was no significant difference between the two WCF techniques. These results are, thus, in line

with those reported in Fathman and Whalley (1990) and in Ashwell (2000), which provides more evidence for the positive role of WCF in improving the overall accuracy of revised texts.

The significance and relevance of the findings yielded from revision studies have, however, generated considerable controversy among critics and advocates of WCF. In his various rebuttals of WCF, Truscott has argued that these studies "offer no measure of changes in students' ability to write accurately, i.e., their learning" (Truscott, 2007, p. 257). In other words, successful revisions do not necessarily imply that those gains will be maintained in new texts (e.g., Truscott & Hsu, 2008). Thus, findings about the usefulness of WCF in revision studies should be rather interpreted as signs of "superficial and possibly transient form of knowledge, with little value for actual use of the language" (Truscott, 1996, p. 345) than as evidence for changes in the "underlying, developing system," i.e., learners' interlanguage (p. 345). Ferris (2004), however, adopts a more nuanced perspective claiming that it is necessary to measure accuracy gains in the revised, as well as in the new texts because the former reflect "student uptake of corrections received" while the latter shows whether WCF has been integrated in learners' "developing competence in the L2" (p. 54).

Although some researchers assert the need to provide learners with opportunities for revision following teachers' WCF (Guénette, 2007; Storch, 2010), others point out that successful uptake of WCF does not necessarily imply that learning, i.e., fine-tuning and restructuring one's linguistic knowledge, has taken place. Uptake is defined as learners' reaction or use of the provided CF (Lyster & Ranta, 1997). It can take different forms, correct revision (incorporation of the correct form), incorrect revision (use of an incorrect form during revision), no revision, deletion or substitution (Ammar *et al.*, 2016; Ferris, 2006; Lyster & Ranta, 1997).

The inadequacy of uptake in measuring learning in the sense of development is foregrounded in research on interaction and oral CF, as well (Ammar & Spada, 2006; Gass, 2003; Lyster, 1998, 2004) and on WCF research (Bitchener & Ferris, 2012; Truscott, 2007). It follows that learners' successful incorporations of their teacher's WCF in subsequent revisions can simply be a sign of "mimicking" to use Gass's (2003) words (p. 236). This is particularly the case if the WCF consists of direct error correction. It is, however, less of an issue in the case of indirect WCF given that the latter does not provide learners with the target-like form, but rather prompts them to self-correct. Consequently, it can be argued that indirect WCF is less likely to result into unanalyzed mimicking.

In all cases, and regardless of WCF types, there is a general consensus in WCF research that "[w]hat is of particular interest to teachers is whether or not students can sustain this improvement on subsequent writing" (Guénette, 2007, p. 44). As such, more appropriate measures of long-term effects, i.e., in new texts, are needed to ascertain its effectiveness on L2 development/ acquisition (Bitchener & Ferris, 2012; Truscott, 2007).

Some of the early studies that examined the long-term effects of WCF include Lalande (1982) and Semke (1984). In the former, sixty intermediate students of German as a foreign language (FL) were divided into two groups: for one group, the teacher used direct WCF (i.e., he provided the correct forms), while the other group received indirect coded CF. Over the course of a ten-week session, students wrote five plot summaries of readings discussed in class; the first and last essays served as pre and post-tests. They received comprehensive WCF on three of the five compositions, i.e., the second, third and fourth assignments, and they were required to rewrite their first drafts incorporating the WCF they received. A comparison of pre-tests and post-tests in the two experimental groups revealed that the indirect WCF group outperformed the

direct WCF one in terms of error reduction; nonetheless, their difference did not reach statistical significance.

Different results emerge, however, from Semke's (1984) study of 141 German as a FL, in which he compared (a) feedback on content, (b) direct WCF, (c) feedback on content combined with direct WCF and (d) indirect coded WCF. Students completed weekly free-writes in class (a total of 10 entries) and received feedback according to their treatment condition. Of the four groups, only the group receiving indirect coded WCF was asked to revise and rewrite their journals. For the pre-test and post-tests, all students completed timed journal entries to help assess their fluency and accuracy and they took timed multiple-choice cloze tests meant to measure their proficiency level. Semke indicates that only the group that received feedback on content improved significantly in all three measures (i.e., in fluency, accuracy and overall proficiency). This led him to claim that WCF, be it direct or indirect, "does not improve students' writing skills in German as a second language, nor does it increase total competency in the language" (p. 200)- an argument that is out of sync with results reported in both Lalande (1982) and Sheppard (1992) (see Table 1). These mixed findings have contributed to the controversy about the efficacy of WCF in improving learners' accuracy in new texts. However, a closer look at those studies designs reveals that their disagreement is largely due to some major methodological limitations, which are discussed next.

Table 1

Representative Studies of Early Research on Written Feedback

Study	Written Feedback Types	Accuracy is measured in…	Effectiveness
Lalande, 1982	1. direct WCF 2. indirect coded WCF	New writing	Yes (no difference between the WCF types)
Sheppard, 1992	1. feedback on content 2. indirect coded WCF	New writing	Yes (no difference between the WCF types)
Ferris & Roberts, 2001	1. indirect coded WCF 2. indirect uncoded WCF 3. no feedback	Revised drafts	Yes (no difference between the WCF types)
Ashwell, 2000	1. content feedback then indirect uncoded WCF 2. indirect uncoded WCF then content feedback 3. content feedback and indirect uncoded WCF 4. no feedback	Revised drafts	Yes (no difference between the three treatment conditions)
Robb, Ross & Shortreed, 1986	1. direct WCF 2. indirect coded WCF 3. indirect uncoded WCF 4. *n* of errors provided in the Margins	Revised drafts	No
Semke, 1984	1. content feedback 2. direct WCF 3. content feedback + direct WCF 4. indirect coded WCF	New writing	No

Different intervening variables. In many of the early inquiries, the experimental groups differ on variables other than the dependent one, i.e., the provided WCF (Ferris, 2004; Guénette, 2007). In Lalande (1982) and Semke (1984) the group receiving indirect WCF had more

advantages in comparison to the other treatments. In Lalande's study, participants in the indirect WCF treatment were encouraged to check grammar books, to consult with their peers and to keep a record of their recurring errors, while those in the direct group were not allowed such opportunities (Lalande, 1982).

In the same vein, only the group that received indirect WCF in Semke (1984) was required to rewrite their compositions, while participants in the other groups were asked to submit new essays. The consequences of this were that the indirect WCF group did not write as "much new material as the other groups" (Semke, 1984, p. 197) and that the other three groups (i.e. direct WCF, feedback on content, and feedback on content coupled with direct error correction) did not get a chance to revise their writing. Furthermore, the fact that the indirect WCF group was eventually provided with direct error correction on their remaining errors in the revised versions makes the comparability between the four treatment groups less forward.

The internal validity in Semke (1984) is further compromised by the presence of an incentive for some groups only (Guénette, 2007). In fact, the groups receiving feedback on form were marked on a ratio of errors to the number of written words, while the content group was graded based on the total number of words written. As explained by Guénette, it is possible that the latter may not have been "worried about losing points, while the other three groups probably needed to write less for fear of making too many mistakes" (p. 50). This is another methodological limitation that accentuates the lack of comparability between the treatment conditions.

Taken together, the different limitations in either study show that the type of feedback provided is not the only criterion distinguishing the various groups. As such, it remains unclear whether any observable advantages of a treatment group can be solely attributable to the WCF

type. This in turn raises questions about the comparability of the results obtained and precludes any generalizations about the effectiveness of particular WCF treatments. Issues of interpretability are also compounded by differences –both qualitative and quantitative– in the writing tasks completed by students. This methodological inconsistency is discussed next.

Different writing tasks. In addition to differences within and across studies on when effectiveness is assessed and how feedback is operationalized, the early studies vary in terms of the writing tasks students completed during the treatment and/or the post-tests. These tasks differ not only in their length, but also in their cognitive and linguistic complexity. For instance, while participants in Lalande (1982) and Sheppard (1992) respectively wrote plot summaries and personal essays in class, those in Semke (1984) and Kepner (1991) submitted journal entries that were written as homework assignments.[7] In both Kepner (1991) and Fazio (2001), the length of the journal entries varied from one paragraph to several pages, and it remains unclear whether or not the interactions between text length and error occurrences were accounted for. Furthermore, and as explained by Ferris (2003), learners are less likely to pay attention to accuracy issues in journal writing than in more academic genres such as plot summaries and argumentative essays- an aspect that is also acknowledged by Fazio (2001). The latter concedes that paying attention to accuracy may have been "overshadowed by the larger communicative value" in journal writing (Fazio, 2001; p. 246). In addition to the obvious differences in topical and situational constraints, it is not clear whether the time spent on task is comparable across studies or not (Storch, 2010; Van Beuningen, 2010).

[7] In Fazio (2001), the journal entries were either completed in class or assigned as homework.

Absence of a control group. Finally, it is important to note that of all the design limitations that affect early inquiries, it is the absence of a control group that constrains their internal validity the most. Having no control group (i.e., a no feedback group) casts doubt on the reported results in these studies. The vast bulk of early WCF studies, particularly those examining its impacts on accuracy in new texts, did not include a real control group, i.e., a control group which completed the same writing tasks without receiving any type of corrective feedback (e.g., Fazio, 2001; Kepner, 1991; Lalande, 1982; Robb *et al.*, 1986; Semke, 1984; Sheppard, 1992. The few exceptions include Ashwell (2000), Fathman and Whalley (1990) and Ferris and Roberts (2001). The absence of a control group has been justified on ethical grounds: given that those studies were conducted in intact classes, it was often felt unfair to withhold feedback from some students (Ferris, 2004).

While it is possible to argue that a group which received comments on content can count as a control/comparison group (e.g., Kepner, 1991; Semke, 1984; Sheppard, 1992), it remains ambiguous whether the provided comments on content were a reaction to parts which are problematic because of their linguistic (i.e., syntactic or lexical) issues or not. Consequently, it is debatable whether such a group is a real control (Ferris, 2003, 2004). Similar reservations are echoed by Bitchener and Storch (2016) who maintain that "findings are only valid if the constructs [referring to the control group] are fully and unambiguously defined" (p. 39). The need for a control group to which no CF is provided is equally underscored by Ferris (2004) and Truscott (1999, 2004).

Whether they are in favor of or against WCF, researchers concur that the research base up until 2004 is remarkably "inadequate" and "insufficient" to use Ferris's (2004) words (p. 50; see also Ferris, 2003; Guénette, 2007; Truscott, 1999, 2004). To account for such inadequacy, Ferris

(2003, 2004) and Guénette (2007) point to other inconsistencies which make the findings of early studies incomparable. One of those dissimilarities consists of whether inter-rater (or intra-rater) reliability is computed or not (Ferris, 2004). Another one refers to who provided the WCF: for example, in Ferris and Roberts (2001) and in Kepner (1991), the researchers annotated learners' texts, while in Lalande (1984) and Chandler (2003), feedback was provided by the participating instructors. Which methodological choice is better remains an indecisive issue in the WCF literature. On the one hand, having teachers provide feedback to their students' essays is more ecologically valid; since both partners are already familiar which each other, the Hawthorne effect is constrained. On the other hand, this choice is not without its complications, for teachers may not always follow the researchers' recommendations. In Ferris (2006), instructors used both direct and indirect WCF although they initially agreed on providing indirect WCF only, leading the researchers to modify their original questions. For this reason, Ferris (2006) called for "having one teacher or researcher provide error feedback for all student subjects to ensure greater consistency in treatment and thus enable assessment of the effects of feedback without this potentially confounding variable" (p. 93).

In light of the aggregate inconsistencies within and across studies, Ferris's (2004) observation that the early studies "compare apples and oranges (and pears, and grapes, and nectarines . . .)" (p. 52) seems to be justified. As such, the reported results, which are unsurprisingly conflicting, cannot be taken as evidence for the effectiveness of WCF or as a proof against its utility. Subsequently, both proponents and advocates of WCF have called for more rigorously designed research, hoping that some of the limitations of early studies would be addressed and more conclusive findings would be reached. These more recent examinations are discussed next.

2.3.3.2 More recent studies on WCF.

Seeking to overcome the aforementioned methodological limitations, different ISLA researchers set out to examine the effectiveness of WCF through more tightly-designed, quasi-experimental studies (Table 2)[8]. With the exception of Truscott and Hsu (2008) who examined the effectiveness of comprehensive indirect WCF only, all recent studies have compared different types of WCF (i.e., at least two other experimental groups are included in addition to the control group). More importantly, unlike Truscott Hs and Hsu who report no accuracy gains in new texts for their WCF group, most- not to say all- recent studies demonstrate that their WCF groups outperform the control group in immediate and/or delayed post-tests (Bitchener & Knoch, 2008; Sheen, 2007; Sheen *et al.*, 2009).

Besides examining the overall effectiveness of WCF, these recent inquiries have also sought to investigate more specific questions such as the relative merits of different operationalizations of WCF (e.g., Bitchener & Knoch, 2008) and to a less extent the possible interactions between WCF type and error category (e.g., Van Beuningen *et al.*, 2012). In general, these studies can be categorized into three groups: (1) few studies compared input-providing (i.e., direct) to output-prompting (i.e., indirect WCF) (Van Beuningen, *et al.*, 2008, 2012); (2) a couple examined focused versus comprehensive (Ellis *et al.*, 2008; Sheen *et al.*, 2009.), (3) while a larger number investigated the differential impacts of focused WCF with or without metalinguistic explanations (e.g., Bitchener & Knoch 2008, 2009, 2010a; Sheen, 2007). In the

[8] Please see page 79.

following section, a brief summary for each group of studies will be provided, which will then be followed by a critical review of their limitations.

. *Direct versus indirect WCF.* Two quasi-experimental studies by Van Beuningen *et al.* (2008, 2012) investigate the effectiveness of WCF on both revised and new texts and examine the differential impacts of direct WCF (i.e., providing the target-like form) and indirect coded WCF. In their more recent study which included a larger group and a delayed post-test, 268 low intermediate, Dutch as a SL, high school learners were divided into four groups: (a) direct WCF, (b) indirect coded WCF, (c) revision with no feedback, and (d) extra-practice. For the pre-test, students completed a receptive vocabulary test- meant to measure their proficiency level- and wrote an expository paragraph about the metamorphosis of an insect. To ensure that learners were familiar with the content of the writing tasks, the topics were "introduced and explained by the researcher" (p. 13). One week later, the first three groups were given back their texts and asked to revise them, while the fourth group was asked to write a new paragraph about the metamorphosis of another insect. The two feedback groups received WCF on grammatical errors, which included "article errors, inflectional errors, word order errors, omissions of necessary elements, additions of unnecessary elements, pronominal errors, and other grammatical errors" (p. 17) and on non-grammatical errors, which comprised lexical, orthographical and pragmatic errors.

A comparison of the revised drafts reveals that unlike the self-correction group, the two WCF groups have significantly improved the overall accuracy of their texts, in that learners in both treatments significantly reduced their grammatical and non-grammatical errors. Furthermore, the direct WCF shows "larger effects than the indirect treatment — when compared to self-correction" (p. 26). As explained by Van Beuningen *et al.*, this was not surprising because

"pupils receiving direct CF only needed to copy the target forms as provided by the researcher" (p. 26).

Results from the immediate and delayed post-tests, administered one and four weeks later, show that both WCF groups outperformed the self-correction and the extra practice group, which proves that WCF results in long-term sustained accuracy gains. More importantly, the initial superiority of direct error correction over indirect WCF was not maintained in the post-tests. It was found that the differential effects of the two WCF techniques were mediated by error type. In the immediate post-test, both WCF types were effective in improving learners' lexical and orthographical accuracy, while only direct WCF helped them significantly improve their grammatical accuracy, when compared to the practice group. In the delayed post-test, the positive effects of direct WCF on grammatical accuracy were still maintained. In contrast, indirect WCF was "significantly more effective in reducing the number of [lexical and orthographical] errors in pupils' writing than either self-correction or writing practice" (p. 30). In other words, the post-tests show that while direct WCF was more effective on grammatical errors, indirect WCF was more successful for lexical and orthographical issues.

These findings do not completely corroborate those reported in Ferris's (2006) descriptive study, in which she examined the effects of teachers' WCF on both revised and new texts. 92 university ESL students took part in this study and completed four three-draft essays over the course of a 15-week semester. They all received WCF on the second drafts of their first, second and third essays, as such there was no control group. Although the initial purpose of the study was to examine the effectiveness of indirect coded WCF, data analysis showed that the teachers did not always adhere to the agreed-upon coding schema. They rather provided direct WCF for what Ferris (2006) classifies as non-treatable errors, which include sentence errors and lexical

errors. Conversely, they used indirect coded WCF for what the researcher categorizes as "treatable" errors. These include errors in verbs, nouns and determiners.

An analysis of learners' revised drafts shows that they were successful at incorporating both direct and indirect WCF, and that their accurate changes reached 88% for the direct WCF and 75% for the indirect WCF. Comparing the first and fourth essays, Ferris (2006) indicates that students showed statistically significant gains in verb errors only, which were marked by indirect WCF. No improvement was made for the other two error categories, i.e., errors in nouns and determiners, although they were also marked by indirect WCF. No gains were either reported in the untreatable errors, which were overwhelmingly marked by direct WCF. Ferris (2006) argues that these findings show "a longitudinal difference in student achievement, based on whether errors were "treatable" or "untreatable," and this distinction may possibly be attributed to widely disparate teacher feedback strategies" (p. 98). Such a conclusion seems unwarranted for the difference was statistically significant for only one out of the five error categories examined.

Furthermore, and due to differences in research methodologies, van Beunigen *et al.*, (2008, 2012) and Ferris (2006) are not straightforwardly comparable. It is plausible to argue that Ferris (2006) did not include a control group, which makes their results less valid than those reported in Van Beuningen *et al.* (2008, 2012). Nonetheless, they also differ from the latter in that they adopted a longitudinal design as they provided WCF more than once, which makes theirs more ecologically valid. "Whether concerns over ecological validity should not supersede those over internal validity," as proposed by Lyster and Ranta (2013, p. 180) remains a contentious issue in oral and written corrective feedback.

Notwithstanding, the scarcity of studies comparing the two WCF types and their different methodological designs preclude any meaningful comparisons, which warrants further research

on the relative merits of both WCF types. Besides underscoring the need for more studies that compare both WCF types, those findings- however inconclusive they are- underscore the necessity to investigate how error types mediate the effectiveness of WCF types, which remains an issue that is relatively unexplored.

Focused versus Comprehensive Direct WCF. In addition to comparing direct to indirect WCF, some studies have also examined the relative merits of providing focused (targeting one linguistic error) versus comprehensive WCF (targeting many forms). These examinations are guided by divergent theoretical claims about the superiority of either type. On the one hand, some researchers have argued that comprehensive WCF is more pedagogically sound in that targeting a limited number of errors would not help learners improve the overall accuracy of their texts (Bruton, 2010; Ferris, 2010; Van Beuningen, 2010). On the other hand, some have posited that focused WCF is better suited at raising learners' awareness and triggering the cognitive processes of noticing and understanding (Bitchener, 2008; Ellis *et al.*, 2008). However, to date, only two studies- Ellis *et al.* (2008) and Sheen *et al.* (2009) - have attempted to empirically test these contradictory theoretical arguments, yet without providing conclusive findings.

In both Ellis *et al.* (2008) and Sheen *et al.* (2009), participants were provided with direct WCF on their determiner errors (in the focused treatment) and on a variety of errors (for the comprehensive treatment). Intermediate English as a second or foreign language learners participated in both studies and were divided into three groups: control, comprehensive direct and focused direct. The two experimental groups received WCF on three tasks in Ellis *et al.* and on two activities in Sheen *et al.* In Ellis *et al.* (2008), the two WCF gained from pre-test to post-tests and outperformed the control group on the delayed post-test. However, there was no difference between focused and unfocused WCF in increasing learners' accuracy. Partially corroborating

results are found in Sheen *et al.* (2009), in that only the focused WCF outperformed the control group in both immediate and delayed post-tests and that there were no significant differences between focused and comprehensive WCF. In light of these findings, and contrary to Ellis *et al.* (2008), Sheen *et al.* (2009) suggested that providing unfocused WCF is not beneficial. They concluded by stating that comprehensive WCF can be "confusing, inconsistent and unsystematic" and might overburden learners (p. 567).

It should be noted, nevertheless, that interpretability of these findings is constrained by some methodological limitations in both inquiries. As argued by Xu (2009), the small number of participants in Ellis *et al.* (2008), which was actually reduced from 49 (18 for focused, 18 for comprehensive and 13 for the control group) to 35 (11 for focused, 13 for comprehensive, and 11 for control) by the end of the experiment, "may not lend much validity to the mixed design (3 groups x 3 writing tests) ANOVAs" used in the study (p. 271). Furthermore, and as acknowledged by the authors, many learners in the unfocused group made fewer errors in their determiner use than those in the focused group, as such they received less corrective feedback. Differences in the amount and quality of WCF between the two treatment groups are also apparent in Sheen *et al.* (2009). In the latter, participants in the focused group had all their determiner errors corrected, those in the comprehensive group received unsystematic WCF in that some of their determiner errors were flagged, while others were not, which might have confused learners and prevented them from accurately understanding when their use was accurate and when it was not (Van Beuningen *et al.*, 2012). In light of the conflicting findings yielded by these two partially flawed studies, a conclusive answer about the superiority of focused or comprehensive WCF is yet out of reach.

Studies on WCF with and without metalinguistic explanations. Most of those studies have investigated different operationalizations of direct WCF (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2010a; Ellis *et al.*, 2008; Sheen, 2007; Sheen *et al.*, 2009; Stefanou & Révész, 2015). In contrast, only two studies have examined indirect WCF with and without metalinguistic clues (Bitchener & Knoch, 2010b; Bitchener *et al.*, 2005). The majority of those inquiries have also studied the effects of WCF for increasing accuracy on specific uses of English determiners, i.e., "a" for first mention and "the" for anaphoric mention (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2010a; 2010b; Ellis *et al.*, 2008; Sheen, 2007; Sheen *et al.*, 2009; Stefanou & Révész, 2015). While they all indicate that WCF groups outperform the control group, they have yielded conflicting findings regarding the relative merits of combining WCF with metalinguistic explanations. Put differently, it is still unclear whether coupling WCF with metalinguistic information would bring about more gains or not.

Most of the studies that have examined direct WCF (i.e., provision of the correct form) with or without metalinguistic explanations were unable to find any statistically significant differences between the various operationalizations of direct WCF. Only Sheen (2007) has lent support in favor of direct WCF with metalinguistic explanation over direct WCF only. In Sheen (2007), 91 intermediate ESL learners were divided into three groups: direct WCF only, direct WCF with metalinguistic comments and a control group. All three groups completed a pre-test, an immediate and a delayed post-test, but only the two WCF groups received two treatment sessions. In contrast, the control group took the tests only, which means that they did not participate in the writing sessions. During a treatment session, the WCF groups completed a dictogloss task, whereby the students read a short story and then rewrote it. In the following class, they received their annotated texts and were asked to look over the feedback, which only

targeted their erroneous use of “a” for first mention and “the” for anaphoric mention. For the pre and post-tests, all students completed three tasks: a speeded dictation task, an error correction task and a writing task (picture-cued narrative). Results of the immediate post-tests show that both WCF groups outperformed the control group in the error correction task, but that there were no statistically significant differences in the speeded dictation or the writing task. In the delayed post-tests, the two WCF groups outscored the control group in the error correction, but only the WCF plus metalinguistic explanations outperformed the control group in the dictation task, and both the control and the direct WCF only groups in the writing tasks. Aggregating the scores of the three tasks, Sheen (2007) indicates that the two WCF groups performed better that than the control group in post-test 1 and that the group receiving metalinguistic feedback outscored the direct WCF group in the delayed post-test. Sheen (2007) suggests that the delayed effect of direct WCF with metalinguistic explanation can be accounted for by the passage of time. Overall, these findings suggest that direct WCF in tandem with metalinguistic explanation is more effective than direct WCF only.

