2011-07-06

Comparing the Effects of Two Forms of Dynamic Corrective Feedback On Four Characteristics of English Language Learner Writing

Judson Mc Kay Hart
Brigham Young University - Provo

Comparing the Effects of Two Forms of Dynamic Corrective Feedback On Four Characteristics of English Language Learner Writing

Judson McKay Hart

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Dr. Neil J. Anderson, chair
Dr. Norman W. Evans
Dr. K. James Hartshorn

Department of Linguistics and English Language

Brigham Young University

August 2011

ABSTRACT

Comparing the Effects of Two Forms of Dynamic Corrective Feedback
On Four Characteristics of English Language Learner Writing

Judson Hart
Department of Linguistics and English Language, BYU
Master of Arts

Particular English language learners have a need to demonstrate high levels of accuracy in their written and spoken language production. Dynamic corrective feedback (dynamic CF) has been shown to facilitate L2 learner written accuracy attainment through providing manageable, meaningful, timely and constant feedback on authentic and frequent written language production. The research of this thesis examined the benefits of altering the dynamic CF model to be responsive to students' spoken production rather than solely their written production and measured the impact this adaptation would have on the established benefits of the instructional strategy on students' gains in written accuracy. The study also looked at the impact of both forms of dynamic CF on students' attainment of written complexity, fluency, and lexical development.

The study included two groups of students whose language proficiency ranged from intermediate-low to advanced-mid who were studying English for academic purposes in an intensive English language program. These students participated in a one-semester Linguistic Accuracy course. Half of the students received the traditional form of dynamic CF in which they received feedback on only their written production, and the other half received a modified version of dynamic CF that provided students with feedback on only their spoken production.

Before and after the treatment, samples of students' written production were collected through a thirty-minute essay test. These writing samples were analyzed for accuracy, fluency, complexity, and lexical development. Changes in each of these variables for both groups were contrasted using a mixed-model repeated measure Analysis of Variance (ANOVA). These tests revealed that there was not a significant difference in terms of the changes in accuracy or lexical development between the two groups; however, participants receiving the modified variation of dynamic CF did perform significantly better on the measurement of written complexity. Also, lower proficiency students receiving the modified version of dynamic CF did significantly better on the measurement of written fluency than students of a similar proficiency receiving the traditional form.

Keywords: [dynamic corrective feedback, L2 learner, writing, speaking, accuracy.]

ACKNOWLEDGEMENTS

I would like to acknowledge all of those whose support and guidance have allowed me to complete this process including my committee members—my chair, Neil J. Anderson and the other committee members Norman W. Evans and K. James Hartshorn. Their expertise and professional examples have greatly impacted the end result of this project. Although too numerous to individually thank here, I would also like to acknowledge the impact and influence of all of the educators behind my TESOL education.

Second, I would like to thank all those associated with the English Language Center who have provided invaluable support to this study, including the teachers and students who participated in the study and those friends and colleagues who helped me through each stage of the research process. In addition, I am indebted to those whose research informed my work here, particularly those studies in dynamic CF that were used as guides and models.

Third, I would like to thank my family. My wife, Jan, and our children, parents and siblings bring unconditional love, support and happiness to my life daily; without them any accomplishment would not be possible or meaningful.

Finally, I give my appreciation to those whose donations and financial support have made my education at Brigham Young University a reality. It truly is a unique institution and I have been extremely fortunate to learn and grow here.

Table of Contents

Index of Tables

Index of Figures

Chapter 1 Introduction

Ambitious Learners Need Accuracy Achievement

The ambition of English language learners can surprise and sometimes puzzle even the most experienced TESOL professional. Many of these learners desire not only to assimilate into English-speaking communities but also to achieve power and prominence in English-speaking contexts. English language classrooms are often filled with would-be diplomats, boardroom executives, surgeons, published academics, attorneys, and politicians. Many of these students plan on using a degree from a university where English is the primary language of instruction to access social and corporate strata beyond their current reach.

While English-language ability alone will not ensure success, insufficient English-language ability in today's global professional communities will surely limit student's accomplishments in the workplace. In light of their own expectations and the standards of the admissions, hiring, and promotion committees awaiting them, instruction that leads these students towards a high level of language refinement is more than a reasonable expectation. Critics of accuracy focused instruction have not provided the TESOL community with a working model for students who demand help in their endeavors of becoming the best, including the most accurate, language producers they can become. If instruction can do nothing to help these learners achieve some degree of the language refinement they desire, language programs should send them away with their pockets unpicked (Folse, 2011).

In spite of the long history of the debate about whether or not error responsive instruction is efficacious, there is sufficient evidence that the right methods of error responsive instruction and error feedback can help learners achieve more accurate language production (Doughty & Varela, 1998; Lyster & Ranta, 1997; Spada & Lightbown, 1993). Although there is not a single

solution or quick fix to all the accuracy problems facing English language learners, there are methods that have demonstrated a marked effect toward accuracy achievement (Bitchener, 2008; Ellis, 1998; Ferris, 1995; Reid, 1997). This study