Such an advantage was not confirmed in other examinations of the relative merits of similar WCF types. In a series of studies, Bitchener and Knoch (2008, 2009a, 2009b, 2010a) compare the relative merits of (1) direct WCF only; (2) direct WCF and written metalinguistic explanation (ME), and (3) direct WCF combined with written and oral ME on low intermediate learners’ accurate use of two functional uses of the English determiners (i.e., “a” for first mention and “the” for subsequent mention). In their 2010a study, for example, 52 low-intermediate ESL university students were assigned to four groups (direct WCF with written and oral ME; direct WCF with written ME; direct WCF only, and a no-feedback group) and completed five tests (a pre-test, an immediate post-test, and three delayed post-tests administered 2, 6 and 10 months

after the experiment). The writing tasks consisted of picture descriptions of different social events (e.g., a family gathering, a sporting event, etc). One week after the pre-test, participants in the treatment groups received their pre-test writing pieces annotated according to their WCF condition. They were then accorded 5 minutes to look over the provided WCF but were not asked to rewrite their paragraphs. Results of all four post-tests indicate that while all treatment conditions led to sustained accuracy gains which continued over a 10-month period, no statistically significant difference was observed between the three options of direct WCF. Put differently, "there was no advantage for any one of the direct feedback options" (Bitchener & Knoch, 2010b, p. 208).

Similar findings are reported in Stefanou and Révész (2015) who have also investigated the effectiveness of direct WCF with or without metalinguistic explanation on the use of English determiners, and who have found that both techniques are equally effective in improving learners' accurate use of the determiner "the" for specific plural references. Taken together, these studies show that at least in the case of determiner use, direct WCF only is "just as effective as the additional provision of written and oral meta-linguistic explanation" (Bitchener & Knoch, 2009a, p. 327). The conflicting findings on the differential impacts of different combinations of direct WCF in the studies cited so far can be attributable to different methodological limitations, which will be explained shortly.

Regarding studies on focused indirect WCF (i.e., WCF that withholds corrections and prompts learners to find the target-like form on their own), a different pattern seems to emerge. Although only two studies (to the best of my knowledge) have examined the effectiveness of three variations of focused indirect WCF, their results are still worth noting.

In Bitchener *et al.* (2005), two variations of indirect WCF are compared: (1) uncoded WCF with written and oral metalinguistic explanations and (2) uncoded WCF with written metalinguistic explanation only. Both conditions were focused on three types of errors (prepositions, the past simple tense, and the definite determiner)[9]. 53 post-intermediate college ESL students were divided into three groups: (1) written metalinguistic explanation and a 5 minute student-researcher individual meetings; (2) written metalinguistic explanations only and (3) no corrective feedback. Participants completed four informal letters (average length = 250 words) and received WCF on the first three tasks. A comparison of accuracy measures of the specific features in task 1 (completed in week 2) and task 4 (completed in week 12) indicates that WCF resulted into statistically significant accuracy gains in the simple past and the definite determiner but not in the use of prepositions, and that written metalinguistic feedback that is coupled with oral feedback is more effective than written metalinguistic feedback only.

Similar findings in favor of coupling indirect WCF with metalinguistic explanations are also underscored in Bitchener and Knoch (2010b). In this study, 63 advanced ESL learners were assigned to one of the following four conditions: (1) indirect uncoded WCF, which took the form of circling only; (2) indirect uncoded WCF plus written ME; (3) indirect uncoded WCF plus written and oral ME and (4) no feedback. WCF was provided only once and targeted learners' erroneous uses of two functional uses of English determiners ("a" for first mention and "the" for anaphoric mention). Three days after the pre-test, which consisted of a picture description of a

[9] It is true that Bitchener *et al*., (2005) classify the WCF they evaluate as "direct explicit written feedback." Yet, a closer look at the example they provide on page 205 reveals that they withheld the correct form for verb tense errors but provided it for prepositions and definite determiners. As such, in this section, only the results pertaining to verb errors will be reported for this error type was targeted by indirect WCF.

social event, the three WCF groups were given back their annotated texts and were allowed "several minutes to consider the feedback" (p. 213); all groups then took the immediate post-test. Results of the immediate post-test show that all treatment groups, regardless of the type of the WCF they received, outperformed the control group. However, only the two groups that received metalinguistic explanations sustained their progress in the ten-week delayed post-test. In other words, those who had their errors circled only and did not receive metalinguistic explanations did not differ significantly from the control group, which means that they could not retain the improved accuracy they showed in the immediate post-test. This long-term effect of WCF that is combined with metalinguistic explanation is similar to the one indicated in Sheen (2007).

A comparison of the two groups of studies (i.e., direct WCF and indirect WCF) suggests, albeit tentatively, that the positive effects of supplying metalinguistic explanation are more pronounced for indirect WCF. Taking into consideration that in most of these studies, WCF targeted two relatively simple uses of English determiners, it is possible that learners receiving direct error correction (without any metalinguistic clues) were able to induce the patterns for accurate use from the corrections, as such no difference was observed between their performance and that of those who received direct WCF plus metalinguistic information (except in Sheen, 2007). In the case of indirect WCF, however, one can possibly argue that withholding the correct form has pushed learners to think more about the source of their erroneous uses and to engage in hypothesis testing. Such cognitive processes might have become more effective when metalinguistic information was supplied.

Tempting as such a hypothesis might be, it should be tempered with a couple of considerations. First, there is an urgent need for more studies that target other linguistic features because drawing conclusions based on inquiries on English determiners only is quite premature.

Second, and in light of the scarcity of studies on indirect WCF, more systematic comparisons between different types of indirect WCF (with or without metalinguistic explanations) are needed to better understand how the impacts of WCF are related to their varying degrees of explicitness. Third, the studies reviewed so far have their own share of methodological limitations, which may account for their conflicting results and which makes reaching any definite conclusion particularly challenging. Some of these methodological limitations are detailed next.

Table 2

Representative Studies of more Recent Research on WCF

Study	WCF types	Findings
Van Beuningen, *et al.*, 2012	1.comprehensive direct 2. comprehensive indirect (coded) 3. Control 1: writing practice 4. Control 2: self-correction	WCF is effective for revised and new texts (in both post-tests) Direct > indirect on grammatical errors Indirect> direct on non-grammatical errors.
Ellis, *et al.*, 2008	1. focused direct only 2. comprehensive direct 3. control	WCF groups > control in immediate and delayed post tests No difference between focused and comprehensive WCF
Sheen, 2007	1. focused direct feedback only 2. focused direct feedback+ written ME 3. control	In the immediate post-test WCF groups > control (no difference between WCF) In the 2-month delayed post-test WCF groups> control Direct+ ME> direct only
Bitchener & Knoch 2008	1. focused direct+ written and oral (ME) 2. focused direct+ written ME 3. focused direct only 4. control	WCF groups > control in immediate and delayed post tests No difference between WCF types in both tests
Bitchner & Knoch, 2010b	1. focused indirect: written ME 2. focused indirect: written+ oral ME 3. focused indirect: circling only 4. control	In the immediate post-test WCF group> control No difference between WCF types In the delayed post-test Only the metalinguistic groups > control

Note. ME: metalinguistic explanations; > : statistically significantly better.

2.3.3.3 Major methodological limitations in recent WCF studies.

Although the more recent inquiries have concurred that WCF is effective on improving learners' accuracy levels, and that their gains are enduring, they have yielded mixed findings about the relative efficacy of different WCF types. The growing research on WCF cannot yet offer definitive answers to more specific questions such as which type of WCF is more effective or which variables and/or interactions between those variables might mediate its effectiveness. It is still unclear whether indirect WCF is more effective than direct WCF or whether WCF that is coupled with metalinguistic explanation is more advantageous than WCF only. Overall findings from these studies do not easily lend themselves to a clear, relatively incontestable interpretation, which can be largely attributed to two design issues common in most recent studies. These are discussed next.

One shot WCF treatment. Most recent investigations employ a one shot treatment. In other words, learners are provided with a single WCF episode on a written draft, after which they are administered the post-tests (few exceptions include Bitchener *et al.*, 2005; Sheen, 2007; Stefanou & Révész, 2015). This is a lingering concern from both practical and acquisitional standpoints. First, such short treatments are incongruent with real-classroom practices given that L2 teachers usually provide corrective feedback on different writing assignments (e.g., Guénette & Lyster, 2013). Second, from an acquisitional perspective, unless learners are provided with repetitive opportunities for knowledge retrieval in meaningful practice as advocated by skill learning theories, their chances for deeper processing, i.e., processing that would lead to less controlled and more automatized knowledge, are restricted (e.g., DeKeyser, 2007). It follows that one shot WCF treatments are less likely to engage learners with the provided WCF, which casts doubt on its meaningfulness and usefulness for learning (Storch, 2010).

*Lack of revision or inadequate revision***.** Limited opportunities for engaging learners with the given WCF are also due to the lack or absence of revision following WCF. This is not necessarily the case in earlier longitudinal studies (Lalande, 1982; Semke, 1984). Yet, in many recent, quasi- experimental studies, participating learners are not required to incorporate the provided WCF. In best case scenarios, they are given short periods of time to look at the provided WCF (e.g., Bitchener, 2008; Bitchener & Knoch, 2009; Sheen, 2007). This is problematic because the potential facilitative role of WCF can be constrained if learners are not encouraged to use the feedback they were given (Ferris, 2004; Guénette, 2007). It is true that studies in which WCF was not followed by revision have shown that it is still effective in improving learners' accuracy. However, studies that have systematically examined the differential impacts of WCF with and without revision have shown that requiring students to revise after WCF can maximize its effects (Chandler, 2003; Shintani, *et al.* 2014). In Chandler (2003), the groups that revised after receiving WCF improved in accuracy from the first to the fifth assignment while no such gains were observed for the group that did not revise. Shintani *et al.* (2014) also compared the accuracy scores of the groups that rewrote their texts after WCF and those that only looked at their WCF. While no difference between the two groups was initially found in the immediate post-test, only the WCF plus revision group was able to outscore the control group in the two-week delayed post-test, which according to Shintani *et al.* "indicates a long-term advantage of requiring learners to rewrite following feedback" (p. 125).

Requiring learners to revise their texts after receiving WCF can enhance its noticeability and would engage them in the deeper processing that is needed for restructuring and fine-tuning their grammar to occur (Ferris, 2004; Shintani *et al.*, 2014). One can possibly argue that this is particularly the case when learners are provided with indirect WCF since they are not given the

correct form but are rather prompted to think about their error and how to fix it. They are thus engaged in problem-solving as they attempt to identify the source of their erroneous use and try out solutions on their own (Ferris, 1999, 2002; Lalande, 1982).

Finally, and in addition to the two aforementioned limitations (i.e., frequency and duration of WCF and lack of revision opportunities), recent WCF studies have not succeeded at addressing other methodological issues that were equally present in early inquiries. The common limitations between the two groups of studies relate to (1) target features, (2) measurement tools and (3) learner individual differences. These lingering issues are discussed next.

2.3.3.4 Common design issues between early and recent WCF studies

Target features. Equally problematic concerns relate to the selection of the targeted errors. In most WCF studies, early or recent, descriptive or rigidly experimental, the pendulum has swung between the two extremes of comprehensive error treatment, in which all erroneous instances of language use are flagged, on the one hand, and the highly focused WCF, which targets one or two specific features, on the other. Arguments against untargeted WCF point out to how time-consuming and discouraging it can be for both teachers and learners (e.g., Sheen, 2007). Those supporting focused WCF claim that targeting specific features enhances noticing and attention, and thus best promotes language development (Ellis *et al.*, 2008). While it is more plausible that learners would attend more to WCF that targets a specific feature than that which addresses all errors, highly focused WCF is problematic for two major reasons. First, it lacks ecological validity for it does not correspond to teachers' actual WCF practices. As reported in studies by Lee (2004, 2008) and Furneaux *et al.*, (2007), L2 instructors often provide WCF comprehensively. Second, focused WCF might enhance the saliency of the studied features, which raises questions about the causes of any possible gains. In other words, would increases in

accuracy result from the WCF itself or rather from strengthened learners' sensitivity to the targeted feature? In addition, studies that employed a highly targeted WCF have examined a very limited range of language features- mostly the determiner system. Thus, it remains unclear whether the reported advantages can be generalized to other error types (Storch, 2010). Furthermore, neither the highly focused WCF nor the comprehensive WCF seems to correspond to teachers' actual practices.

Measurement issues. The controversies surrounding the use of highly-focused or comprehensive WCF are further compounded by inconsistencies in measuring accuracy gains. In the comprehensive WCF studies, for example, inaccuracy ratios are often computed "as the total number of errors divided by the total number of words written" as it is the case on Truscott and Hsu (2008, p. 297) and Van Beuningen *et al.* (2012). These ratios, however, would not necessarily capture change that might (or might not) have resulted from the provided WCF (Bruton, 2009). Take for example a learner whose pre-test contains three errors of inflectional noun endings but whose post-test include only one inflectional noun ending error and two punctuation errors. In both tests, she would end up with similar error ratios, which would misleadingly suggest that she had not made any progress.

Another issue that is also prevalent in comprehensive WCF studies refers to the variety of accuracy measures implemented. Some studies computed ratios of errors to total number of words (e.g., Ferris & Roberts, 2001; Van Beuningen *et al.*, 2012), others examined ratios of error free T-units[10] to the total number of T-units (e.g., Robb *et al.*, 1986) while few assessed mean

[10] T-units refer to one main clause and any subordinate clauses. For example, "pour faire cette visite, il faut beaucoup d'argent" is one T-unit that consists of two clauses (Gunnarson, 2012, p. 252).

scores for accuracy (e.g., Kepner, 1991). These different accuracy measures make comparisons across studies particularly difficult.

It is true that this less of an issue in the highly-focused WCF studies (i.e., studies that examined accuracy of one linguistic feature). However, while some of those inquiries compared percentages of correct usage in obligatory contexts (e.g., Bitchener, 2008; Bitchener & Knoch, 2009), their adopted measures do not always account for the number of overused instances, which also casts doubt on how precise these measures are (see, however, Sheen, 2007; Shintani *et al.*, 2014).

Learner individual differences. In addition to the aforementioned limitations, empirical research on WCF has to date overlooked the moderating effects of learners' individual differences. This is surprising given that both SLA and L2 writing researchers concur that "individual differences may be hypothesized to exert influence on how students process feedback, the extent to which they notice gaps in their knowledge" and the extent to which they benefit from corrective feedback (Kosmos, 2012; p. 400). Despite different calls to examine the possible interactions between learner internal variables and WCF types (R. Ellis, 2010; Storch, 2010), empirical WCF research is still lagging behind. Two learner internal factors, proficiency level and language analytic aptitude, merit further attention.

Proficiency. In most WCF studies, proficiency is assumed based on learners' educational level instead of reliable measurement tools (e.g., Bitchener, 2008; Ferris & Roberts, 2001; Sheen, 2007). Guénette (2007) explains that such an assessment is not fined-grained enough because students in the same class "can vary widely in their command of English grammar, in their familiarity with the structures … and in their background in formal instruction" (p. 42).

It should be noted here that defining and operationalizing L2 proficiency are highly contentious issues in the field of SLA. As affirmed by Housen, Kuiken & Vedder (2012) "L2 proficiency is not a unitary construct" (p. 1). In fact, L2 proficiency can be described in three distinct components, i.e., accuracy, fluency and syntactic maturity, the operationalization and measurement of each have been marked by inconsistencies (Housen, *et al.*, 2012). As mentioned before, accuracy has been computed differently across studies. It was measured either in terms of overall error frequency (e.g., Van Beuningen, *et al.*, 2012), of error frequency of pre-determined features (e.g., Stefanou & Révész, 2015), or the two combined. Fluency, which refers to the speed and ease by which a learner produces oral and written output, is quantified in writing research as the number of words produced during a timed assignment (e.g., Chandler, 2003). Syntactic maturity, the third dimension of L2 proficiency, is operationalized in terms of subordination and can be measured by number of clauses per T-unit (e.g., Van Beuningen *et al.*, 2012). In a nutshell, because proficiency is a componential construct, it cannot simply be inferred from learners' institutional level. Furthermore, using institutional placement as an index of proficiency is problematic because what is intermediate in one setting might be placed as low or high intermediate in another setting. This in turn would make comparability between studies more challenging and would limit generalizability of the findings.

Similar concerns about "the lack of attention to L2 proficiency as a moderating variable in the field of SLA" are raised by Sasayama (2015, p. 76). It is argued that operationalizing proficiency based on institutional level or on impressionistic judgment is both inaccurate and impractical (Sasayama, 2015; Tremblay, 2011; Tremblay & Garrison, 2010). Conceding that standardized tests, which provide more accurate estimates for proficiency, are often quite expensive and "time-consuming to be administered in experimental settings", Sasayama explains

that different "short-cut estimates" have been developed (p. 77). Also described as reduced redundancy tests, these measures include cloze tests and elicited imitation tests. A cloze test often takes the form of fill-in the blank exercise (e.g., Tremblay, 2011), while elicited imitation tests often require learners to listen to sentences and repeat them as accurately as possible (e.g., Sasayama, 2015). According to Sasayama, both tests exhibit high reliability and provide practical and efficient measures of proficiency.

It is true that few WCF studies have begun to use other short-cut estimates such as receptive vocabulary tests (e.g., Van Beuningen, *et al.*, 2012) and written multiple-choice grammar tests (e.g., Stefanou & Révész, 2015), which is a positive step forward. However, because learners in these studies were found to be homogenous in terms of their proficiency, these studies did not provide insights into how the differential impacts of WCF may vary across different proficiency levels. Investigating how different WCF types would interact with various proficiency levels is an issue worth examining, especially in light of findings from oral CF literature suggesting that proficiency is a mediating factor (Ammar & Spada, 2006).

Language analytic abilities. Proficiency is not the only learner individual difference for which future empirical research needs to account. Individual cognitive factors such as learners' aptitude and its interactions with various types of WCF are also ignored in most studies on WCF (Kang & Han, 2015).

Aptitude for language learning refers to the special ability or potential that learners have for learning foreign languages (Carroll, 1962, 1981). Caroll (1962) distinguished between four subcomponents: (1) phonetic coding ability which he defined as "the ability to perceive phonetic distinctions" (p. 96), (2) rote memory or "the capacity to learn a large number of these associations in a relatively short time" (p. 129), (3) grammatical sensitivity, i.e. "sensitivity to

the functions of words in a variety of contexts" (p. 129) and (4) inductive language learning ability which refers to "the ability to infer linguistic forms, rules, and patterns from new linguistic content" (p. 130). The last two subcomponents, i.e., grammatical sensitivity and inductive language learning, were collapsed by Skehan (1998) into one category, which he named language analytic ability.

Out of the numerous studies on WCF, only two -- both of which evaluate the effectiveness of direct WCF on the use of English determiners -- have studied the moderating impacts of two aspects of learners' language aptitude. Sheen (2007) examines the effectiveness of direct WCF on learners' accuracy in relation to their language analytic ability, while Stefanou and Révész investigates the moderating role of grammatical sensitivity and metalinguistic knowledge. Both inquiries underscore a significantly positive association between learners' accuracy gains and the aptitude subcomponent in question. Particularly in Sheen (2007), the positive correlation was stronger in the group that received direct WCF combined with metalinguistic explanations than in the WCF only group. To what extent such patterns of interaction between WCF type and learners' language analytic ability would be observed in studies on other types of WCF (e.g., indirect WCF) that target other linguistic features (other than determiners) warrants further research.

In a nutshell, although WCF studies have yielded mounting evidence for the effectiveness of WCF on sustained accuracy, they have resulted into conflicting findings with regards to the differential impacts of various types of WCF. More precisely, the lingering concerns in WCF literature consist of (1) the type of WCF (i.e. direct vs indirect; with or without metalinguistic explanations); (2) the mediating effect of error type; (3) the moderating effects of proficiency and

language analytic aptitude. Other variables that transpire from the extant literature concern (4) the frequency of WCF, (5) the resultant revision and (6) the writing task. (see Figure 3)

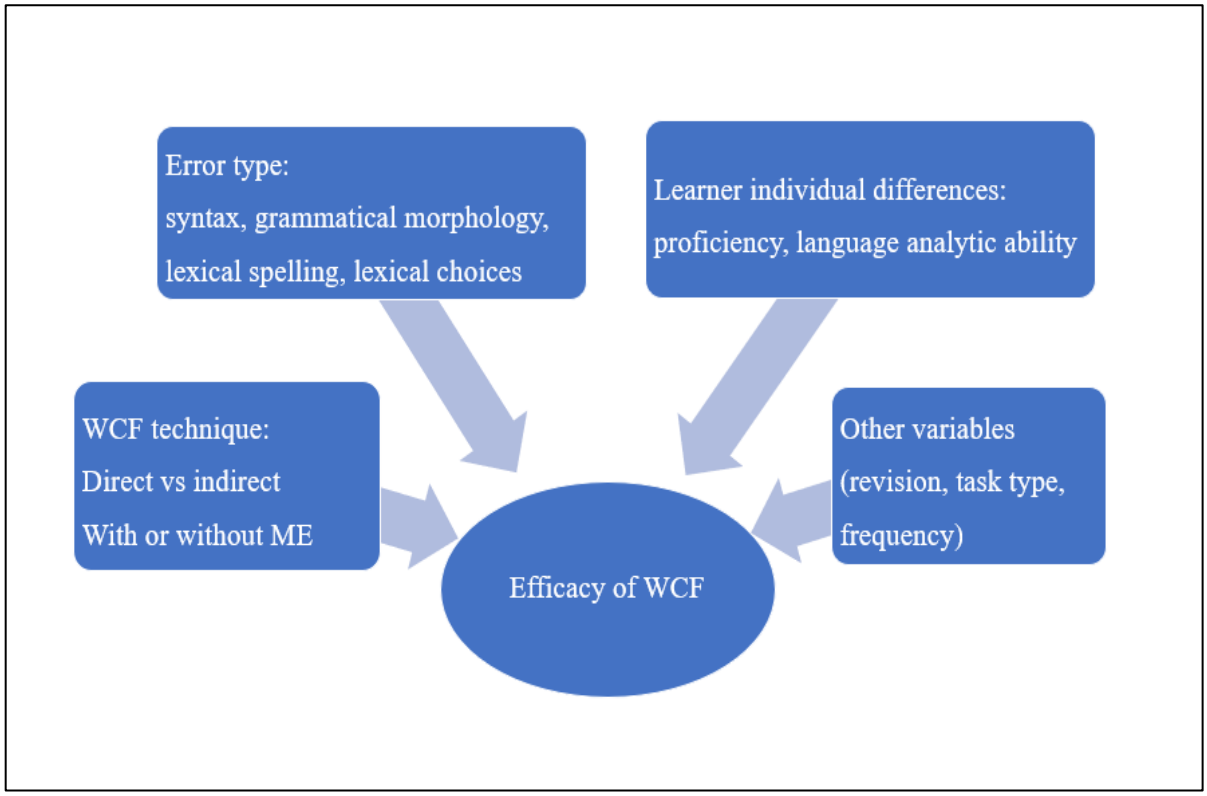


Figure 3. Variables impacting the efficacy of WCF

2.4 Specific research questions

In light of the conflicting findings about the differential impacts of WCF techniques (e.g., Bitchener & Knoch, 2010b; Sheen, 2007), the first research question is as follows:

RQ1: What are the differential impacts of three WCF techniques: direct, indirect only, and indirect plus metalinguistic explanations on L2 accuracy?

Because only few studies (i.e., Bictchener *et al.*, 2005; Benson & DeKeyser, 2018; Ferris, 2006; Shintani *et al.*, 2014; Van Beuningen *et al.*, 2012) have examined the interactions between error category and WCF technique and have yielded inconclusive results, the second research question is:

RQ2: How are the differential effects of direct versus indirect WCF on L2 accuracy moderated by error category?

Given that learners' individual differences have been rarely accounted for and in light of the theoretical arguments suggesting that the effectiveness of FFI hinges upon learners' varying proficiency and language analytic ability, the third question as follows:

RQ3: How are the differential effects of three WCF techniques on L2 accuracy moderated by learner proficiency and language analytical ability?

Chapter 3

Methodology

3 Methodology

The present study sets out to examine the relative merits of three different WCF techniques: one direct technique (i.e, providing the correct form without metalinguistic clues) and two indirect techniques (i.e., indirect only and indirect with metalinguistic clues) and to determine the mediating effects of error category, learner proficiency level and language analytic ability. The coming sections provide a description of its methodology, more specifically of (1) the context in which this quasi-experimental study took place; (2) the teachers and learners who participated in this study; (3) the different experimental conditions that were compared; (4) the language features that were targeted; (5) the experimental treatment that was implemented; (6) the data collection instruments which were used; (7) the experimental procedure that was followed and finally (8) the data analysis which was undertaken.

3.1 Research context

The present study took place in four secondary level-three[11] Enriched French classes in Quebec. According to the MEES's *Progression of Learning in Secondary School* (2010), Enriched French classes, as opposed to core French classes, are generally meant for anglophone learners who have either studied in French immersion programs in their elementary schools and/ or who have achieved advanced scores at the end of their regular FSL program in elementary school. Besides taking their FSL class, learners in the Enriched French program are encouraged to enroll in content classes, such as social sciences, which are taught in French. It is assumed that by the end of their schooling in the Enriched French program, learners would have "acquired

[11] Secondary level-three is equivalent to grade 9.

different strategies as well as vast vocabulary, general and specialized, that is relevant to other disciplines taught in French" (MEES, 2010, personal translation).

The Enriched French program adopts a communicative approach to language teaching. In other words, and as promoted by the MEES's "Progression of Learning," it aims to help learners deepen their knowledge and use their skills in communicative situations that are both "meaningful and authentic" (p. 5; personal translation). For this reason, learners are exposed to "multiple occasions for using French in diverse contexts so that they can achieve a bilingual-like competency" (p. 5; personal translation).

3.2 Participants

Three FSL teachers and their respective classes participated in this study (a total of 4 classes and 93 students). Precautions were made to recruit instructors from different high-schools in different school boards in the boroughs of Montreal in order to reduce potential data contamination. As a result, three teachers from two different school boards were retained. Given that intact classes participated in this study, it was not possible to control for gender make-up. Each one of the four classes was assigned to an experimental condition as described in the following section.

3.3 Experimental conditions

The study seeks to compare the effectiveness of three different WCF techniques on FSL learners' acquisition. The four participating classes constituted three experimental groups and a comparison group. During the experimental intervention, the comparison group did not receive any form of WCF. Each one of the three experimental classes was assigned to one WCF technique that corresponded the most to their teacher's WCF regular practices (see Table 3).

Table 3

Number of Participants per Experimental Condition

Experimental condition	*n* (at the onset)	*n* (at the end)
Comparison (no WCF)	22	14
Direct	20	17
Indirect only	25	16
Indirect + metalinguistic explanations	26	18
Total *n*	93	65

The first experimental group received direct WCF, i.e., the provision of the correct form. The second experimental group was provided with indirect WCF in the form of underlining and/or circling. The third experimental group was given indirect WCF plus metalinguistic explanations, i.e., visual or textual metalinguistic indications (see Table 4). It should be noted that all feedback instances, including the metalinguistic explanations (ME), were provided in French. These conditions were chosen for different reasons. First, and as indicated in Ammar *et al*'s (2016) descriptive study, French as a second language (FSL) teachers including enriched French teachers flag their learners' errors primarily with indirect WCF (58.66% of the cases) and to a lesser degree with direct WCF (37,47%). In other words, both direct and indirect WCF are used in FSL classrooms. More importantly, and unlike French mainstream teachers who prefer coded WCF, FSL teachers, both elementary and secondary, prefer non-coded WCF that takes the form of underlining: in fact, 88.9% of their indirect WCF is provided via underlining (p. 12; p. 58). Second, and as highlighted in the previous literature review, conflicting findings have

emerged from the few inquiries comparing direct to indirect WCF (i.e., Van Beuningen *et al.*, 2012 versus Ferris, 2006), which warrants further research into the relative merits of either WCF technique. The scarcity of empirical studies comparing the relative merits of different operationalizations of indirect WCF is even more noticeable (the only exceptions include Bitchener & Knoch, 2010b and Chandler, 2003).[12] Consequently, and in attempt to fill in this gap in WCF research and to address both ecological and methodological concerns, a third type of WCF, i.e., indirect plus ME, is included. Its ecological validity draws from FSL secondary learners' preference for obtaining metalinguistic clues; in fact, 40.5% of the interviewed students held favorite views about metalinguistic WCF (Ammar *et al.*, 2016; p. 68). Moreover, including this third type of WCF, is motivated by the scarcity of studies examining the effectiveness of indirect WCF with metalinguistic explanations and by the inconclusive findings regarding the potential advantage of supplying metalinguistic clues on the other hand (e.g., Bitchener *et al.*, 2005, Bitchener & Knoch, 2009; Sheen, 2007; Stefanou & Révész, 2014).

[12] Although Ferris and Roberts (2001), Robb et al., (1986) have also investigated two types of indirect WCF, they focused only on their impacts on improving learners' revisions of their annotated drafts. In other words, they have not evaluated their effectiveness on long-term accuracy.

Table 4

Experimental Treatments

Experimental Condition	Example
Direct WCF	sa meilleure amis sa meilleure amie
Indirect only	sa meilleure amis
Indirect plus ME	sa meilleure amis (singulier ou pluriel? féminin ou masculin?)
Comparison: no WCF	sa meilleure amis

3.4 Targeted features

In order to answer the second research question, i.e., how the effectiveness of the WCF techniques varies across error type, the WCF provided targeted different errors, which are selected based on Ammar *et al.*'s (2016) findings. Among other things, Ammar *et al.* indicate that FSL secondary school students' most frequent errors correspond to two major linguistic categories, namely grammatical morphology and syntax. More specifically, their errors correspond to agreement in the noun phrase (ANP), agreement in the verb phrase (AVP) — both of which are subcategories of grammatical morphology — as well as noun phrase (NP) structure, verb phrase (VP) structure and homophones, which are syntactical errors. In light of these findings and in accordance with Boivin and Pinsonneault's (2016) taxonomy, the following linguistic features were targeted in the present study. Grammatical morphology errors included agreement within the noun phrase (gender and plural marking on the noun, pre-determiner agreement, determiner agreement, and adjective agreement) and agreement within the verb

phrase (agreement between the subject and the verb, the subject and the past participle, the subject and the subject complement, as well as the object and the past participle). Syntactical errors included noun phrase structure (absence of the head noun or of the determiner; error in the placement of the adjective), verb phrase structure (the choice and position of auxiliary) and grammatical homophones[13] (see Table 5)

[13] The categorization of the errors targeted in the present study is inspired by — yet slightly departs from — Boivin and Pinsonneault's (2016) taxonomy.

Table 5

Targeted Features

Error Category	Error Sub- category	Specific targeted forms	Examples
Grammatical morphology	Agreement in the noun phrase	Gender and number marking on the noun	* beaucoup de barrière
		Pre-determiner agreement	* tout ces émotions
		Determiner agreement	* la bricolage
		Adjective agreement	* deux jeune garçons
	Agreement in the VP	Agreement between the subject and the verb	* plusieurs personnes peut
		Agreement between the subject and its complement	* Ils deviennent très triste
		Agreement between the subject and the past participle	* Ils sont devenue pessimist
		Agreement between the object and past-participle	* elle les a changé
Syntax	Structure of the NP	Absence of the head noun	* elle aidait plusieurs
		Absence of the determiner	* Sa famille et amis

	Error in the placement of adjective	* un âge jeune
Structure of the VP	Choice of auxiliary	* Elle a devenu
	Position of the auxiliary (mostly with pronominal verbs)	* elle a s'installé
Homophones	Grammatical homophones	* Elle leurs a montré

N.B. All examples are taken from participating students' texts.

3.5 Experimental intervention

Students in the experimental and comparison groups completed three writing cycles. Each cycle consisted of three stages. First, students completed a text-reconstruction task. Following Sheen (2007), text-reconstruction tasks were used in this study to control for ideational dimensions (i.e., topical content and lexical complexity). In total, five different text-reconstruction tasks were designed for the purposes of this study. The same reconstruction task served as both pre-test and immediate post-test (see Appendix A); another reconstruction task was used for the delayed post-test (see Appendix B), and the remaining three tasks were used during the experimental intervention (see Appendix C).

All five tasks were designed by a group of researchers, which included the author of this dissertation, another doctoral candidate and their supervisor. The task-design procedure consisted of three steps. First, and after consulting the participating teachers, the research group decided on some of the themes that would be of interest to high-school learners. The themes kept included art as a tool of change, humanitarian engagement, young Quebecers' aspirations and the refugee crisis. Second, the researchers searched for newspaper articles that dealt with those themes and made a few modifications that included simplifying complex sentences or lexical items and uniformizing text lengths. Two other graduate research assistants analyzed the proposed texts to identify the nouns, verbs, adjectives and homophones. This step was undertaken to ensure (1) that all five texts had comparable distributions of nouns that vary in number and gender and of verbs that require either the auxiliary 'to be' or 'to have' and (2) that the other forms (such as predeterminers and homophones) occur almost equally across all texts. In the case of some discrepancies (e.g., the use of pronominal verbs), more adaptations were undertaken (e.g., changing "voyager" with "se rendre à", changing "être" with "se sentir"). Third, the author of this

dissertation and the other doctoral candidate drafted the instructions and guidelines that were handed to students and instructors, respectively (see Figure 5).

All written productions were collected at the end of each class period. All students' texts, except for those written by the comparison group, were then annotated by the author of this dissertation and another research assistant with either direct WCF, underlining or metalinguistic clues depending on the experimental condition to which learners were assigned. Third, and given the importance of revision in accentuating the effectiveness of WCF (Chandler, 2003; Shintani *et al.*, 2014), students were given their texts back and were asked to revise their errors. Those in the treatment conditions were asked to incorporate the provided WCF; they were explicitly instructed to revise the marked errors without rewriting the whole texts. The comparison group was asked to re-read their texts and make any changes they deemed necessary (without rewriting the whole text either). Following Van Beuningen *et al*., (2008) and Shintani *et al.*, (2014), all groups were given 15 to 20 minutes to revise their drafts, after which their drafts were collected again. During the writing and revision phases, students were not allowed to use dictionaries or any other grammar resources.

3.6 Data collection instruments

Apart from the text-reconstruction tasks that were used to measure students' mastery of the targeted linguistic features before the intervention started (pre-test), immediately after it ended (immediate post-test) and three weeks later (delayed post-test), a language analytical ability task was used to gauge students' language analytical abilities.

Language analytic ability tasks: This instrument was based on the Words in Sentences subset of Carroll's (1967) MLAT test. The Words in Sentences subset is meant to measure

learners' sensitivity to grammatical structure and their inductive language. In this subset, learners are asked to read a key sentence and to identify the word in a second sentence that functions the same way as the capitalized word in the key sentence. The language analytic ability tasks used in this study sought to overcome two major limitations of the Words in Sentences subset: (1) the Words in the Sentences subset is not in line with the principles of Modern Grammar, which analyzes sentences in terms of syntactic phrases rather than individual words and (2) the Words in the Sentences subest exhibits a misleading and confusing conflation between function (e.g., subject) and word class (e.g., noun phrase). The latter is illustrated in example 9 from the MLAT test.

Example 9

In each of the following questions, we will call the first sentence the key sentence. One word in the key sentence will be underlined and printed in capital letters. Your task is to select the letter of the word in the second sentence that plays the same role in that sentence as the underlined word in the key sentence. Sample: JOHN took a long walk in the woods. Children in blue jeans were singing and dancing in the park. A B C D E You would select "A." because the key sentence is about "John" and the second sentence is about "children."

The language analytic ability tasks consisted of two sections: one for functions and one for word classes. The "Functions" task includes 12 items, two for each one of the following functions: *adverbials, direct objects, indirect objects, subjects, subject complements* and *verbs*.

The “Word Class” comprises 16 items, two for each one for the following classes: *nouns, adjectives, adverbs, verbs, prepositions, determiners, pronouns, conjunctions*.

In the “Functions” task, students were asked to identify which word or group of words has the same function as the word or group of words that is bolded and capitalized in the numbered key sentence (see example 10)

Example 10: Instructions of the “Functions” task

In the following sentences, identify which word or group of words has the same function as the word or group of words that is **BOLDED AND CAPITALIZED** in the numbered key sentence. The functions examined in this test are adverbials, direct objects, indirect objects, subjects, subject complements and verbs.

Example

0. **THIS TEST** might be fun.

- She never arrives late.

A B C D

Answer: A. Both are subjects.

In the “Word Class” task, participants were asked to identify which word belongs to the same family as the word that is bolded and capitalized in the number key sentence (see example 11).

Example 11: Instructions of the "Word Class" task

In the following sentences, identify which word belongs to the same family/is of the same nature as the word that is **BOLDED AND CAPITALIZED** in the numbered key sentence. The word classes examined in this test are *adjectives, adverbs, coordinators, determiners, nouns, prepositions, pronouns and verbs*.

0. I lost **MY** wallet.

- The teacher is absent today.

A B C D

Answer: **A**. Both words are determiners.

It should be noted that earlier versions of both tasks (i.e., the function and word class tasks) were piloted with English-speaking experts (i.e., L2 professors at Concordia University) and with similar groups of high- school learners (n = 190) (Appendix E). A first item analysis indicated that the reliability coefficient was 0.57, which allowed the researcher to identify the items that did not discriminate well (for example items 5 and 9 in Appendix E). In light of this analysis, revisions were made, and the problematic items were modified. Subsequently, revised versions of both tasks were then used with the groups participating in this study and with different grade level students (n = 185) (see Appendix D). The reliability coefficient of this instrument is r = .83.

3.7 Procedure

The study included four phases: a pre-test, a six-week intervention (two weeks for each writing cycle), an immediate post-test (one week after the treatment) and a delayed post-test (three weeks after the treatment). One week before the experimental intervention, all learners from the experimental and comparison conditions completed the pre-test (i.e., text reconstruction task) and took the language analytic ability tasks. The experimental intervention, which started

one week later, consisted of three writing cycles; each writing cycle was spread over two weeks and included the text-reconstruction task, the provision of WCF (for the experimental conditions) and revision during the subsequent week. The comparison group participants did not receive WCF but were asked to revise their drafts (see Figure 4).

Week 1	Day 1: Language Analytic Aptitude Tasks Day 2: Pret-test: writing task 1 (text-reconstruction 1)
Week 2	Intervention began: Writing task 2 (text-reconstruction 2)
Week 3	WCF (for the three experimental groups) Revising task 2 (for all four groups)
Week 4	Writing task 3 (text-reconstruction 3)
Week 5	WCF (for the three experimental groups) Revising task 4 (for all four groups)
Week 6	Writing task 4 (text-reconstruction 4)
Week 7	WCF (for the three experimental groups) Revising task 4 (for all four groups)
Week 8	Immediate post-test (same writing task 1)
Week 11	Delayed post-test (text- reconstruction 5)

Figure 4. Experimental procedure

For each one of the five writing tasks (evaluation and intervention), learners were given a handout that explains the instructions in French and provides them with a series of key nouns, verbs, adjectives, determiners and pre-determiners that were used in the text they would listen to (see Figure 5). All words, which were taken from the texts learners would listen to, were provided in their bare forms (nouns and adjectives in the singular form; verbs in the infinitive). Students were requested to use a certain number of words from each form category to increase the chances of using the target features and to maximize the comparability of the experimental conditions. The teacher read out loud the instructions, explaining that students would listen to a text three times (once to understand its general ideas and twice to take notes of the main elements), after which they would have to construct the text using a minimum of 40 words from the list provided. Before reading the text to her students, the teacher checked if they had any questions about the meanings of the provided words or if they had any other clarification requests. Students were reminded that they would not have access to their dictionaries or grammar handouts and that their writing is to be completed individually. To control for the length of their compositions, learners were asked to write a minimum of 250 words. Teachers spent 10 minutes in average to read aloud the texts three times, after which learners were given 45 minutes to complete their text-reconstruction tasks.

- You will listen to a text about two women, Françoise Collin and Saundra Baly. These two women met each other at an association called "La Rue Des Femmes''.
- You will then write an essay in which you reconstruct the text to which you listened (a minimum of 250 words). You need to follow the following steps:
  1. First, listen to the text to understand its general ideas
  2. Second, listen to the text again to take note of the elements that you consider necessary for the text-reconstruction
  3. Listen to the text one last time to make sure you have all the necessary elements
  4. Write your text-reconstruction using at least 40 words of the following list:

- Patience, femme, barrière, jeu, découverte, secours, efficacité, souffrance, peinture, émotion, artiste, organisme, œuvre, guérison, énergie **(use at least 10 words)**
- Démuni, compulsif, quotidien, grand, talentueux, personnel, lumineux, conscient, emprisonné, fondamental, déconnecté, agressif, dangereux, bon, négatif, communicatif **(use at least 10 words)**
- Fréquenter, fouler, s'éloigner, s'initier, débuter, passer, s'installer, mettre, se retrouver, transformer, changer, devenir, exprimer, se parler, écouter **(use at least 10 words)**
- Leur, leurs, a, à, on, ont, ses, ces, c'est, s'est, là, la, l'a **(use at least 10 words)**
- Quelques, plusieurs, tout, chaque, aucun, chacun **(use at least 10 words)**

Figure 5. A translated example of the instructions of the text-reconstruction task

The three writing tasks completed during the intervention phase were annotated by the author of this dissertation and another doctoral candidate. The author of this dissertation provided direct WCF and indirect WCF (Appendix F and Appendix G, respectively), while the other doctoral candidate gave indirect WCF with metalinguistic clues (Appendix H). To ensure that the treatment was implemented as intended, the two researchers made photocopies of five students' pre-tests, separately identified learners' errors and provided feedback. The discrepancies and

ambiguous cases were then discussed with their supervisor.[14] They then double-checked one another's feedback on all students' texts to minimize any unintentional omissions. A total of 223 students' texts were annotated throughout the intervention, and each one of the texts was verified by both researchers.

It is true that having research assistants provide WCF may reduce the ecological validity of this study. However, it was primarily motivated by practical reasons. It was decided not to ask teachers to provide feedback on their students' writing texts so as not to add to their workload, to make their participation more manageable, and most importantly, to limit the potential variability between teachers. As highlighted by Ferris (2006), providing feedback by the researchers "ensures greater consistency in treatment and enables assessment of the effects of feedback without this potentially confounding variable [of teacher variability]" (p. 93), accentuating thus its methodological validity.

Besides providing the WCF on learners' texts, the author of the present study and two other research assistants attended most class periods during which learners either completed or revised their writing tasks. During the revision sessions, learners were given between 15 and 20 minutes and were asked to revise their texts individually; they were not allowed to check their dictionaries or any other resources. The use of these tools was excluded to better tease apart the effects of WCF.

---

[14] Some of the ambiguous cases included examples such as (**sa* addiction, **au* rivière, chacune **a s'amusé,* elle *à *découvrié)

3.8 Data analyses

To measure learners' performance in the pre and post-tests, the researcher identified and coded all targeted errors in all pre-test and post-tests copies (see examples 12 and 13 below). To ensure inter-rater reliability, two research assistants identified all targeted errors in 10% of the data subset (n = 20 copies); the inter-rater reliability was 97%. Another doctoral candidate, who specializes in French grammar analysis, verified all the coding done by the researcher. The few discrepancies that were found were then discussed and resolved after consulting with the dissertation supervisor.

Examples 12 and 13:

- Les femme ont devenu : 2 errors (1 for noun agreement and 1 for the choice of auxiliary)
- Deux garçons qui sont considéré aggressive: 3 errors (1 in the past participle agreement and 2 in the adjective agreement: wrong gender and wrong number).

Following Ferris (2006) and van Beuingen *et al.*, (2012), two error ratios were used. The first one corresponds to overall inaccuracy and is computed by first dividing the number of total errors by the number of total words and then multiplying the result by 100. The second one corresponds to inaccuracy per error category and is calculated by first dividing the number of errors in that category by the number of total words in a text and then multiplying the result by 100 (see Table 6).

Table 6

Error and accuracy ratios

Ratio	Formula
Error ratio	(Total number of errors / total number of words) * 100
Overall accuracy	100 - error ratio
Error ratio per category	(Incorrect use of targeted forms within a given category/ total number of words)*100
Accuracy per error category	100 - error ratio per category

Learners' language analytic ability was determined based on their scores in the language analytical ability tasks. First, learners' individual scores were calculated as follows: 1 point for a correct answer; 0 for a wrong answer. Second, ratios were computed by dividing learners' individual scores by possible total points (i.e., 28).

Learners' proficiency levels were also determined using their scores at the pre-test. In each one of the experimental groups, learners who scored higher than their respective group overall mean were classified as higher proficiency learners, while those who scored below their group mean were classified as lower proficiency learners.

To answer the first research question, which examines the differential impacts of three WCF techniques (i.e., direct, indirect only, indirect plus metalinguistic clues) on L2 development, a repeated-measures 3 (time)* 4 (group) ANOVA with time (pre, post1, post-2) as within subject variable and experimental condition (no WCF, direct WCF, indirect WCF only, indirect WCF plus metalinguistic clues) as between subject variable was run on overall accuracy scores. To answer the second research question, which explores the moderating effect of error categories, mixed methods ANOVA with error category (3 levels) and time (3 levels) as within subject

variables and experimental condition (4 levels) as between subject variable was conducted. To answer the third question, which focuses on the moderating effects of proficiency and language analytical ability, multiple linear regressions for each experimental group were carried out. In all analyses, SPSS (version 6) was used and the significance α was kept at 0.05.

More details about the analyses and the results are reported next.

Chapter 4

Results

4 Results

The fourth chapter consists of the descriptive and statistical analyses undergone to answer the three research questions examined in this quasi-experimental study. Consequently, it is divided into three major parts, one part for each research question. The first part consists of the descriptive and statistical analyses carried out to explore the effectiveness of three WCF techniques, i.e., direct, indirect only WCF, and indirect WCF that is combined with metalinguistic explanations (henceforth, indirect + ME) on learners' overall accuracy. The second part includes the descriptive and statistical analyses conducted to examine the responsiveness of three error categories (agreement in the noun phrase, agreement in the verb phrase and homophones) to each one of the WCF techniques over time. Last, the third part describes the regression analyses carried out to explore the mediating effect of learners' proficiency and language analytic abilities on the effectiveness of WCF techniques.

All analyses were conducted using analyses of variance in a combination of fully between, repeated measure and mixed design.

4.1 The effectiveness of WCF techniques on overall accuracy

The first section aims at deciphering whether treatment groups outperformed each other and the control group. Furthermore, the study examines whether the post-test scores for each group varied based on the treatment they received and whether the effects of the treatment persisted over a second follow-up testing session.

4.1.1 Checking ANOVA assumptions

Before running these analyses, we first examined whether the data for each experimental condition and across all groups met the assumptions for ANOVA (Meyers, Gamst & Guarino,

2013). The first two assumptions concern how the dependent (overall accuracy) and independent variables (WCF technique) were operationalized. The dependent variable was measured at the continuous level (from 0 to 100) at three points of time (pre, post-test 1, post-test 2), and the independent variable consisted of four categorical levels (no WCF, direct, indirect only and indirect + ME). The third assumption requires that no significant outlier be present in any of the conditions. To detect univariate outliers, all learners' overall scores were transformed to z-scores; cases with absolute z-scores that are larger than 3.29 can be potential outliers (Tabachnick & Fidell, 2013). All standardized scores for the overall scores fell below the threshold of 3.29; as such, no univariate outlier was detected. The fourth assumption refers to the normal distribution of the dependent variable (overall scores) across the independent variables. To check univariate normality, the skew and kurtosis indexes for all variables were computed separately in each group (see Table 7). None of the variables, in any of the four groups across the three testing times, had an absolute value of skew index greater than 3 or an absolute value of kurtosis greater than 10 (Kline, 2009).

Finally, to check for the sphericity assumption, SPSS was used to run Mauchly's test of sphericity, which tests whether the variances of the differences between all combinations of related groups are equal. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 3.021$, $p = .221$.

Table 7

Descriptive Statistics of Overall Accuracy Scores

		N	Mean	SD	Min	Max	Skewness	Kurtosis
Pre-test accuracy	Comparison	14	83.90	6.13	73.61	92.55	-0.32	-0.75
	Direct	17	84.14	4.16	77.36	92.36	0.29	-0.52
	Indirect only	16	86.14	3.50	81.22	91.22	0.00	-1.60
	Indirect + ME	18	86.78	4.38	79.10	94.19	-0.30	-0.91
Post-test 1 accuracy	Comparison	14	83.89	5.26	77.42	94.24	0.62	-0.65
	Direct	17	83.85	2.90	77.92	88.54	-0.47	-0.19
	Indirect only	16	87.60	3.10	83.62	94.16	1.14	0.56
	Indirect + ME	18	88.25	4.14	80.42	94.12	-0.26	-0.91
Post-test 2 accuracy	Comparison	14	80.68	6.08	65.63	89.72	-1.04	1.96
	Direct	17	84.28	6.69	70.13	93.06	-0.75	0.01
	Indirect only	16	85.83	4.45	77.60	93.16	-0.10	-0.39
	Indirect + ME	18	87.35	4.71	79.08	94.27	-0.20	-1.10

4.1.2 Pre-test results

As shown on table 7, descriptive statistics of overall scores at the pre-test for each one of the four experimental conditions indicated that the indirect only and indirect + ME groups had higher means than the comparison and the direct group.

A one-way ANOVA on pre-test overall scores revealed no overall effect $F(3, 61) = 1.612$, $p = .196$, $\eta^2 = .073$. Subsequent pairwise comparisons (Tukey HSD) between the groups also showed no statistically significant differences between any of the four groups. Consequently, all four groups were comparable at the onset of the study (see Table 8).

Table 8

Pairwise Comparisons of Pre-Test Overall Mean Scores

		Mean Difference	Std. Error	Sig.	Hedge's g
Comparison	Direct	-.24	1.65	.88	-.04
	Indirect only	-2.24	1.67	.19	-.44
	Indirect + ME	-2.88	1.63	.08	-.53
Direct	Indirect only	-2	1.59	.21	-.50
	Indirect + ME	-2.64	1.55	.09	-.60
Indirect only	Indirect + ME	-.64	1.57	.68	-.15

Note. For all effect sizes calculated for between group comparisons, Hedge's *g* was used as it outperforms Cohen's *d* when sample sizes are smaller than 20 (Kline, 2009).

4.1.3 Post-test results

Since the pretest scores were equivalent across groups, it is possible to evaluate the effectiveness of the treatment conditions by comparing post-test scores to the pre-test scores. The

goal is to determine which, if any, treatment conditions perform better at post-test 1 and 2 compared to the pre-test (where everyone is equivalent).

A repeated-measures 3 (time)* 4 (group) ANOVA with time (pre, post 1, post 2) as within subject-variable and experimental condition (no WCF, direct WCF, indirect only, indirect + ME) as between subject variable was run on overall accuracy scores. Results showed no statistically significant main effect of time $F(2, 61) = 2.84$, $p = .062$, $\eta^2 = .02$, nor a statistically significant interaction between time and experimental condition $F(6, 61) = 1.50$, $p = .184$, $\eta^2 = .08$. However, a main effect of experimental condition was found $F(3,61) = 4.77$, $p = .005$, $\eta^2 = .19$.

The absence of a statistically significant main effect of time does not dictate one to abandon post-hoc comparisons. In fact, main effects seldom provide useful information pertaining to individual group differences (Kline, 2009). This is especially true when groups are formed by a small sample size ($n < 20$). As such, to fully understand the effects of time, it is critical to analyse whether differences over time occurred within each group separately. In addition, one needs to evaluate whether group differences occur within time points.

Separate analyses were then conducted to (1) examine the within-group differences over time for each experimental condition individually and (2) explore the potential between-group difference at each one of the testing times.

4.1.4 Change over time for each group

To answer the first question (i.e., whether the experimental groups changed over time), a one-way repeated measures ANOVA for each experimental condition was performed.

For **the comparison** (no WCF) group, descriptive statistics (see Table 7) indicated that the mean scores at the pre-test and post-test 1 were almost the same, but they decreased at post-test 2. A one-way repeated measures ANOVA revealed no statistically significant differences between

participants' overall scores at the three times $F(2, 26) = 3.04$, $p = .065$, $\eta^2 = .18$. Pairwise comparisons showed no statistically significant difference between the overall scores at the pre and post-tests (see Table 9). This suggests that learners' accuracy did not change statistically over the course of the study.

Table 9

Pairwise Comparisons between the Overall Accuracy Scores of the Comparison Group

Time		Mean Difference	Std. Error	Sig.
Pre-test	Post 1	.02	1.42	.99
Pre-test	Post 2	3.22	1.55	.06
Post 1	Post 2	3.21	1.54	.06

For **the direct WCF group**, descriptive statistics showed a small decrease for overall accuracy from pre-test to post-test 1 and a small increase from post-test 1 to post-test 2 (see Table 7). A one-way repeated measures ANOVA revealed no statistically significant differences between the overall scores at the three times $F(2, 32) = .051$, $p = .95$, $\eta^2 = .003$. Likewise, pairwise comparisons showed no statistically significant differences at the three time points (see Table 10). In other words, the overall accuracy for learners receiving direct WCF did not change statistically over the course of the study.

Table 10

Pairwise Comparisons between the Overall Accuracy Scores of the Direct WCF Group

Time		Mean Difference	Std. Error	Sig.
Pre-test	Post 1	.30	.99	.77
Pre-test	Post 2	-.14	1.43	.93
Post 1	Post 2	-.43	1.65	.80

For the **indirect WCF only** group, a comparison of the group means at the three testing times showed some increase from pre-test to post-test 1 and some decrease from post-test 1 to post-test 2 (see Table 7). In other words, the initial increase from pre-test to post-test 1 did not carry over to post-test 2. A one-way repeated measures ANOVA also revealed no statistically significant differences between the overall scores at the pre-test, the immediate post-test or the delayed post-test $F(2, 30) = 1.71$, $p = .198$, $\eta^2 = .10$. Put differently, the change in overall accuracy for learners receiving indirect WCF only over the course of the study (i.e., from the pre-test to the post-tests) was too small to reach statistical significance.

Table 11

Pairwise Comparisons between the Overall Accuracy Scores of the Indirect WCF only Group

Time		Mean Difference	Std. Error	Sig.
Pre-test	Post 1	-1.46	1.02	0.17
Pre-test	Post 2	.31	1.11	0.78
Post 1	Post 2	1.77	0.93	0.08

For the **indirect WCF + ME group**, a comparison of the mean scores revealed that the overall scores increased the pre-test to post-test 1, but they decreased from post-test 1 to post-test

2 (see Table 7). Yet, unlike the other groups, scores at the post-test 2 were higher than those at the pre-test. A one-way repeated measures ANOVA indicated statistically significant differences between the overall scores at the pre-test, post-test 1 or post-test 2 $F(2, 34) = 3.48$, $p = .042$, $\eta^2 = .16$. Pairwise comparisons showed that the increase from pre-test to post-test 1 was statistically significant $p = .032$, yet the increase from pre-test to post-test 2 was not significant $p = .302$; neither was the decrease from post-test 1 to post-test 2 (see Table 12).

Table 12

Pairwise Comparisons between the Overall Accuracy Scores of the Indirect + ME Group

Time		Mean Difference	Std. Error	Sig.
Pre-test	Post 1	-1.47*	.63	.03
	Post 2	-.57	.53	.30
Post 1	Post 2	.90	.10	.10

* The mean difference is significant at the .05 level.

Overall, the separate within-group analyses show that while both the direct and comparison groups decreased from pre-test to post-test 1, both indirect only and indirect + ME groups increased in overall accuracy from pre-test to post-test 1 (see Figure 6). The change reached statistical significance for the indirect + ME group only. A comparison between scores at the post-test 1 and post-test 2 showed that overall accuracy decreased for all groups (except for the direct group), but such a decrease did not reach statistical significance for any of the four groups. Finally, a comparison of the mean scores at the pre-test and post-test 2 for all groups showed that overall accuracy for the comparison group decreased quite remarkably (from 83.90 to 80.63), remained almost the same for the direct and the indirect only groups, and slightly increased for

the metalinguistic group (from 86.78 to 87.35). These changes from pre-test to post-test 2 did not reach statically significant difference for any of the groups.

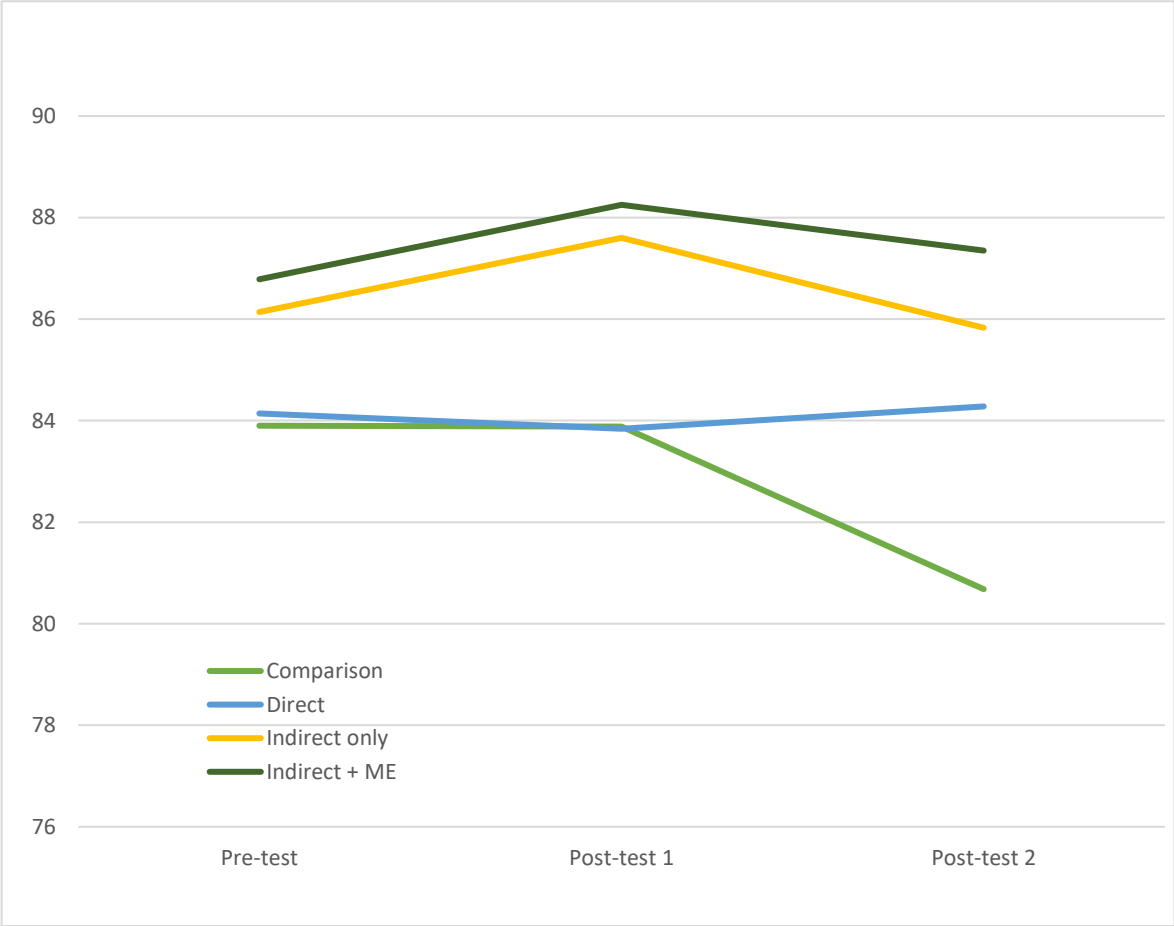


Figure 6. Learners' overall accuracy over time

4.1.5 Between- group differences

As mentioned earlier, a one-way ANOVA on pre-test overall accuracy scores showed no statistically significant differences between all four groups $F(3, 61) = 1.612$, $p = .196$, $\eta^2 = .073$. Pairwise comparisons between the groups also indicated no statistically significant difference between any of the four groups at the onset of the study.

To explore if the groups remained comparable after the treatment, one-way ANCOVAs were performed on post-test 1 and post-test 2 scores. ANCOVA was deemed useful because it " increases the power of an *F* test for a main effect" (Tabachnick & Fidell, 2013, p. 197). The one-way ANCOVA on post-test 1 scores showed that when the initial variance at the pre-test was controlled for, there were statistically significant differences between the four groups at post-test 1 $F(3, 61) = 12.92$, $p = .000$, $\eta^2 = .463$. Post-hoc comparisons indicated that both the indirect only and metalinguistic groups differed statistically significantly from both the direct and the comparison groups (see Table 13). No statistically significant differences were detected between the comparison and the direct group on the one hand, or the indirect only and indirect + ME groups on the other hand.

Table 13

Pairwise Comparisons between Overall Mean Scores at Post-Test 1

Experimental condition		Mean Difference	Sig.	Hedge's *g*
Comparison	Direct	.15	.90	.009
	Indirect only	-2.65*	.03	- .85
	Indirect + ME	-3.00*	.02	- .912
Direct	Indirect only	-2.81*	.02	-1.22
	Indirect + ME	-3.157**	.01	-1.19
Indirect + ME	Indirect only	.35	.76	.17

* The mean difference is significant at the .05 level.
**The mean difference is significant at the 0.01 level

The one-way ANCOVA on post-test 2 scores showed that when the initial variance at the pre-test was controlled for, there were statistically significant differences between the four groups at post-test 2 $F(3, 61) = 2.816$, $p = .047$, $\eta^2 = .123$. Post-hoc comparisons indicated that while

there were no statistically significant differences between the three WCF groups, all the WCF treatments outperformed significantly the comparison group (see Table 14).

Table 14

Post-hoc Comparisons between Overall Mean Scores at Post-Test 2

Experimental condition		Mean Difference	Sig.	Hedge's *g*
Comparison	Direct	-3.43*	.04	- .52
	Indirect only	-3.59*	.04	- .66
	Metalinguistic	-4.66**	.01	- .85
Direct	Indirect only	-0.16	.92	-.02
	Metalinguistic	-1.23	.44	-.85
Metalinguistic	Indirect only	1.07	.5	.22

* The mean difference is significant at the .05 level.
**The mean difference is significant at the 0.01 level

Put together, the separate within-subject and between-subject analyses suggest that only indirect WCF techniques (indirect only and indirect + ME) brought about accuracy gains from pre-test to post-test 1, and that those gains reached statistical significance only for the group receiving indirect plus metalinguistic explanations. The short-time effects of indirect WCF were also affirmed through the between- group comparisons, which showed that when controlling for initial variation at the pre-test, the two indirect WCF groups significantly outperformed the direct and the comparison groups (see Figure 7). At post-test 2, all groups' overall accuracy scores decreased. The decrease from post-test 1 to post-2 as opposed to the improvement from pre-test to post-test 2 can be attributed to a test-practice effect and to task familiarity. At post-test 1,

students completed the same text-reconstruction task as the one at the pre-test; however, at post-test 2 they completed a different text-reconstruction task. In spite of this decrease, a comparison of students' performance at pre-test and at post-test 2 indicated that, when controlling for initial variation at the pre-test, the three experimental conditions outperformed in statistically significant manner the comparison group. This suggests that the effectiveness of WCF (as compared to no WCF) was maintained at post-test 2.

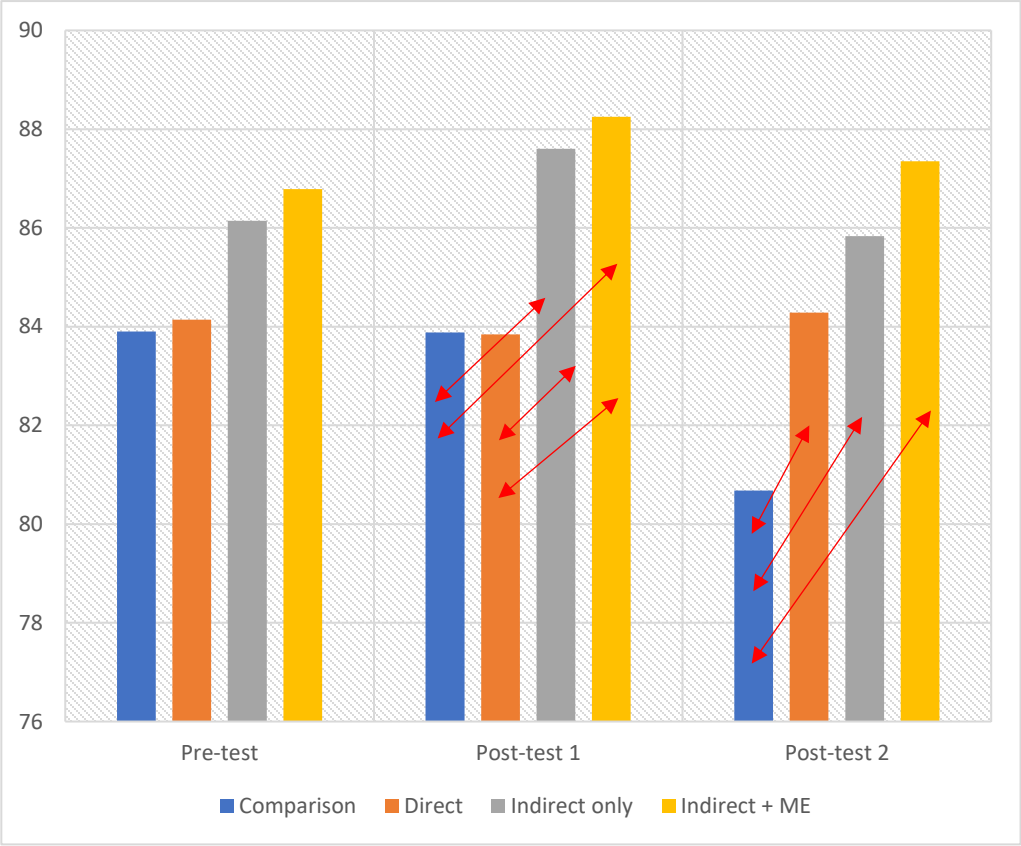


Figure 7. Comparing group overall accuracy at the three testing times
The red arrows signal statistically significant differences between the experimental conditions.

4.2 The moderating effects of error categories

The second research question is about the amenability of different error categories, i.e., agreement in the noun phrase, agreement in the verb phrase, and homophones, to the WCF types, i.e., no WCF, direct WCF, indirect only WCF, indirect plus ME WCF (see Table 15).

A mixed methods ANOVA with error category (3 levels) and time (3 levels) as within subject variables and experimental condition (4 levels) as between subject variable was conducted. Results showed that the error category was statistically significant $F(2, 59) = 23.78$, $p = .000$, $\eta^2 = .281$ that time had a statistically significant effect $F(1.641, 100.08)$[15]$= 4.64$, $p = .011$, $\eta^2 = .072$ and that there was a statistically significant interaction between time and error category $F(3.5, 213.57) = 19.58$, $p= .000$, $\eta^2 = .05$.

However, there was no interaction effect between error category and experimental condition $F(6,120) = 1.50$, $p = .182$, $\eta^2 = .071$, nor was there an interaction effect between error category, time and experimental condition $F(12,177) = 1.346$, $p = .193$, $\eta^2 = .064$. This implies that the potential effect of the different feedback conditions did not vary across error categories. Overall, results of the mixed-methods ANOVA suggest that the effect of time on accuracy scores was different across the error categories, regardless of the experimental condition.

An inspection of the graphs generated by SPSS of the mean scores in the different error categories from pre-test to the post-tests for each one of the experimental conditions indicates that while accuracy scores in the agreement in the noun phrase and in the verb phrase decreased

[15]The significance value of the Mauchy's test (for time) indicates that the main effect of time has violated the assumption of sphericity (p= 0.001). As such any effect involving time was corrected.

from pre-test to post-test 2, the accuracy in the use of homophones increased from pre-test to post 2 (see Figure 8).

Table 15

Descriptive Statistics for Accuracy per Error Sub-category

Experimental condition	Time		N	Mean	SD	Min.	Max.	Skew.	Kurtosis
Comparison	pre-test	ANP	14	93.98	2.74	88.62	99.01	-.48	.24
		AVP	14	95.66	1.81	93.06	99.21	.38	-.2
		Homophone	14	95.3	3.13	88.89	98.94	-.87	-.26
	post 1	ANP	14	94.21	2.84	90.29	99.48	.4	-.22
		AVP	14	96.47	1.44	93.53	97.97	-.96	-.42
		Homophone	14	94.64	2.06	91.47	98.43	.15	-.68
	post 2	ANP	14	90.83	4.63	81.25	96.97	-.43	-.55
		AVP	14	94.52	2.25	90.34	98.54	-.5	.21
		Homophone	14	96.35	2.66	90.78	100	-.32	-.14
Direct	pre-test	ANP	17	95.41	2.4	90.81	99.21	-.24	-.65
		AVP	17	96.66	1.67	93.7	98.92	-.55	-.57
		Homophone	17	93.05	3.04	84.91	97.22	-1.18	1.89
	post 1	ANP	17	94.8	2.27	89.95	98.96	-.28	.27
		AVP	17	96.09	1.93	91.67	99.4	-.47	.83
		Homophone	17	94.07	2.96	88.33	98.96	-.14	-.51
	post 2	ANP	17	92.76	4.4	81.82	98.37	-.91	.8
		AVP	17	95.84	1.78	92.31	98.6	-.4	-.11
		Homophone	17	96.1	3.28	86.54	100	-1.78	3.68

Indirect only	pre-test	ANP	16	95.5	2.73	89.84	99.24	-.63	-.28
		AVP	16	96.51	1.21	94.44	98.9	.18	-.43
		Homophone	16	95.27	1.8	91.27	97.32	-.72	-.31
	post 1	ANP	16	95.59	2.26	91.64	99.35	.04	-1.01
		AVP	16	96.82	1.68	93.72	99.72	-.51	-.17
		Homophone	16	95.86	1.88	91.81	98.73	-.41	-.02
	post 2	ANP	16	93.09	4.02	82.79	98.94	-1.35	2.17
		AVP	16	96.02	2.78	90.18	100	-.4	-.5
		Homophone	16	96.4	3.02	89.34	100	-1.43	1.53
Indirect + ME	pre-test	ANP	18	95.18	2.55	89.55	98.8	-1.26	1.1
		AVP	18	96.93	1.27	94.53	98.84	-.25	-1.02
		Homophone	18	95.59	2.67	89.16	98.84	-1.11	.76
	post 1	ANP	18	95.25	1.88	91.16	98.27	-.82	.28
		AVP	18	97.51	1.18	94.56	98.96	-.75	.57
		Homophone	18	96.72	2.49	91.61	99.35	-.57	-.91
	post 2	ANP	18	94.62	2.18	90.05	97.81	-.38	-.72
		AVP	18	96.17	1.98	92.6	99.25	-.16	-.72
		Homophone	18	97.42	1.67	93.75	99.7	-.62	-.19

Note. ANP: agreement in the noun phrase; AVP: agreement in the verb phrase; min.: minimum; max.: maximum; skew.: skewness

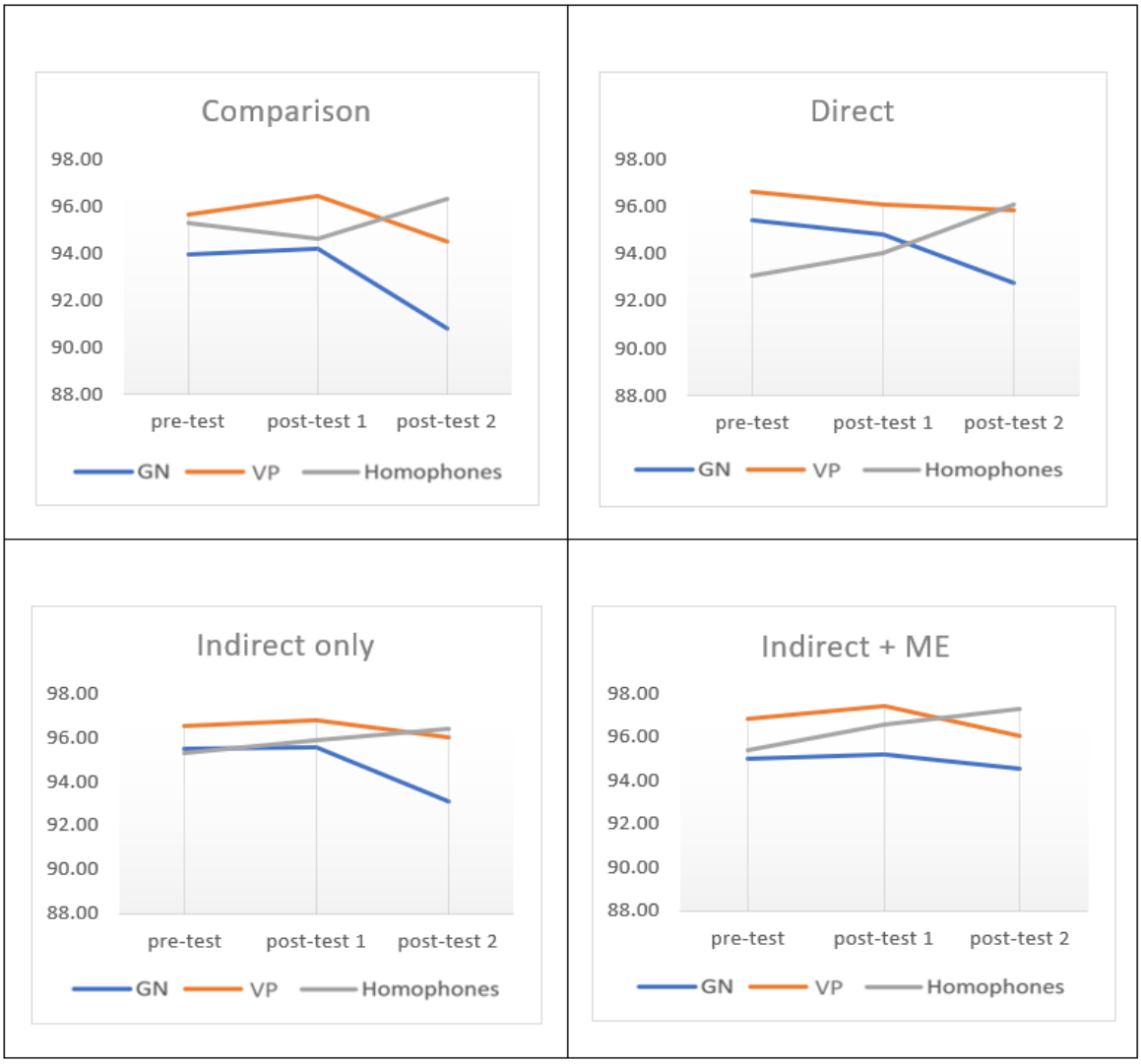


Figure 8. Accuracy per error category in the experimental groups

Considering the pattern of results in Figure 8 and the statistically significant interaction between error category and time as well as the main effect of time, error categories appear to vary across time points but only within experimental conditions. In order to determine whether error categories improved over time within groups, and whether a specific error category is better than

another, subsequent repeated measures ANOVAs for each error category and for each group were then carried out. These are presented next.

4.2.1 The comparison group

A one-way repeated measures ANOVA was conducted to compare the effect of time on scores of ANP scores in the comparison group. Results showed a significant effect of time $F(2, 26) = 6.13$, $p = .007$, $\eta^2 = .321$. Subsequent pairwise comparisons indicated that the decrease from pre-test (M = 93.98, SD = 2.73) to post-test 1 (M = 94. 21, SD = 2.84) was not statistically significant (see Table 16). However, the mean scores at the pre-test and at the post-test 1 were statistically higher than the mean scores at the post-test 2 (M = 90.82, SD = 4.62). In other words, learners' accuracy scores of the ANP in the comparison group decreased significantly from the beginning to the end of the study.

Table 16

Pairwise Comparisons between the ANP Scores for the Comparison Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-.23	.59	.7
	Post 2	3.15*	1.35	.03
Post 1	Post 2	3.38**	1.15	.01

*The mean difference is significant at the .05 level.

**The mean difference is significant at the 0.01 level

To examine whether the scores of the agreement in the verb phrase changed over time, a one-way repeated measures ANOVA was conducted to compare the effect of time on scores of the agreement in the verb phrase in the comparison group. Results showed a statistically

significant effect of time $F(2, 26) = 4.105$, $p = .02$, $\eta^2 = .24$. Subsequent pairwise comparisons indicated that the decrease from pre-test ($M = 95.65$, $SD = 1.81$) to post-test 1 ($M = 96.47$, $SD = 1.44$) was not statistically significant, neither was the decrease from pre-test to post-test 2. Nonetheless, the decrease from post-test 1 to post-test 2 ($M = 94.51$, $SD = 2.25$) was statistically significant.

Table 17 *Pairwise Comparisons between the Agreement in the Verb phrase Scores for the Comparison Group*

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-.82	.67	.25
	Post 2	1.14	.67	.11
Post 1	Post 2	1.95*	.71	.02

*The mean difference is significant at the .05 level.

To examine whether the scores of homophones changed over time, a one-way repeated measures ANOVA was conducted to compare the effect of time on scores of homophones in the comparison group. Results show no significant effect of time $F(2,26) = 2.80$, $p = .079$, $\eta^2 = .178$. Pairwise comparisons indicate that the decrease from pre-test ($M = 95.30$, $SD = 3.13$) to post-test 1 ($M = 94.63$, $SD = 2.05$) and the increase from pre-test to post-test 2 ($M = 96.35$, $SD = 2.65$) were not statistically significant ($p = .36$ & $.20$, respectively). The increase from post-test 1 to post-test 2, however, was statistically significant ($p = .02$).

Table 18

Pairwise Comparisons between the Homophone Scores for the Comparison Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	.67	.71	.36
	Post 2	-1.05	.79	.21
Post 1	Post 2	-1.71*	.69	.03

*. The mean difference is significant at the .05 level.

Overall, the comparison group (i.e., no WCF group) had its scores of the agreement in the NP and the agreement in the VP decrease statistically significantly from the beginning to the end of the study, however, its scores of homophones decreased slightly from pre-test to post-test 1 and then increased on post-test 2; the increase from post-test 1 to post-test 2 reached statistical significance. However, the change from pre-test to post-test 2 was not significant.

4.2.2 The direct WCF group

In order to determine whether error categories improved over time within the direct group and whether a specific error category is better than another, three separate one-way repeated measures ANOVA and their respective post-hoc comparisons were carried out.

First, a one-way repeated measures ANOVA was conducted to compare the effect of time on scores of the ANP. Results indicate a significant effect of time $F(2, 32) = 4.18$, $p = .024$, $\eta^2 = .207$. Subsequent pairwise comparisons indicated a statistically significant decrease from the pre-test ($M = 95.41$, $SD = 2.40$) to post-test 2 ($M = 92.75$; $SD = 4.39$). The decreases from pre-test to post-test 1 and from post-test 1 to post-test 2 did not reach statistical significance.

Table 19

Pairwise Comparisons between the ANP Scores for the Direct Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	.61	.78	.44
	Post 2	2.66*	1.05	.02
Post 1	Post 2	2.05	1.04	.07

*The mean difference is significant at the .05 level.

Second, a one-way repeated measures ANOVA was conducted to compare the effect of time on the verb phrase agreement scores shows a non- significant effect of time $F(2, 32) = 1.06$, $p = .35$, $\eta^2 = .063$. Subsequent pairwise comparisons reveal no significant differences between the mean scores at the pre-test and post-tests.

Table 20

Pairwise Comparisons between the AVP Scores for the Direct Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	.57	.55	.32
	Post 2	.82	.64	.21
Post 1	Post 2	.25	.54	.64

Finally, to examine whether the scores of homophones changed over time, a one-way repeated measures ANOVA was conducted to compare the effect of time on scores of homophones in the direct group. Results show a significant effect of time $F(2,32) = 13.36$, $p = .0$, $\eta^2 = .44$. Subsequent pairwise comparisons indicate statistically significant increases from pre-

test (M = 93.05, SD = 3.03) and post-test 1 (M = 94.07, SD =2.95) to post-test 2 (M = 96.09, SD =3.28).

Table 21

Pairwise Comparisons between the Homophone Scores for the Direct Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-1.02	.53	.07
	Post 2	-3.04**	.58	.00
Post 1	Post 2	-2.02**	.68	.01

*. The mean difference is significant at the .05 level.
**. The mean difference is significant at the 0.01 level

Overall, participants in the direct group had their scores in the ANP decrease statistically from the pre-test to post-test 2, while their scores of homophones have increased statistically during the same period. No significant change was observed for their scores of the agreement in the verb phrase.

4.2.3 The indirect only group

In order to determine whether error categories improved over time within the Indirect only group and whether a specific error category was better than another, three separate one-way repeated measures ANOVA and their respective post-hoc comparisons were carried out.

First, a one-way repeated measures ANOVA on the ANP scores showed a significant effect of time $F(2, 30) = 4.32$, $p = .02$, $\eta^2 = .22$. Subsequent pairwise comparisons indicated statistically significant decreases from the pre-test (M = 95.50, SD = 2.73) and post-test 1 (M = 95.58, SD = 2.26) to post-test 2 (M = 93.08, SD = 4.02) (see Table 22).

Table 22

Pairwise Comparisons between the ANP Scores for the Indirect only Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-.08	.71	.91
	Post 2	2.42*	1.04	.03
Post 1	Post 2	2.50*	1.11	.04

*. The mean difference is significant at the .05 level.

Second, a one-way repeated measures ANOVA on the verb phrase agreement scores showed a non- significant effect of time $F(2, 30) = .608$, $p = .55$, $\eta^2 = .039$. Subsequent pairwise comparisons revealed no significant differences between the mean scores at the pre-test and post-tests (see Table 23).

Table 23 *Pairwise Comparisons between the Agreement in the Verb phrase Scores for the Indirect only Group*

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-.30	.58	.61
	Post 2	.49	.66	.47
Post 1	Post 2	.79	.90	.39

Last, a one-way repeated measures ANOVA on the homophones scores showed a non-significant effect of time $F(2, 30) = 1.76$, $p = .18$, $\eta^2 = .105$. Subsequent pairwise comparisons revealed no significant differences between the mean scores at the pre-test and post-tests (see Table 24).

Table 24

Pairwise Comparisons between the Homophones Sc/ores for the Indirect only Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-.59	.56	.31
	Post 2	-1.12	.59	.07
Post 1	Post 2	-.54	.64	.42

Overall, participants in the indirect only group had their scores in the agreement in the noun phrase decrease statistically from the pre-test to post-test 2, while no significant change was observed for their scores of the agreement in the verb phrase or the homophones.

4.2.4 The indirect + ME group

In order to determine whether error categories improved over time within the indirect + ME group and whether a specific error category was better than another, three separate one-way repeated measures ANOVA and their respective post-hoc comparisons were carried out.

First, a one-way repeated measures ANOVA on the ANP scores showed a non- significant effect of time $F(2, 34) = .89$, $p = .42$, $\eta^2 = .05$. Subsequent pairwise comparisons revealed no significant differences between the mean scores at the pre-test and post-tests (see Table 25).

Table 25

Pairwise Comparisons between the ANP Scores for the Indirect + ME Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	- .07	.49	.89
	Post 2	.57	.61	.37
Post 1	Post 2	.63	.46	.18

Second, a one-way repeated measures ANOVA on the verb phrase agreement scores showed a significant effect of time $F(2, 34) = 5.20$, $p = .01$, $\eta^2 = .23$. Subsequent pairwise comparisons indicated that the increase from pre-test ($M = 96.93$, $SD = 1.26$) to post-test 1 ($M = 97.50$, $SD = 1.17$) was not statistically significant, yet, the decrease from post-test 1 to post-test 2 ($M = 96.16$, $SD = 1.98$) was statistically significant ($p = .01$). The small change from pre-test to post-test 2, however, was not statistically significant.

Table 26

Pairwise Comparisons between the AVP Scores for the Indirect + ME Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-.57	.30	.08
	Post 2	.77	.44	.10
Post 1	Post 2	1.34**	.49	.01

**The mean difference is significant at the 0.01 level

Finally, a one-way repeated measures ANOVA on the homophones scores showed a significant effect of time $F(2,34) = 7.02$, $p = .003$, $\eta^2 = .292$. Subsequent pairwise comparisons revealed statistically significant increases from pre-test ($M = 95.58$; $SD = 2.67$) to post-test 1 ($M = 96.71$, $SD = 2.48$) and from pre-test to post-test 2 ($M = 97.41$, $SD = 1.66$).

Table 27

Pairwise Comparisons between the Homophones Scores for the Indirect + ME Group

Time		Mean Difference	Std. Error	Sig.
Pre	Post 1	-1.13*	.45	.02
	Post 2	-1.83*	.57	.01
Post 1	Post 2	- .70	.46	.14

*The mean difference is significant at the .05 level.

Overall, changes in scores of the agreement in the verb phrase and of the agreement in the verb phrase did not reach statistical significance from pre-test to post-test 2; yet, the increases in homophone accuracy reached statistical significance from pre-test to post-test 1 and from pre-test to post-test 2.

To sum up, three major findings are obtained from both descriptive statistics and statistical analyses. First, regardless of the WCF group, scores in the agreement in the verb phrase (AVP) were the highest (especially at the pre-test and post-test 1). At post-test 2, they were second to scores of homophones- the differences between the two are minimal. Second, the ANP scores increased from pre-test to post-test 1 for all groups except for the direct group; these changes did not reach statistical significance. The ANP scores decreased from pre-test to post-test 2 for all groups and these decreases reached statistical significance for all groups except for the indirect + ME group. Third, only scores of homophones underwent steady increases from pre-test to post-test 1 to post-test 2 for all groups. The increase from pre-test to post-test 1 reached statistical significance for the indirect + ME group only; the increase from pre-test to post-test 2 reached statistical significance for the direct and indirect + ME groups.

4.3 The moderating effects of proficiency and language analytical ability

To explore the potential moderating effects of proficiency (determined based on learners' overall accuracy scores at the pre-test) and language analytical ability (computed based on their performance at the language analytical ability task), we conducted multiple linear regressions for each experimental group. The outcome variable consisted of an average of learners' overall scores at both post-tests; the predicting variables consisted of learners' proficiency[16] (a categorical variable) and their scores at the language analytical abilities (a continuous variable). Descriptive statistics for each group can be found in Table 28.

To verify whether the four groups were comparable in terms of their language analytical abilities, one-way ANOVA was carried out on their scores of the language analytical abilities tasks. The results indicated no significant differences between the groups F (3, 61) = .454, p = .71.

[16] Proficiency is recoded as a dummy variable (0, 1); 0 for scores below the group mean and 1 for scores above the group mean.

Table 28

Learners' individual differences for all experimental conditions

Condition	Variable		*N*	*Mean*	*SD*
Comparison	Accuracy		14	82.29	4.90
	Aptitude		14	74.49	20.24
	Proficiency	High	7		
		Low	7		
Direct	Accuracy		17	84.07	3.87
	Aptitude		17	67.44	17.31
	Proficiency	High	8		
		Low	9		
Indirect only	Accuracy		16	86.72	3.36
	Aptitude		16	72.77	15.42
	Proficiency	High	8		
		Low	8		
Indirect + ME	Accuracy		18	87.80	4.29
	Aptitude		17	70.16	19.42
	Proficiency	High	9		
		Low	9		

4.3.1 The comparison group:

The multiple linear regression to predict learners' performance at the post-test based on their proficiency and language analytical abilities was not significant $F(2, 11)= 3.14$, $p = .08$. Neither proficiency nor language analytical ability was a statistically significant predictor for learners' overall post-test scores.

4.3.2 The direct group

The multiple linear regression to predict learners' performance at the post-test based on their proficiency and language analytical abilities was not significant $F(2, 14) = .92$, $p = .42$.

Neither proficiency nor language analytical ability was a statistically significant predictor for the direct group learners' overall post-test scores.

4.3.3 The indirect only group

The multiple linear regression to predict learners' performance at the post-test based on their proficiency and language analytical abilities was not significant F (2, 13) = .95, p = .41. Neither proficiency nor language analytical ability was a statistically significant predictor for the indirect only group learners' overall post-test scores.

4.3.4 The indirect + ME group

The multiple linear regression to predict learners' performance at the post-test based on their proficiency and language analytical abilities was significant F (2, 14) = 9.80, p = .02, with an adjusted R^2 of .58. Only proficiency was found to be a significant predictor of learners' overall accuracy at the post-tests (β = .747, p = .00). Language analytical ability was not a significant predictor of learners' overall accuracy at the post-tests (β = .12, p = .5).

In general, neither proficiency nor language analytical ability seemed to be a significant predictor for overall accuracy after the experimental intervention for learners receiving no WCF, direct WCF or indirect only. However, proficiency was a significant predictor for overall accuracy of learners receiving indirect WCF + ME.

Table 29

Summary of multiple regression analyses for variables predicting learners' accuracy at the post-tests

	Comparison			Direct			Indirect only			Indirect + ME		
Variable	B	SE B	β	B	SE B	B	B	SE B	β	B	SE B	β
Proficiency	5.35	2.61	.56	2.10	1.90	.28	1.75	1.71	.27	6.09	1.41	.75
Lang aptitude	.02	.07	.07	-.05	.06	-.22	.04	.06	.19	.02	.03	.12
R^2		.36			.34			.13			.58	
F		3.15			.93			0.95			9.83	
p		.08			.42			.41			.002*	

* The significance level is .05.

Chapter 5

Summary of results and discussion

5 Summary of results and discussions

In this final chapter, results of this empirical study are summarized and discussed as they relate to each one of the three research questions. Theoretical and methodological implications are also highlighted. Last, limitations of the study and possible directions for future research are discussed.

5.1 Differential impacts of different WCF techniques on overall accuracy

The first research question examined the relative merits of three WCF techniques on learners' overall accuracy in new texts (as opposed to accuracy in revised drafts). The WCF techniques consisted of direct WCF (i.e., providing the target-like form), indirect WCF only and indirect WCF plus metalinguistic explanations.

Results indicated that while the four groups were statistically comparable at the pre-test (i.e., before the experimental intervention), the indirect WCF groups, i.e., the groups receiving indirect WCF only and indirect WCF plus ME, significantly outscored both the comparison group (i.e., the no WCF group) and the direct WCF group at the immediate post-test. No statistically significant differences were found between the two indirect WCF groups or between the comparison and the direct WCF groups either. A closer look at the four groups' performance at the immediate post-test showed that while the mean scores of the comparison group remained almost the same[17], those of the direct WCF decreased, albeit not significantly, and those of both indirect groups improved from pre-test to the immediate post-test and reached statistical

[17] The fact that the comparison group maintained the same performance can be explained by a test practice effect or can simply be attributed to the fact that they completed and revised the same writing tasks as the three WCF groups.

significance only for the group receiving metalinguistic clues. These results indicated that (a) only the indirect WCF groups have benefitted from the provided WCF and (b) indirect WCF that is combined with metalinguistic explanations is the only technique that leads to statistically significant improvement.

By the time of the delayed post-test for which learners completed a new reconstruction task that is different from the one completed at the pre-test and the immediate post-test, different result patterns have emerged. A comparison of learners' performance at the pre-test and the delayed post-test shows (a) a slight increase for the indirect + ME WCF and the direct WCF groups, (b) a minimal, almost negligible, decrease for the indirect only group and (c) a remarkable decrease that approached statistical significance (p = .06) for the comparison (no WCF) group. Although it might have been tempting to presume that, based on the results of the immediate post-test, the direct WCF was the least effective, results of the ANCOVA on the delayed post-test scores show that both direct and indirect WCF groups outperformed significantly the comparison group, which suggests that, on the long-run, providing WCF is more beneficial than withholding it. The absence of significant differences between the three WCF techniques at the delayed post-test should be interpreted with caution: while it is true that the direct WCF group outscored significantly the comparison group, it was still performing lower than the two indirect WCF groups. More important, the indirect + ME WCF group maintained its initial superiority (the mean scores at the delayed post-test for the indirect + ME, indirect only, and the direct WCF groups were respectively 87.35, 85.83 and 84.22).

Taken together, results of the immediate and delayed post-tests indicate that (1) providing WCF- regardless of its operationalization- is better than withholding it; (2) its effects seem to be more pronounced with time, and (3) indirect WCF, specifically indirect + ME, is more beneficial

than direct WCF on the short run and on the long run given that the indirect + ME WCF group is the only group that maintained its quantitative and qualitative superiority from the pre-test to the immediate and delayed post-tests.

Overall, the results pertaining to the effectiveness of WCF compared to no WCF run against Truscott's (2007) claims about the uselessness of corrective feedback and are in line with findings reported in most WCF studies, whether the feedback provided was comprehensive or highly focused. In Van Beuningen *et al.* (2012), which examined comprehensive WCF, for example, both the direct and indirect group outperformed the self-correction group (which is similar to the comparison group in the current study)[18], suggesting that WCF results in long-term sustained accuracy gains. Likewise, many focused WCF studies (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2010a; Ellis *et al.*, 2008; Sheen, 2007; Sheen *et al.*, 2009; Stefanou & Révész, 2015) have also indicated that WCF groups outscored the control group.

As far as the differential impacts of WCF techniques, the current study suggests that indirect feedback is more effective than direct feedback and that indirect feedback plus ME is more beneficial than indirect only (although the differences are not statistically significant). Those results provide evidence, albeit tentative, for the superiority of output-eliciting techniques over the input-providing ones. As argued by Lalande (1982), it is plausible that indirect WCF allows for "guided learning and problem solving" (p. 140) and pushes learners to think about their errors and to try out other alternatives (Ferris, 1999, 2002). Furthermore, adding metalinguistic explanations is more effective than indirect only. Indirect WCF, especially when it

[18] The self-correction group in van Beuningen et al. (2012) and the comparison group in this study completed the same writing tasks as the WCF groups but they revised their texts without having received WCF.

takes the form of metalinguistic explanations, is more likely to trigger the processes of hypothesis testing promoting thus "the type of reflection that [would] foster long-term acquisition" (Bitchener & Knoch, 2008, p. 415).

Results showing the absence of statistically significant differences between the two indirect WCF (i.e., indirect only and indirect + ME) are comparable to the findings reported in Bitchener and Knoch (2010b). In their investigation of the effects of circling versus circling and metalinguistic explanations on the accurate use of determiners, Bitchener and Knoch indicate no difference between the two indirect techniques. It should be noted, however, that while the current study also shows no statistically significant difference between the WCF techniques, its descriptive results indicate that the group receiving metalinguistic explanations outperformed the group whose errors were simply underlined. Furthermore, the metalinguistic group is the only indirect group that improved statistically significantly from the pre-test to the immediate post-test and that maintained its increase on the delayed post. Those results resemble the ones shown in Karim and Nassaji (2018) who explain that "underline + metalinguistic WCF performed slightly better than underline only WCF group; however, the difference was not significant." (p. 15).

Notwithstanding, these comparisons need to be interpreted with caution given that the current study and the two inquiries by Bitchener and Knoch and Karim and Nassaji differ in terms of the WCF focus, with the former adopting a highly focused feedback and the latter opting for comprehensive WCF. It is possible to argue that targeting one specific feature (i.e., the indefinite determiner and the definite determiner for anaphoric use in Bitchener & Knoch) may have had accentuated the saliency of underlining only. When different features are targeted, as is the case in the current study and in the one by Karim and Nassaji, underlining only can become

harder to interpret and metalinguistic explanations can help learners diagnose their errors and better understand their nature.

Another explanation for the lack of statistically significant differences between the two indirect techniques can be attributed to the small *n* in each group and to the nature of metalinguistic clues provided. Given that the metalinguistic explanations were provided by a research assistant (to ensure consistency across all treatment conditions), it is possible that learners have found some of these clues different from the metalanguage normally used by their teachers. Although learners were provided with WCF three times over the course of eight weeks and were given the chance to ask clarification questions during their revision sessions, it is possible that they might have needed more time to become more acquainted with the metalanguage used by the researchers. Future inquiries might thus consider providing participating learners with training to ensure that they understand and become more used to the metalinguistic clues and provide feedback over a longer period. These claims need to be empirically validated in follow-up studies.

In general and given the scarcity of studies that compared indirect WCF with and without metalinguistic feedback, no conclusive findings can be drawn about the potential superiority of metalinguistic explanations. Similar lack of straightforward results is indeed found in studies comparing different operationalizations of direct WCF. With the exception of Sheen (2007) whose study lends support to the effectiveness of direct WCF that is combined with metalinguistic explanation over direct WCF only, most other studies (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2010a; Ellis *et al.*, 2008; Sheen *et al.*, 2009; Stefanou & Révész, 2015) have shown that at least in the case of determiner use or simple past, direct WCF only is "just as

effective as the additional provision of written and oral meta-linguistic explanation" (Bitchener & Knoch, 2009a, p. 327).

It should be noted, however, that exact comparisons between these studies and the current one cannot be made because of their methodological differences. First, these studies differ in the type and operationalization of metalinguistic feedback: some provided handouts with an explanation of the rules but did not indicate where errors occurred (e.g., Shintani *et al.*, 2014; Shintani & Ellis, 2015), others underlined the errors and provided brief grammatical explanations (Benson & DeKeyser, 2018), while most studies have combined metalinguistic explanations with the provision of the correct form (Bitchener, 2008; Bitchener & Knoch, 2008, 2009a, 2010a; Ellis *et al.*, 2008; Sheen, 2007; Sheen *et al.*, 2009; Stefanou & Révész, 2015). Second, unlike the current study, most of the recent investigations have focused on a limited range of error categories (mostly specific determiner uses and to a less extent tense-related errors). The results might have been different had the feedback targeted different error categories. Findings of the current study suggest that when different language features are considered simultaneously, providing metalinguistic clues seems to be more beneficial than indirect WCF only or than direct WCF. Yet, such finding patterns need to be corroborated by future studies that examine other error categories (e.g., lexical choices and sentence structure), which brings us to the second research question of the present study.

5.2 The moderating effect of error category

The second research question investigated how the effectiveness of WCF techniques can potentially be mediated by error categories. The error categories targeted in this analysis include agreement in the noun phrase, agreement in the verb phrase and grammatical homophones — all of which are grammatical, rule-governed, categories. The second research question was thus an

attempt to move beyond the simplistic dichotomy of treatable and untreatable errors given that such a distinction is "an ad hoc one with no clear theoretical basis" that fails to recognize how "rule-governed features vary enormously in their complexity" (Shintani *et al*, 2015, p. 109).

A mixed methods ANOVA with error category (3 levels) and time (3 levels) as within subject variables and experimental condition (4 levels) as a between subject variable indicated statistically significant effects for both error category and time and a statistically significant interaction between time and error category. However, there was no interaction effect between error category and experimental condition, nor was there an interaction effect between error category, time and experimental condition. This suggests that the potential effect of the different feedback conditions did not seem to vary across error categories. Yet, a closer scrutiny at each group separately reveals that learners' accuracy scores in the three error categories was not always the same across the WCF conditions.

In all four conditions (i.e., no WCF, direct, indirect only, indirect + ME), scores in the agreement in the verb phrase underwent minimal changes from the pre-test to the immediate and delayed post-tests. These changes, however, did not reach statistical significance[19]. Nonetheless, a slightly more nuanced picture emerged for both agreement in the noun phrase (ANP) and for homophones.

First, concerning learners' mean scores in ANP, none of the four groups changed significantly from the pre-test to the immediate post-test. However, the scores at the delayed post-test plummeted statistically significantly for all groups except for the metalinguistic group.

[19] The mean difference between scores at the the pre-test and the two post-tests ranged between ± .30 and ± .70.

In other words, only the group receiving metalinguistic clues was able to avoid significant attrition in its ANP scores.

Second, concerning learners' scores in homophones, all four groups underwent steady increases from the pre-test to the immediate and delayed post-tests. More importantly, while both the direct and metalinguistic groups have reached statistically significant increases from the pre-test to the immediate post-test, only the metalinguistic group was able to carry out such a statistically significant increase to the delayed post-test, as well. Put differently, the improvement in the accurate use of homophones was durable for the metalinguistic WCF group only.

Two tentative conclusions can be drawn from these result patterns. First, the facts that the decrease in ANP was statistically significant for all groups except for the metalinguistic WCF and that the increase in homophones reached statistical significance only for the metalinguistic WCF group further underscore the superiority of combining WCF with metalinguistic clues (as opposed to indirect only or providing the target form only). Second, error categories seem to respond differently to WCF techniques. In fact, while scores in the verb phrase agreement remain almost unaltered across all four groups, those for homophones have increased remarkably- at least in the direct and metalinguistic WCF groups.

The lack of significant variation in the verb phrase agreement mean scores can be attributed to a ceiling effect as indicated by the high scores achieved in this error category for all groups (which range from 95.66 to 96. 51). It should also be noted here that only errors in agreement were included in the accuracy scores. In other words, derivational errors such as *j'ai prendu* instead of *j'ai pris* were not considered.

The remarkable decrease for agreement in the noun phrase, especially from the beginning to the end of the study, compared to the steady improvement for homophones can be explained

by three tentative hypotheses. First, at post-test 2, learners wrote longer texts and used structurally longer and richer noun phrases, which may have increased the likelihood of making errors. Second, one can argue that learners may have prioritized ideational content over formal accuracy during post-test 2; however, the differing levels of success across the three error categories implies that formal accuracy was not altogether sidelined. It is possible that learners' attentional resources were not divided equally across the three targeted errors, which may explain their better performance for homophones. Although it is true that no psychometric measures (e.g., eye tracking) were used to assess how learners directed their attentional resources as they produced their texts, one can speculate that learners may have prioritized forms that are traditionally perceived as easily fixable (Ammar *et al.*, 2016). Third, and along the same lines, the differing levels of achievement in the use of homophones and ANP can be attributed to the varying levels of complexity of the two error categories. As explained by DeKeyser (2005, 2016), complexity can be determined by at least three factors: complexity of form, complexity of meaning and complexity of the form-meaning relationship. Complexity of form refers to the number of choices available, complexity of meaning has to do with the level and degrees of abstraction (e.g., grammatical gender in French), and complexity of form-meaning mapping corresponds to the transparency between a form and its meaning (e.g., gender and number markings on nouns and adjectives are redundant in French). Drawing on DeKeyser's explanations, it is possible to argue grammatical homophones are less complex than agreement in the noun phrase.

The grammatical homophones that were targeted in the current study are mostly binary (*à* versus *a*, *la* versus *l'a*, infinitive verbs ending in *-er* versus participles ending in *-é*), which might have accentuated the corrective efficacy of direct and metalinguistic feedback. It is plausible to argue that while indirect only WCF is not enough in helping learners understand the nature of

their errors (particularly when different errors are underlined), direct WCF, which juxtaposes the correct and erroneous forms, can be easier to process especially when it targets simpler forms. Furthermore, the binary nature of the grammatical homophones targeted in this study may have also been made more salient by the metalinguistic explanations which often took the form of questions such as (*as-tu besoin d'une préposition ou de l'auxilaire avoir ici*? *Do you need a preposition or an auxiliary here*?).

Compared to homophones, agreement in the noun phrase is also characterized by the complexity of meaning (noun gender is a major learning difficulty for French L2 learners) and the complexity of its form-meaning mapping (gender and plural markings on nouns and adjectives are also encoded in determiners or can be inferred from the context). As a result, the agreement in the noun phrase can be more complex for learners because of the number of sub categories they have to consider (pre-determiner, determiner, noun, adjective) and because of the redundancy of gender and number endings. From a pedagogical perspective, errors in the noun phrase agreement are persistent learning problems for French L2 learners (Ammar *et al.*, 2016). Results of this study suggest that such errors can be less amenable to corrective feedback, especially when it takes the form of either direct WCF or underlining only. Neither technique was effective at countering the remarkable decrease in the ANP mean scores from the pre-test to the delayed post-test. In contrast, learners receiving metalinguistic feedback have at least maintained their initial scores. This might imply that even more complex language features can be affected positively by metalinguistic feedback. More studies on equally complex features are needed to confirm the potential effectiveness of metalinguistic explanations on complex features and to examine possible interactions between WCF techniques and other error categories.

Potential interactions with other mediating variables, such as learner individual differences, are to be further explored. A tentative attempt in such a direction has also been made in the current study. This is summarized next.

5.3 The moderating effects of learner individual differences

The final research question examined in this study was motivated by theoretical arguments which suggest that "individual differences may be hypothesized to exert influence on how students process feedback, the extent to which they notice gaps in their knowledge" and the extent to which they benefit from corrective feedback (Kosmos, 2012, p. 400). Against this backdrop, the third research question investigated to what extent the differential effects of three WCF techniques on L2 development were moderated by learner proficiency and language analytic ability.

The obtained results indicated a relatively moderate correlation between proficiency and learners' overall accuracy only for the group receiving metalinguistic clues, suggesting that more proficient learners have benefitted more from the metalinguistic feedback. No similiar correlations were found for either the direct WCF or the underlining groups, which might imply that the effects of the two techniques were comparable across proficiency Furthermore, no significant correlations were found between language analytic ability (LAA) and learners' overall accuracy in any of the WCF treatment conditions, which does not lend clear support for the claim that the efficacy of WCF can be mediated by learners' LAA.

These results, albeit tentative, imply that the role that LAA plays in written corrective feedback remains unclear. Such lack of clarity is further underscored by the inconclusive results obtained from the few studies that have systematically examined the mediating effect of LAA. The four inquiries that have investigated the mediating role of LAA have yielded conflicting

findings. For example, Sheen (2007) who examined the effectiveness of two direct WCF techniques (with or without metalinguistic information) on the acquisition of determiners found positive correlations between LAA and both types of WCF (direct only and direct metalinguistic feedback) and indicated that these correlations were stronger for the direct metalinguistic feedback group. On the contrary, Stefanou and Révész (2015), who also investigated the effectiveness of the same WCF techniques (direct vs direct metalinguistic WCF) on the acquisition of determiners revealed that participants with greater LAA are more likely to achieve gains in the direct feedback only group and that no such positive correlations were evident for the direct metalinguistic group. Similar results are found in Benson and DeKeyser (2018), who compared direct WCF and indirect metalinguistic WCF on the simple past and present perfect, and who indicated that learners with greater LAA are more likely to achieve gains in the direct feedback group than in the indirect metalinguistic group. Results from Benson and DeKeyser's study are in partial contrast to the ones reported in Shintani and Ellis (2015), who compared the effects of direct WCF and indirect metalinguistic WCF on determiners and past hypothetical conditional and who found that learners with higher LAA benefited more from both types of corrective feedback than learners with weaker LAA.

The inconsistent results reported in these studies and in the current examination can be explained by their methodological differences. Aside from the differences in feedback focus (comprehensive or selective) and in its form (direct, direct + ME, indirect only, indirect + ME), the five studies- including the current one- have used different tools to measure learners' language analytic abilities. Sheen (2007) used a language analysis test that required learners to induce grammatical rules of an artificial language based on a set of sentences that were accompanied by their English (learners' second language) translation. A similar instrument was used in Shintani and Ellis (2015), except that the sentences in the artificial language were

presented with their Japanese (learners' first language) translation. Benson and DeKeyser (2018) used a sub-set of Meara's (2005) LLAMA computer-based aptitude test that measures grammatical inferencing also using an artificial language (yet, without providing translations to participants). Last, both the present investigation and the one by Stefanou and Révesz (2015) used tasks that were based on adaptations of the MLAT's words-in-sentences subcomponent and were conducted in learners' first language (Greek in the former and English in the latter), yet, the instrument in the current study was longer (28 items compared to 15 items in Stefanou and Révesz). It is also plausible to argue that even though the instrument used in the present study was validated with a similar sample, it still exhibits some limitations such as the occasional overlap between functions and word classes (particularly as far as verbs are concerned). In a nutshell, and given the fact that LAA is measured differently across these studies, comparing their results cannot be straightforward.

Another important difference between the current study and those that examined the mediating effect of LAA is the number of participants. The small *n* in the current study (an average of 16 per treatment condition compared to an average of 30 participants per condition in the other studies) might have precluded any significant interactions between WCF technique and learner individual differences. In other words, different results might have emerged had the groups been larger. In general, results of the current study remain tentative at best, but they lend more support to Shintani and Ellis's claim that "further exploratory correlational studies are needed to provide a clearer picture of when and under what conditions LAA plays a role in how learners process feedback on their writing" (p. 118).

5.4 Pedagogical implications

Notwithstanding its methodological limitations, some of which are inherent to classroom-based research, this study yields some useful pedagogical implications. First, findings of its first research question, which examined the relative merits of three WCF techniques, reveal that providing WCF is more beneficial than withholding it. Thus, L2 teachers are encouraged to signal their learners' errors and to provide them with adequate feedback. With regards to specific WCF techniques, this study suggests that indirect WCF is generally more effective than direct WCF, and that combining indirect WCF with metalinguistic clues seems to be the most adequate technique. Indirect metalinguistic feedback helps leaners not only detect their errors but also better diagnose them. Consequently, L2 instructors are encouraged to prioritize indirect metalinguistic feedback. It is thus important to develop learners' metalanguage and to teach them how to mobilize it during the writing and the revising processes. Along those lines, it is incumbent on L2 teachers to develop and fine-tune their metalinguistic knowledge, as well.

Results pertaining to the second research question, which focused on the amenability of three error categories to WCF, suggest that not all errors are equally amenable to WCF. Unlike agreement in the noun phrase and in the verb phrase, grammatical homophones were more responsive to WCF (especially when the latter was combined with metalinguistic clues). It is true that the lack of positive change for errors in agreement in the noun phrase can be discouraging, however, both L2 teachers and researchers need to be reminded that L2 acquisition is a complex and nonlinear process. In fact, and as argued by Thornbury (2001), "learner's grammar restructures itself as it responds to incoming data. There seems to be periods of little change alternating with periods of a great deal of flux and variability, and even some backsliding" (cited in Menez, 2013, p. 407).

With regards to the third question, which addressed the mediating effects of learners' proficiency and language analytic ability, this study suggests that more proficient learners might benefit more from metalinguistic indirect feedback. However, given that the small number of participants precludes any conclusive results, such findings are to be interpreted with caution. Further studies with larger samples and which examine other learner individual variables are needed to shed more light on the possible interactions between WCF techniques and learners' differences.

Finally, some of the methodological choices that were implemented in this study can also inform L2 teachers' pedagogical practices. For instance, using text-reconstruction tasks have the potential of minimizing the cognitive load of generating propositional content and can, thus, help learners pay more attention to language forms. Besides providing learners with ideational and linguistic input, these writing tasks create meaningful opportunities to use specific language forms. In terms of the amount and frequency of the provided WCF, L2 teachers are encouraged to choose a mid-focused approach by targeting specific forms; such an approach is less overwhelming for both teachers and learners. Finally, and beside prioritizing indirect metalinguistic feedback, teachers need to engage their learners into revising their texts after receiving WCF. As argued by Ferris (2004) and Guénette (2007), the facilitative role of WCF is constrained if learners are not encouraged to use the feedback they were given.

5.5 Summary, limitations and future directions

The present study was designed to examine three relatively unexplored research questions. First, it investigated the differential impacts of three WCF techniques, particularly direct WCF, indirect only and indirect plus metalinguistic explanations on French L2 accuracy. It is, thus, an attempt to address some of the limitations of previous studies by adopting a mid-focused

approach, which is more ecologically valid. Unlike most recent studies, the corrective feedback in the current study did not just target one or two discrete language features. Rather, it was provided as a reaction to a range of error categories and their respective subcategories (e.g. errors in the noun phrase agreement included errors in determiners, nouns, and adjectives). To further increase the ecological validity of the provided treatment, this study did not adopt a one-shot treatment design; rather, written corrective feedback was provided on three different text-reconstruction tasks. More important, and in line with theoretical arguments and empirical evidence in favor of revision, all learners (including those who did not receive any WCF) were required to revise their texts.

Comparisons of learners' overall accuracy before and after the experimental intervention lend strong support to the usefulness of WCF, which in turn further undermines Truscott's (2004, 2007) claims. With respect to the relative merits of specific WCF techniques, this study implies that indirect WCF is more beneficial than direct WCF. While the results concerning these two techniques in the extant WCF literature are not conclusive, findings from this study corroborate theoretical arguments that are in favor of output-eliciting over input-providing techniques (Ferris, 2001; Lalande, 1982). With regards to different operationalizations of indirect WCF, this study lends more support to the superiority of metalinguistic feedback over underlining only and over direct WCF only, as well. A more comprehensive study can also include direct WCF that is combined with metalinguistic clues as well as other types of indirect WCF (for example coded WCF). Besides comparing other types of WCF, future studies might consider providing learners with training on how to process and to incorporate the provided feedback.

The second research question that motivated the current study examined the responsiveness of three error categories (two are morphological and one syntactic) to different

WCF techniques. While the results were mixed, this study has shed some light on the possible interactions between complexity and WCF effectiveness (unlike agreement in the noun phrase, homophones have benefitted from WCF). These interactions merit further attention and future examinations might need to examine other, and equally problematic, error categories such as lexical errors and sentence structure problems.

The third research question is a tentative exploration of the mediating effect of some learner factors (mostly proficiency and language analytic ability). The results, however, did not clearly establish whether and to what extent the efficacy of WCF techniques is mediated by these factors. The lack of clear correlations between WCF techniques and the two learner individual differences examined in this study can be partially attributed to how they were measured. Future research should include more rigorous measures of proficiency that consider not only accuracy but also complexity and fluency. Furthermore, and to obtain more comparable results, future inquiries might adopt more uniform measurements of language analytic ability. More important, very little is known about how WCF interacts with other individual factors such as motivation and learner's perceived self-efficacy.

Finally, and although this study has sought to address different methodological issues, it has its own share of limitations, as well. The biggest one is the small number of participants that remained in each one of the experimental conditions. Adopting a conservative approach by only including learners that completed all three tests and that participated in at least two out of the three writing and revision sessions has resulted into smaller groups, which can account for the absence of more significant results.

A second limitation of this study might be attributed to the relatively large number of writing tasks participants were asked to complete from the beginning to the end of the study.

Besides completing the language analytic ability tasks and a questionnaire on their perceptions on French and on writing, participants had to write six text reconstruction tasks and to revise three of those six tasks. Given that learners were not used to write this extensively, it is possible that they felt either bored or overwhelmed by the end of the study.

Third, and with respect to the writing task, text-reconstruction tasks were chosen to control for propositional content and to lighten the cognitive processes involved in generating ideas and organizing texts. These tasks, however, were not piloted with similar learners before the experimental intervention — an aspect that can be addressed in future inquiries so as learners' affective and cognitive engagement in the different writing tasks would be taken into consideration. While many participants seemed to have appreciated such writing tasks, few of them- especially the more proficient ones- have expressed their desire for more creative writing assignments. As such, future inquires might include less contrived writing tasks.

Fourth, some potentially confounding variables were not controlled for. For example, because teachers volunteered to participate in this study, their previous experience in teaching French or teaching writing was not accounted for, neither were their teaching practices. Notwithstanding, different methodological measures were undertaken to attenuate the possible variations between participating teachers. First, the WCF in this study was provided by the author of this dissertation and another research assistant in order to control for the potential influence of teachers' grammatical knowledge and of their usual WCF practices. To ensure consistency across the experimental conditions, all WCF instances in the three writing tasks were double-checked by the author of this study and the research assistant. Furthermore, the author of this dissertation and three other research assistants attended and monitored the revision sessions to ensure that learners revised under similar conditions.

Finally, this study does not tease apart the effects of practice only (as is the case with Van Beuningen *et al.*, 2012). Future inquires might thus include a real control group that does not participate in the writing tasks of the experimental intervention. Last, this study has only focused on learners' performance in new texts (in the immediate and delayed post-tests). It would have been insightful to examine how learners processed and incorporated the provided WCF in their revised texts, and whether there are any correlations between their success in the revision tasks and their performance in the new writing tasks.

References

Alamargot, D., & Chanquoy, L. (2001). *Through the Models of Writing: With Commentaries by Ronald T*. New York: Springer Science & Business Media.

Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research*, *12*(2), 183-210.

Ammar, A., Daigle, D., & Lefrancois, P. (2016). *La rétroaction corrective écrite dans l'enseignement secondaire au Québec: Effets du contexte d'apprentissage, du type d'erreurs et de la compétence des apprenants à l'écrit.* Québec : Fonds de Recherche Société et Culture.

Ammar, A., & Hassan, R. M. (2018). Talking it through: Collaborative dialogue and second language learning. *Language learning*, *68*(1), 46-82.

Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28(04), 543-574.

Amrhein, H. R., & Nassaji, H. (2010). Written corrective feedback: What do students and teachers think is right and why?. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, *13*(2), 95-127.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, *48*(1), 35-48.

Anderson, J. R. (2015). *Cognitive psychology and its implications.* New York: Worth Publishers.

Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method?. *Journal of second language writing*, *9*(3), 227-257.

Becker, A. (2006). A review of writing model research based on cognitive processes. In A.S. Horning, & A. Becker (Eds) *Revision: History, theory, and practice* (pp. 25-49). Parlor Press, LLC.

Benson, S., & DeKeyser, R. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research*, 00(0), 1-25.

Bitchener, J. (2012). A reflection on 'the language learning potential'of written CF. *Journal of Second Language Writing*, *21*(4), 348-363.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 409–431.

Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal, 63*, 204–211.

Bitchener, J., & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*, 193 214.

Bitchener, J., & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, *19*, 207–217.

Bitchener, J., & Storch, N. (2016). *Written Corrective Feedback for L2 Development* (Vol. 96). Multilingual Matters.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*, 227–258.

Boivin, M.C. & Pinsonneault, R. (2016). Un modèle théorique d'articulation de l'enseignement de la grammaire et de l'écriture. In M. Depeursinge, S. Florey, N. Cordonier, S. Aeby Daghé, J.-F. de Pietro (Eds.), Actes du colloque « L'enseignement du français à l'ère du numérique », 29, 30, 31 août 2013, (pp. 108-120). Lausanne : Haute école pédagogique du canton de Vaud.

Boivin, M.C. & Pinsonneault, R. (2014). *Un modèle didactique d'articulation de la grammaire et de l'écriture pour favoriser le transfert des connaissances grammaticales en situation de production écrite chez les élèves du secondaire.* Québec : Fonds de Recherche Société et Culture.

Bruton, A. (2009). Designing research into the effects of grammar correction in L2 writing: Not so straightforward. *Journal of Second Language Writing*,*18*(2), 136-140.

Bussière, P., Cartwright, F., Crocker, R., Ma, X., Oderkirk, J., & Zhang, Y. (2001). *Measuring up: The performance of Canada's youth in reading, mathematics, and science, OECD PISA Study – first results for Canadians aged 15*. Ottawa, ON: Human Resources Development Canada, Council of Ministers of Education, Canada, and Statistics Canada. http://www.pisa.gc.ca/pisa/81-590-xpe.pdf

Byrnes, H., & Manchón, R. (Eds.) (2014). *Task-based Language Learning: Insightsfrom and for L2 Writing.* Amsterdam: John Benjamins.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of second language writing*, *12*(3), 267-296.

Chartrand, S. (2013). Quelles finalités pour l'enseignement grammatical à l'école? Une analyse des points de vue des didacticiens du français depuis 25 ans. *Formation et Profession, 20*, 3. [En ligne]: http://crifpe.ca/formationetprofession.

Chartrand, S., & Lord, M. (2013). L'enseignement de la grammaire et de l'écriture au secondaire québécois. Principaux résultats d'une recherche descriptive. *Schweizerische Zeitschrift für Bildungswissenschaften*, *35*(3), 515-539.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language*, *35*(1), 26-58.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Cumming, A. (1990). Metalinguistic and Ideational Thinking in Second Language Composing. *Written Communication*, 7(4), 482–511.

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language learning*, *46*(3), 529-555.

DeKeyser, R. (1996). Exploring automatization processes. *Tesol Quarterly*, *30*(2), 349-357.

DeKeyser, R. (2005). What makes learning second-language grammar difficult? A review of issues. *Language learning*, *55*(S1), 1-25.

DeKeyser, R. (2007a). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories*

in second language acquisition (pp. 97–113). Mahwah, NJ: Erlbaum.

DeKeyser, R. (2007b). Situating the concept of practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1–18). New York: Cambridge University Press.

DeKeyser, R. (2016). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition*, *38*(2), 353-363.

DeKeyser, R., & Criado, R. (2013). Automatization, skill acquisition, and practice in second language acquisition. In C.A. Chappelle (Ed.) *The Encyclopedia of Applied Linguistics* (pp. 323-331). Maiden, MA: Blackwell.

DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46(4), 613-642.

De Larios, J. R., Murphy, L., & Marín, J. (2002). A critical examination of L2 writing process research. In *New directions for research in L2 writing* (pp. 11-47). Amsterdam: Springer.

Doughty, C., and Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197-262). Cambridge: Cambridge University Press.

Ellis, R. (1994). *The study of second language acquisition*. Oxford University.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT journal*,*63*(2), 97-107.

Ellis, R. (2010). Epilogue. *Studies in Second Language Acquisition*, *32*(02), 335-349.

Ellis, R. (2012). *Language teaching research and language pedagogy*. Oxford: Wiley Blackwell.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, *36*(3), 353-371.

Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. *Second language writing: Research insights for the classroom*, *9*, 178-190.

Fazio, L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority language students. *Journal of Second Language Writing, 10* (4), 235–249.

Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, *31*(2), 315-339.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of second language writing*, *8*(1), 1-11.

Ferris, D. (2003). *Response to student writing: Implications for second language students.* Mahwah, NJ: Erlbaum.

Ferris, D. (2004). The ‘‘grammar correction’’ debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime . . .*?). Journal of Second Language Writing*, 13(1), 49–62.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short-and long-term effects of written error correction. In K., Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing: Contexts and Issues* (pp. 81- 104). Cambridge: Cambridge University Press.

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition*, *32*(02), 181-201.

Ferris, D. & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process, and practice* (2nd ed.). Mahwah, NJ: Erlbaum.

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.

Fillmore, L. W., & Snow, C. E. (2000). What teachers need to know about language. *Center for Applied Linguistics, Washington, DC.*

Fitzgerald, J. (1987). Research on revision in writing. *Review of educational research, 57(4), 481- 506.*

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing.*College composition and communication, 32(4), 365-387.*

Flower, L., Hayes, J. R., Carey, L., Shriver, K., & Stratman, J. (1986). Detection, diagnosis and the strategies of revision. *College Composition and Communication* , 37 (1), 1655.

Fortune, A. (2005). Learners' use of metalanguage in collaborative form-focused L2 output tasks. *Language Awareness, 14(1), 21-38.*

Furneaux, C., Paran, A., & Fairfax, B. (2007). Teacher stance as reflected in feedback on student writing: An empirical study of secondary school teachers in five countries. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(1), 69-94.

Galbraith, D. (2009). Cognitive models of writing. *German as a foreign language, (2-3), 7-22.*

Goldenberg, C., & Coleman, R. (2010). *Promoting academic achievement among English learners: A guide to the research.* Thousand Oaks, CA: Corwin.

Grégoire P. (2012). *L'impact du traitement de texte sur l'écriture d'élèves québécois du secondaire.* Communication présentée au colloque 2012 de l'Association francophone pour le savoir (ACFAS), Montréal, Québec.

Guénette, D. (2007). Is feedback pedagogically correct?: Research design issues in studies of feedback on writing. *Journal of second language writing*,*16*(1), 40-53.

Guénette, D., & Lyster, R. (2013). Written corrective feedback and its challenges for pre-service ESL teachers. *Canadian modern language review*,*69*(2), 129-153.

Guthrie, J., Schafer, W., & Hutchinson, S. (1991). Relations of document literacy and prose literacy to occupational and societal characteristics of young black and white adults. *Reading Research Quarterly*, 26, 30–48.

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing*, *11*(4), 329-350.

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J. F., & Carey, L. (1987). 5 Cognitive processes in revision. *Advances in Applied Psycholinguistics: Volume 2, Reading, Writing, and Language Learning*. New York: Cambridge University Press.

Housen, A., & Pierrard, M. (2005). *Investigations in instructed second language acquisition*, Vol. 25. Berlin: Walter de Gruyter.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 1-20.

Hyland, K., & Hyland, F. (2006). *Feedback in second language writing: Contexts and issues*. Cambridge university press.

Izumi, S. (2013). Noticing and L2 development: Theoretical, empirical and pedagogical issues. In J. Bergsleithner, S. Frota, & J. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 25-38). Honolulu: University of Hawai'i, National Foreign Language Resource Center.

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A Meta-Analysis. *The Modern Language Journal*, *99*(1), 1-18.

Karim, K., & Nassaji, H. (2018). The revision and transfer effects of direct and indirect comprehensive corrective feedback on ESL students' writing. *Language Teaching Research*, 1-21.

Kartchava, E. (2013). The place of corrective feedback within the major SLA theories. *US-China Foreign Language*, *11*(2), 136- 151.

Kennedy, L. (2013, December 10). Literacy doesn't mean what you think. *The Huffington Post*. http://www.huffingtonpost.ca/lindsay-l-kennedy/literacy-skills_b_4413500.html

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, *75*(3), 305-313.

Kline, R. B. (2009). *Becoming a behavioral science researcher: A guide to producing research that matters*. New York: Guilford Press.

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390-403.

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

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

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Lalande, J. F. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66*, 140–149.

Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing*, *13*(4), 285-312.

Lee, I. (2007). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing*, *17*(2), 69-85.

Leeman, J. (2007). Feedback in L2 learning: Responding to errors during practice. *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*, 111-137.

Le Ferrec, L. (2008). Littératie, relations à la culture scolaire et didactique de la lecture- écriture en français langue seconde. In J.-L. Chiss (Ed.), *Immigration, École et didactique du français* (p. 101-145). Paris: Les Éditions Didier.

Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals*, 24, 203–218.

Leki, I., Cumming, A. H., & Silva, T. (2008). *A Synthesis of research on second language writing in English.* New York: Taylor & Francis.

Lévesque, C. (2015, September 8). Illiteracy in Quebec becoming a massive problem in the province. *Huffington Post*. Retrieved from http://www.huffingtonpost.ca/2015/09/08/illiteracy-in-quebec_n_8100450.html

Li, S. (2013). The interface between feedback type, L2 proficiency, and the nature of the linguistic target. *Language Teaching Research, 18*(3), 373-396.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In L. Ortega (Ed.), Understanding second language process (pp.27-44). New York: Routledge.

Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching. *Studies in second language acquisition*, *12*(04), 429-448.

Lightbown, P. M., & Spada, N. (1994). An innovative program for primary ESL students in Quebec. *TESOL quarterly*, *28*(3), 563-579.

Lightbown, P. M., & Spada, N. (1999). *How languages are learned*. Oxford: Oxford University Press.

Liu, Q. & Brown, D. (2015). A methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *The Journal of Second Language Writing, 30*, 66-81.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, and C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). Amsterdam: John Benjamins.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of language acquisition: Vol. 2. Second language acquisition* (pp. 413-468). New York: Academic Press.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in second language acquisition*, *26*(03), 399-432.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach.* Amsterdam: John Benjamins.

Lyster, R. & K. Saito (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition 32*(2), 265–302.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language teaching*, *46*(1), 1-40.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied linguistics*, *27*(3), 405-430.

Manchòn, R. M. (2011). Writing to learn the language: Issues in theory and research. In R. M. Manchòn (Ed.), *Learning-to-write and writing-to learn in an additional language* (pp. 61–82). Amsterdam: John Benjamins.

Manchòn, R.M., & Roca de Larios, J. (2007). Writing-to-learn in instructed language learning

contexts. In E. Alcòn Soler & M.P. Safont Jordà (Eds.), *Intercultural language use and language learning* (pp. 101-121). Dordrecht: Springer.

Manchòn, R. M., Roca de Larios, J., & Murphy, L. (2009). The temporal dimension and problem-solving nature of foreign language composing processes: Implications for theory. In R.M. Manchòn (Ed.) *Writing in foreign language contexts: Learning, teaching, and research* (pp. 102-129). Bristol: Multilingual Matters.

McLaughlin, B. (1990). "Conscious" versus "unconscious" learning. *TESOL Quarterly*, 24, 617-634.

Menezes, V. (2013). Second language acquisition: Reconciling theories. *Open Journal of Applied Sciences*, *3*(07), 404-412.

Meyers, L. S., Gamst, G. C., & Guarino, A. J. (2013). *Performing data analysis using IBM SPSS*. New Jersey: John Wiley & Sons.

Miller, G. (1988). The challenge of universal literacy. *Science*, 241, 1293–1299.

Ministère de l'Éducation et de l'Enseignement Supérieur (2004). *Québec education program: Secondary school education, cycle one*. Québec: Gouvernement du Québec.

Ministère de l'Éducation et de l'Enseignement Supérieur (2010). *Progression des apprentissages au secondaire Français, langue seconde (programme enrichi)*. Québec: Gouvernement du Québec.

Ministère de l'Éducation et de l'Enseignement Supérieur (2017). *Référentiel d'intervention en écriture.* Québec : Gouvernement du Québec.

Mitchell, R., & Myles, F., (2002). *Second language learning theories*. New York: Oxford University.

Murnane, R., & Levy, F. (1996). *Teaching the new basic skills: Principles for helping children to thrive in a changing economy*. New York: Free Press.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder.

Ortega, L. (2012). Epilogue: Exploring L2 writing–SLA interfaces. *Journal of Second Language Writing*, *21*(4), 404-415.

Pica, T. (1994). Research on negotiation: What does it reveal about second language learning conditions, processes, and outcomes? *Language Learning, 44*(3), 493–527.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, *21*(4), 375-389.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*, 83–95.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*(01), 27-67.

Robinson, P. (Ed.). (2002a). *Individual differences and instructed language learning* (Vol. 2). Amsterdam: John Benjamins.

Robinson, P. (2002b). Introduction: Researching individual differences and instructed learning. In P. Robinson (Ed.) *Individual differences and instructed language learning* (pp. 1-12). Amsterdam: John Benjamins.

Rutherford, W., & Sharwood Smith, M. (1985). Consciousness-raising and universal grammar. *Applied linguistics*, *6*(3), 274-282.

Sasayama, S. (2015). *Validating the assumed relationship between task design, cognitive complexity, and second language task performance* (Unpublished doctoral dissertation). Georgetown University, Washington, USA.

Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. *Advances in applied psycholinguistics*,*2*, 142-175.

Schachter, J. (1991). Corrective feedback in historical perspective. Second Language Research, 7, 89-102.

Schmidt, R. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*(2), 129-158.

Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 21-42). New York: Oxford University Press.

Schmidt, R. (1994a) Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review, 11*, 11–26.

Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–210). San Diego, CA: Academic Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning*, *9*, 1-63.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–33). New York: Cambridge University Press.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: a case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: conversation in second language acquisition* (pp. 237–326). MA: Newbury House.

Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *The Modern Language Journal*, *85*(2), 244-258.

Schwartz, B. **(**1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in second language acquisition 15(*2), 147– 164.

Schwartz, B. & Gubala-Ryzak, M. (1992). Learnability and grammar reorganization in L2A: against negative evidence causing the unlearning of verb movement. *Second language research 8(*1), 1–38.

Segalowitz, N., & Lightbown, P. M. (1999). Psycholinguistic approaches to SLA. *Annual Review of Applied Linguistics*, *19*, 43-63.

Semke, H. (1984). The effects of the red pen. *Foreign Language Annals*, 17, 195–202.

Sharwood Smith, M. (1981). Consciousness-Raising and the second language learner. *Applied linguistics*, *2*(2), 159-168.

Sharwood Smith, M. (1986). Comprehension versus acquisition: Two ways of processing input. *Applied linguistics*, *7*(3), 239-256.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of determiners. *TESOL Quarterly*, *41*(2), 255-283.

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, *37*(4), 556- 569.

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*(1), 103-131.

Shintani, N., & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing?. *System*, *49*, 110-119.

Silva, T., & Leki, I. (2004). Family matters: The influence of applied linguistics and composition studies on second language writing studies—Past, present, and future. *The Modern Language Journal*, *88*(1), 1-13.

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

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.) *Individual differences and instructed language learning* (pp. 69-74). Amsterdam: John Benjamins.

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

Smith, M. S. (1981). Consciousness-Raising and the second language learner. *Applied linguistics*, *2*(2), 159-168.

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching* 30, 73−87.

Spada, N., & Lightbown, P. M. (1989). Intensive ESL programmes in Quebec primary schools. *TESL Canada Journal*, *7*(1), 11-32.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A Meta-Analysis. *Language learning*, *60*(2), 263-308.

Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal*, *99*(2), 263-282.

Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, *10*(2), 29-46.

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

Swain, M. (1991). French immersion and its offshoots: Getting two for one. In B. Freed (Ed.), Foreign language acquisition: Research and the classroom (pp. 91-103).

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, 51, 158-164.

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

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). Cambridge: Cambridge University Press.

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

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, 82, 320-337.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th Ed.). Boston, MA: Pearson.

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition*, 33(3), 339-372.

Tremblay, A., & Garrison, M. D. (2010). Cloze tests: A tool for proficiency assessment in research on L2 French. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected*

proceedings of the Second Language Research Forum 2008 (pp. 73-88). Somerville, MA: Cascadilla Press

Trofimovich, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 171-195). New York: Oxford University Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, *46*(2), 327-369.

Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*,*8*(2), 111-122.

Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of second language writing*, *13*(4), 337-343.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of second language Writing*, *16*(4), 255-272.

Truscott, J., & Hsu, A. Y. P. (2008). Error correction, revision, and learning. *Journal of second language writing*, *17*(4), 292-305.

Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing, and translation from L1 into L2. *Journal of Second Language Writing*,*5* (3), 271-294.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL International Journal of Applied Linguistics*, 156, 279–296.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1-41.

VanPatten, B., & Williams, J. (2007). Input processing in adult SLA. In B. VanPatten, & J. Williams (Eds), *Theories in second language acquisition: An introduction* (pp.115-135). New York: Routledge.

White, L. (1991). Adverb placement in second language acquisition: some effects of

positive and negative evidence in the classroom. *Second language research 7(*2), 133– 161.

White, L. (1996). Universal grammar and second language acquisition: Current trends and new directions. In W. C. Ritchie, & T. K. Bhatia (Eds), *Handbook of language acquisition: Vol.2. Second language acquisition* (pp. 85-120). New York: Academic Press.

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Williams, J. (2012). The potential role (s) of writing in second language development. *Journal of Second Language Writing*, *21*(4), 321-331.

Xu, C. (2009). Overgeneralization from a narrow focus: A response to Ellis *et al*.(2008) and Bitchener (2008). *Journal of Second Language Writing*, *18*(4), 270-275.

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics, 34*(3), 344–368.

Appendix A
Text-reconstruction task for the pre-test and the immediate post-test

Consigne pour le rappel du texte « L'art comme issue de secours »

- Vous allez écouter un texte qui parle de deux femmes Françoise Collin et Saundra Baly. Ces deux femmes se sont rencontrées dans un organisme qui s'appelle '' La Rue Des Femmes''.
- Vous allez ensuite écrire le rappel de ce texte (un minimum de 250 mots). Pour ce faire, vous allez suivre les étapes suivantes :
  1. Écoutez le texte une première fois pour comprendre le thème général;
  2. Écoutez le texte une deuxième fois pour prendre notes des éléments qui vous semblent nécessaires pour rédiger votre rappel;
  3. Écoutez le texte une dernière fois pour vous assurer d'avoir tous les éléments nécessaires;
  4. Rédigez votre rappel en utilisant au moins 40 mots de la liste suivante.

- Patience, femme, barrière, jeu, découverte, secours, efficacité, souffrance, peinture, émotion, artiste, organisme, œuvre, guérison, énergie **(10 mots au minimum)**
- Démuni, compulsif, quotidien, grand, talentueux, personnel, lumineux, conscient, emprisonné, fondamental, déconnecté, agressif, dangereux, bon, négatif, communicatif **(10 mots au minimum)**
- Fréquenter, fouler, s'éloigner, s'initier, débuter, passer, s'installer, mettre, se retrouver, transformer, changer, devenir, exprimer, se parler, écouter **(12 mots au minimum)**
- Leur, leurs, a, à, on, ont, ses, ces, c'est, s'est, là, la, l'a **(5 mots au minimum)**
- Quelques, plusieurs, tout, chaque, aucun, chacun **(3 mots au minimum)**

Appendix B
Text-reconstruction task for the delayed post-test

Consigne pour le rappel du texte « *Les réfugiés syriens : survivre aux horreurs de la guerre et chercher un meilleur avenir* »

- Vous allez écouter **trois fois** un texte qui parle de la situation alarmante des réfugiés syriens et des efforts fournis pour les aider.
- Vous avez ensuite 45 minutes pour écrire le rappel de ce texte (un minimum de 250 mots). **Laissez une interligne double lors de votre rédaction.**
- Pour ce faire, vous allez suivre les étapes suivantes :
    1. Écoutez le texte une première fois pour comprendre le thème général;
    2. Écoutez le texte une deuxième fois pour prendre notes des éléments qui vous semblent nécessaires pour rédiger votre rappel.
    3. Écoutez le texte une dernière fois pour vous assurer d'avoir tous les éléments nécessaires;
    4. Rédigez votre rappel en utilisant au moins **40 mots** de la liste suivante. Attention : les mots sont donnés dans leurs formes brutes. Assurez-vous de les mettre dans la forme appropriée dépendamment de vos phrases.

- rêve, inquiétude, difficulté, problème, environnement, perception, effort, engagement, mission, affection, volontaire, situation, expérience, réfugié, frontière, humiliation, territoire, aventure, désespoir **(10 mots au minimum)**
- alarmant, public, ambitieux, pessimiste, défavorisé, divers, violent, incessant, différent, humanitaire, incapable, innombrable, dangereux, impuissant, anxieux, fatal, sécurisant, sûr **(10 mots au minimum)**
- réussir, s'aggraver, se détériorer, se déplacer, s'impliquer, dénoncer, favoriser, fournir, perdre, s'adapter, travailler, retourner, s'éloigner, parcourir, traverser, subir, survivre, risquer, échapper, soutenir **(12 mots au minimum)**
- Leur, leurs, a, à, on, ont, ses, ces, c'est, s'est, là, la, l'a, ce, se, et, est **(5 mots au minimum)**
- Quelques, plusieurs, tout, chaque, aucun, chacun **(3 mots au minimum)**

Appendix C
Text-reconstruction tasks for the experimental intervention

Task 1

Consigne pour le rappel du texte « Les jeunes québécois : entre rêves et inquiétudes »

- Vous allez écouter **trois fois** un texte qui parle d'un sondage qui s'adresse aux jeunes et qui révèle leurs rêves et inquiétudes.
- Vous avez ensuite 45 minutes pour écrire le rappel de ce texte (un minimum de 250 mots). **Laissez une interligne double lors de votre rédaction.**
- Pour ce faire, vous allez suivre les étapes suivantes :
  1. Écoutez le texte une première fois pour comprendre le thème général;
  2. Écoutez le texte une deuxième fois pour prendre notes des éléments qui vous semblent nécessaires pour rédiger votre rappel.
  3. Écoutez le texte une dernière fois pour vous assurer d'avoir tous les éléments nécessaires;
  4. Rédigez votre rappel en utilisant au moins **40 mots** de la liste suivante. Attention : les mots sont donnés dans leurs formes brutes. Assurez-vous de les mettre dans la forme appropriée dépendamment de vos phrases.

- rêve, inquiétude, valeur, soutien, décrochage, difficulté, problème, environnement, changement, pollution, disparition, perception, rendement, effort, engagement **(10 mots au minimum)**
- jeune, inscrit, important, familial, conventionnel, alarmant, public, capable, ambitieux, constant, mauvais, climatique, pessimiste, négatif, satisfaisant, social **(10 mots au minimum)**
- chercher, valoriser, réussir, chuter, s'aggraver, se détériorer, se déplacer, recycler, aider, penser, s'impliquer, assumer, dénoncer, estimer, favoriser, fournir **(12 mots au minimum)**
- Leur, leurs, a, à, on, ont, ses, ces, c'est, s'est, là, la, l'a, ce, se, et, est **(5 mots au minimum)**
- Quelques, plusieurs, tout, chaque, aucun, chacun **(3 mots au minimum)**

Task 2

Consigne pour le rappel du texte '' Mission humanitaire : témoignage de Marie Fouré ''

- Vous allez écouter **trois fois** un texte qui parle d'un sondage qui s'adresse aux jeunes et qui révèle leurs rêves et inquiétudes.
- Vous avez ensuite 45 minutes pour écrire le rappel de ce texte (un minimum de 250 mots). **Laissez une interligne double lors de votre rédaction.**
- Pour ce faire, vous allez suivre les étapes suivantes :
  5. Écoutez le texte une première fois pour comprendre le thème général;
  6. Écoutez le texte une deuxième fois pour prendre notes des éléments qui vous semblent nécessaires pour rédiger votre rappel.
  7. Écoutez le texte une dernière fois pour vous assurer d'avoir tous les éléments nécessaires;
  8. Rédigez votre rappel en utilisant au moins **40 mots** de la liste suivante. Attention : les mots sont donnés dans leurs formes brutes. Assurez-vous de les mettre dans la forme appropriée dépendamment de vos phrases.

- mission, volontariat, orphelinat, problème, comportement, affection, volontaire, moniteur, directive, situation, respect, expérience **(10 mots au minimum)**
- défavorisé, divers, difficile, violent, verbal, incessant, impressionnant, différent, intense, réduit, social, psychique, laborieux, bénéfique, reconnaissant **(10 mots au minimum)**
- se rendre, perdre, abuser, s'adapter, se dérouler, travailler, réaliser, suivre, retourner, connaître, montrer, devoir, disparaître, se montrer, participer **(12 mots au minimum)**
- Leur, leurs, a, à, on, ont, ses, ces, c'est, s'est, là, la, l'a, ce, se, et, est **(5 mots au minimum)**
- Quelques, plusieurs, tout, chaque, aucun, chacun **(3 mots au minimum)**

Task 3

Consigne pour le rappel du texte « L'histoire de survie d'un réfugié syrien »

- Vous allez écouter **trois fois** un texte qui parle d'un réfugié syrien, Rachid, qui a fui la guerre en Syrie et qui a entamé un voyage dangereux vers l'Europe.
- Vous avez ensuite 45 minutes pour écrire le rappel de ce texte (un minimum de 250 mots). **Laissez une interligne double lors de votre rédaction.**
- Pour ce faire, vous allez suivre les étapes suivantes :
  9. Écoutez le texte une première fois pour comprendre le thème général;
  10. Écoutez le texte une deuxième fois pour prendre notes des éléments qui vous semblent nécessaires pour rédiger votre rappel.
  11. Écoutez le texte une dernière fois pour vous assurer d'avoir tous les éléments nécessaires;
  12. Rédigez votre rappel en utilisant au moins **40 mots** de la liste suivante. Attention : les mots sont donnés dans leurs formes brutes. Assurez-vous de les mettre dans la forme appropriée dépendamment de vos phrases.

- aide, réfugié, frontière, humiliation, détention, territoire, aventure, obscurité, bateau, voyage, soutien, désespoir, avenir, solution, paix, éducation **(10 mots au minimum)**
- humanitaire, incapable, innombrable, dangereux, total, différent, nombreux, impuissant, entassé, vieux, anxieux, chanceux, fatal, sécurisant, sûr **(10 mots au minimum)**
- s'éloigner, parcourir, traverser, subir, embarquer, conduire, menacer, s'opposer, survivre, risquer, échapper, se déplacer, soutenir, se débrouiller, soulager **(12 mots au minimum)**
- Leur, leurs, a, à, on, ont, ses, ces, c'est, s'est, là, la, l'a, ce, se, et, est **(5 mots au minimum)**
- Quelques, plusieurs, tout, chaque, aucun, chacun **(3 mots au minimum)**

Appendix D
Measurement of language analytical abilities (revised version)

Part 1: Word class

Task: In the following sentences, identify which word belongs to the same family/is of the same nature as the word that is **BOLDED AND CAPITALIZED** in the numbered key sentence. The word classes examined in this test are *adjectives, adverbs, coordinators, determiners, nouns, prepositions, pronouns and verbs*.

0. I lost **MY** wallet.

- The teacher is absent today.

A B C D

Answer: **A**. Both words are determiners.

1. It's possible for almost anyone to **RECORD** a great song at home or in a professional studio.
- I've never been able to get anything done around city hall without bribing someone.
A B C D
Answer: …….

2. I bought her all kinds of things such as stuffed animals, clothing, **AND** jewelry.
- My dog enjoys being bathed but hates getting his nails trimmed.
A B C D
Answer: ……

3. Has **ANYONE** seen my handbag? I cannot find it.
- In democratic countries like Canada, none is above the law.
A B C D
Answer: …….

4. She has been waiting for her results **IMPATIENTLY**.
- From now on, I will try to never disappoint you.
A B C D
Answer: ……

5. Great Expectations is my **FAVORITE** book.
- The angry politician was talking about the necessity of adopting stricter measures.
A B C D
Answer: ……..

6. They decided to go **TO** France for their honeymoon.
- Dwarf couple marries in a fairy-tale ceremony after they met while performing in Snow White.
A B C D
Answer: ……

7. Among my personal highlights is the **MEMORY** of travelling to Kenya.
- This short documentary emphasizes the importance of not texting while driving.
A B C D

Answer: ……

8. Parents are usually **PROUD** of their children's achievements.
- His fans were so disappointed that he did not win the race.
A B C D
Answer: …….

9. Justin Bieber will have a concert **IN** Toronto.
- She left him some money on the Table.
A B C D
Answer:…….

10. At that point, they decided to call the **EXPERIMENT** to a halt.
-The call you received must have annoyed you.
A B C D

Answer: …….

11. **IT** is the most beautiful rose I have ever seen.
- Can I go with you to the movies?
A B C D
Answer: ……

12. The boys wish they could **GO** to Hawaii.
- The teacher's wish list may include school supplies and gift cards for her favorite places.
A B C D
Answer: ……

13. I enjoyed **OUR** dinner last night.
- My heart nearly burst with joy when I finally brought them back from the hospital.
A B C D
Answer: ……

14. The school magazine comes out **WEEKLY**.
- Unfortunately, your submission to this journal was not accepted.
A B C D

Answer: ……

15. My cat loves having his head scratched **BUT** hates getting his claws trimmed.
- She studied hard for her final exam, yet she couldn't get the grade she wanted.
  A B C D
Answer: ……

16. **THE** Montreal Canadiens are a professional ice hockey team based in Montreal.
- The CBC broadcasts a series of public service announcements.
  A B C D
Answer:…….

****

Part 2: Functions

Task: In the following sentences, identify which word or group of words has the same function as the word or group of words that is **BOLDED AND CAPITALIZED** in the numbered key sentence. The functions examined in this test are *adverbials, direct objects, indirect objects, subjects, subject complements* and *verbs*.

Example
0. **THIS TEST** might be fun.

- She never arrives late.

A B C D

Answer: **A**. Both are subjects.

1. Their daughter **IS COMPLAINING** about the harshness of her first-grade teacher.
Montreal prohibited cigarettes in all indoor public spaces.
A B C D
Answer: ……

2. Because of the Cuban government, the book remained **UNPUBLISHED**.
For different reasons, this seems a good idea.
A B C D
Answer: ……..

3. One should not drive **FAST** on snowy days.
Galileo worked hard to convince his generation that the earth revolves around the sun.
A B C D

Answer: ……..

4. We wanted to go out, but we were **TOO TIRED**.
The Greek athletes felt confident in their victory because of their extensive training,.
A B C D
Answer: ……..

5. I bought a new laptop **FOR MY BROTHER.**
Mary should never have borrowed money from her co-workers.
A B C D
Answer: ……..

6. **MARY** is happy.
Sophia can tell that her sister must have had a bad day from the look on her face
A B C D
Answer: ……..

7. He sent an invitation **TO ALL HIS CLASSMATES**.
Big corporations often give a lot of money to political candidates.
A B C D
Answer:……..

8. German politicians **HAVE RESPONDED** to outbreaks of racial violence with demands for tighter immigration controls.
Despite their economic difference, all the other eastern states share the same pattern of
A B C D
high anti-foreign violence.

Answer: ……..

9. She ordered **HER CHILDREN** to stay quiet.
They have completed the renovations in such a short time!
A B C D

Answer: ……

10. After spending so many years in Hawaii, **SHE** decided to go to Toronto.
The talented artist has traveled to nearly a hundred countries for millions of miles.
A B C D

Answer: …….

11. The officer gave me **A TICKET**!
When she went away to college, Emily wrote to her father the most beautiful letter he
A B C D
had ever received.

Answer: ……

12. Many refugees want to go back to their countries **AFTER THE END OF THE WAR.**

During their campaigns, many politicians would talk about their achievements.
A B C D
Answer: …….

Appendix E
Measurement of language analytical abilities (initial version)

Part 1: Word class

Task:

One word in the key sentence (numbered sentence) will be bolded and **PRINTED IN CAPITAL LETTERS**. Your task is to select the word in the second sentence that belongs to the same family/ is of the same nature (e.g., noun, adjective, adverb etc.) as the underlined word in the key sentence.

I lost **MY** wallet.

The teacher is absent today.

A B C D

Answer: The correct answer is A

Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

1. It's possible for almost anyone to **RECORD** a great song at home or in a professional studio.
- I've never been able to get anything done around city hall without bribing someone.
A B C D

Answer: …….
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

2. I bought her all kinds of things such as stuffed animals, clothing, **AND** jewelry.
- My dog enjoys being bathed but hates getting his nails trimmed.
A B C D

Answer: ……
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

3. Has **ANYONE** seen my handbag? I cannot find it.
- In democratic countries like Canada, none is above the law.
A B C D

Answer: …….
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

4. She has been waiting for her results **IMPATIENTLY**.
- From now on, I will try to never disappoint you.
A B C D

Answer: ……
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

5. Great Expectations is my **FAVORITE** book.
- The angry politician was talking about the necessity of adopting stricter measures.
A B C D

Answer: ……..
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

6. **AFTER** that effort, everyone doubts whether she can win.
- Dwarf couple marries in a fairy-tale ceremony after they met while performing in Snow White.
A B C D
Answer: ……
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

7. Among my personal **HIGHLIGHTS** is the memory of travelling to Kenya.
- This short documentary emphasizes the importance of not texting while driving.
A B C D
Answer: ……
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

8. The **LAST** time I visited my grandparents was two months ago.
- His fans were so disappointed that he came last in the race.
A B C D
Answer: …….
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

9. Justin Bieber will have a concert **IN** Toronto.
- She left him some money on the Table.
A B C D
Answer:…….
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunction

10. At that point, they decided to call the **EXPERIMENT** to a halt.
-The call you received must have annoyed you.
A B C D

Answer: …….
Justification:
Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

11. **THIS** is the most beautiful rose I have ever seen.
-This expensive bag is not mine.
A B C D
Answer: ……
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

12. The boys wish they could **GO** to Hawaii.
- The teacher's wish list may include school supplies and gift cards for her favorite places.
A B C D
Answer: ……
Justification:
Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

13. I enjoyed **OUR** dinner last night.
- My heart nearly burst with joy when I finally brought them back from the hospital.
A B C D

Answer: ……
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

14. The school magazine comes out **WEEKLY**.
- Unfortunately, your submission to the weekly journal was not accepted.
A B C D
Answer:
Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

15. My cat loves having his head scratched **BUT** hates getting his claws trimmed.
- She studied hard for her final exam, yet she couldn't get the grade she wanted.
  A B C D

Answer: ……

Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

16. **THE** Montreal Canadiens are a professional ice hockey team based in Montreal.
- The CBC broadcasts a series of public service announcements.
  A B C D

Answer:…….

Justification: Both are (circle one): nouns; adjectives; adverbs; verbs; prepositions; determiners; pronouns; conjunctions

Part 2: Functions

Task

One word or a group of words in the key sentence will be bolded and printed in capital letters. Your task is to select the letter of the word/ group of words in the second sentence that **plays the same role/function (e.g., subject, direct object, indirect object, etc.)** in that sentence as the bolded word/group of words in the key sentence.

First, select an answer; then, provide a justification.

E.g. **THIS TEST** might be fun.

She never arrives late.

A B C D

Answer: The right answer is A.

Justification: Both are (circle one): (subject;) direct object; indirect object; verb; subject complement; object complement; adverbial

1. Their daughter **IS COMPLAINING** about the harshness of her first-grade teacher.
Montreal prohibited cigarettes in all indoor public spaces.
A B C D
Answer: ……
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

2. Because of the Cuban government, the book remained **UNPUBLISHED**.
For different reasons, this seemed a good idea.
A B C D
Answer: ……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

3. One should not drive **FAST** on snowy days.
Galileo worked hard to convince his generation that the earth revolves around the sun.
A B C D
Answer: ……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

4. We wanted to go out, but we were **TOO TIRED**.
The Greek athletes felt confident in their victory because of their extensive training,.
A B C D
Answer: ……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

5. They decided to call their new baby **JOY**.
He asked, "Have you named Mr. Jones temporary chairman?"
A B C D
Answer: ……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

6. I bought a new laptop **FOR MY BROTHER.**
Mary should never have borrowed money from her co-workers.
A B C D
Answer: ……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

7. **MARY** is happy.
Sophia can tell that her sister must have had a bad day from the look on her face
A B C D
Answer: ……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

8. He sent an invitation **TO ALL HIS CLASSMATES**.
Big corporations often give a lot of money to political candidates.
Answer:……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

9. The explorers found the old building **EMPTY**.
They elected the hardworking twins co-captains of the entire team.
A B C D
Answer: ……
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

10. German politicians **HAVE RESPONDED** to outbreaks of racial violence with demands for tighter immigration controls.

Despite their economic difference, all the other eastern states share the same pattern of
A B C D
high anti-foreign violence.

Answer: ……..
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

11. He ordered **HER CHILDREN** to stay quite.

They have completed the renovations in such a short time!
A B C D

Answer: ……
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

12. After spending so many years in Hawaii, **SHE** decided to go to Toronto.

The talented artist has traveled to nearly a hundred countries for millions of miles.
A B C D

Answer: …….
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

13. The officer gave me **A TICKET**!

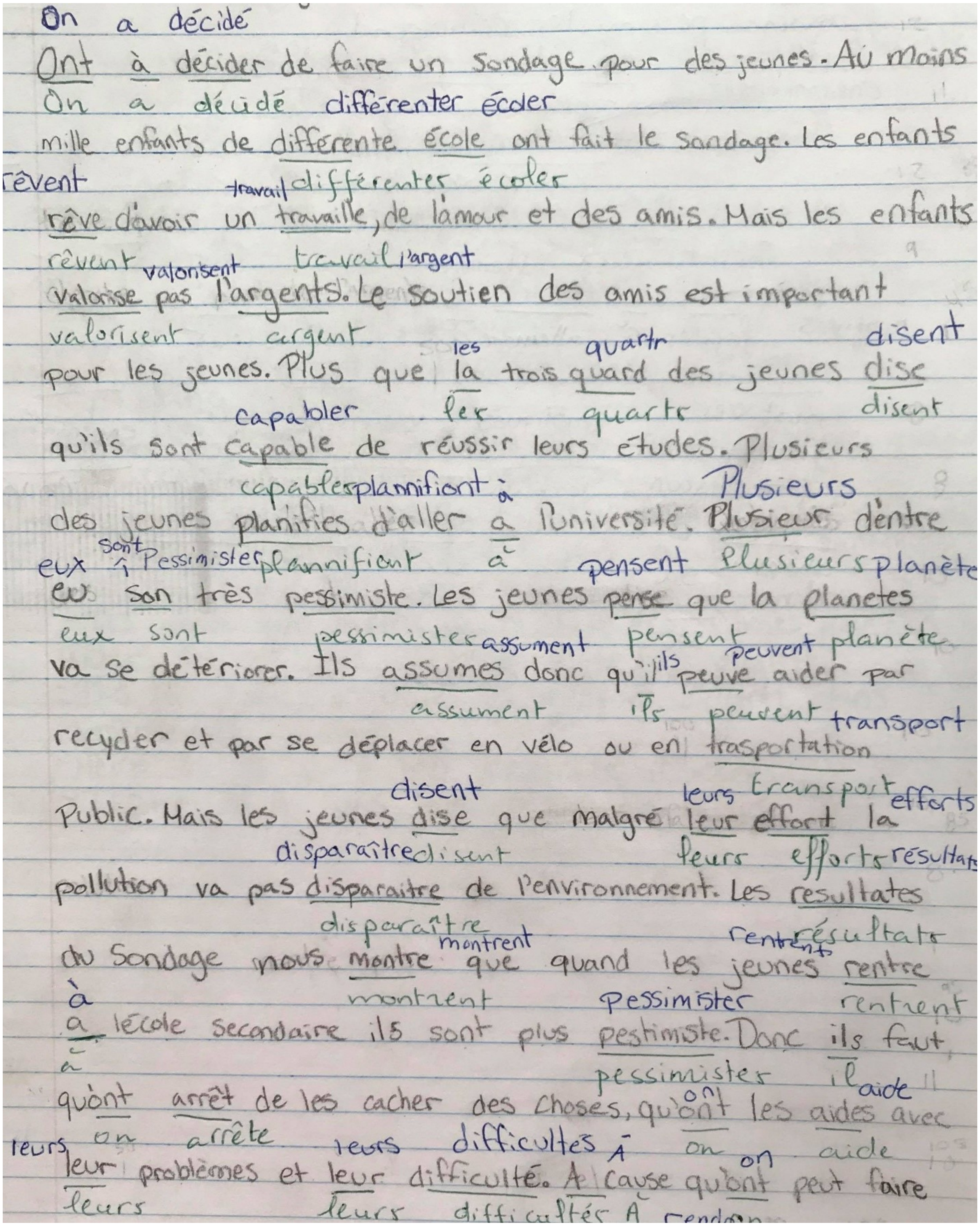
When she went away to college, Emily wrote to her father the most beautiful letter he
A B C D
had ever received.

Answer: ……
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

14. Many refugees want to go back to their countries **AFTER THE END OF THE WAR.**

During their campaigns, many politicians would talk about their achievements.
A B C D
Answer: …….
Justification: Both are (circle one): subject; direct object; indirect object; verb; subject complement; object complement; adverbial

Appendix F
A student's text annotated with direct WCF



On a décidé
Ont à décider de faire un Sondage pour des jeunes. Au moins
On a décidé différenter écoler
mille enfants de differente école ont fait le sondage. Les enfants
rêvent travail différenter écoler
rêve d'avoir un travaille, de l'amour et des amis. Mais les enfants
rêvent valorisent travail l'argent
valorise pas l'argents. Le soutien des amis est important
valorisent argent
les quartr disent
pour les jeunes. Plus que la trois quard des jeunes dise
capabler les quartr disent
qu'ils sont capable de réussir leurs études. Plusieurs
capablers planifient à Plusieurs
des jeunes planifies d'aller a l'université. Plusieur d'entre
eux sont Pessimister planifient à pensent Plusieurs planète
eu son très pessimiste. Les jeunes pense que la planetes
eux sont pessimister assument pensent peuvent planète
va se déteriorer. Ils assumes donc qu'il ils peuve aider par
assument ils peuvent transport
recycler et par se déplacer en vélo ou en trasportation
disent leurs transport efforts
Public. Mais les jeunes dise que malgré leur effort la
disparaître disent leurs efforts résultats
pollution va pas disparaitre de l'environnement. Les resultates
disparaître montrent rentrent résultats
du Sondage nous montre que quand les jeunes rentre
à montrent Pessimister rentrent
a l'école secondaire ils sont plus pestimiste. Donc ils faut
à pessimister il aide
qu'ont arrêt de les cacher des choses, qu'ont les aides avec
leurs on arrête leurs difficultes À on on aide
leur problèmes et leur difficulté. A cause qu'ont peut faire
leurs leurs difficultés A rendom...

Appendix G
A student's text annotated with indirect WCF only

Les jeunes québécois : entre rêves

et inquiétudes

Il y avait un sondage pour les jeunes en 28 écoles primaires. Mille huit cent soixante seize jeunes ont participées ce va
on participés. Se sondage vas montrer les

conventionnel
rêves et inquiétudes par les enfants. Dans les familles conventionnels

ça
sa démontre que les jeunes valorisent l'amour et des enfants.

affirment
En chacun cas, ils n'ont pas le loisir. Plus de 3/4 x jeunes affirme

à sont leur
qu'ils réussissent bien a l'école et sommes ambitieux. Leurs perception

change au milieu des rêves et l'inquiétudes.

tous ont
Quelques enfants ou presque toutes les enfant on aussi des inquiétudes.

planète
Plusieurs, ne voient pas l'avenir mais ils pensent que la planète est

mauvais veux
en mauvaise conditions. La majorité de 10-18 ans veut que la

prochains à
planète détériore dans les prochaines 20 ans. Ils disent ça a cause que

température
Il y a de la pollution d'eau, x changement de temperature.

difficultés ça
Beaucoup de difficultés et problèmes viennent entre sa. La

négative
perception des jeunes est plutôt négatif et pessimiste.

Appendix G
A student's text annotated with indirect + ME WCF

Rachid est un homme syrien qui travailleit en Syrie dans la domaine humanitaire. Sa femme était enceinte avec leur fils, Adom, et Rachid ne voulait pas que son fils passe lieu
son enfance dans un ~~lieux~~ dangereux où les enfants
As-tu besoin du pluriel ici?
doivent travailler et n'ont pas d'accès à l'éducation. Alors, lui
la
et sa femme partent pour ~~l~~ Jordanie. En Jordanie, Rachid est incapable
accord avec le sujet déterminant
de trouver un travail qui sera assez stable pour soutenir une
famille. Alors avec l'aide des passeurs, il décide de
risquer sa vie pour faire un voyage vers un pays sûr; la
entassé
Turquie. Il passe des heures innombrables ~~entassés~~ dans un véhicule
qui est entassé?
avec plusieurs autres réfugiés. Quand ils arrivent à leurs
leur
accord avec le nom
destination, ils embarquent un bateau pour traverser la frontière
de la étaient
et s'éloigner ~~du~~ guerre. Leurs instructions ~~était~~ d'aller vers une
accord avec le nom accord avec le sujet
lueur loin dans l'obscurité, mais personne ne savait comment
conduire un bateau. Alors, ils devaient se débrouiller car c'était
était
leur seule chance à survivre. Chacun des personnes ~~étaient~~
accord avec le sujet
anxieux d'arriver, et ils sont arrivés Rachid et sa femme
choix de l'auxiliaire revoir l'accord avec le sujet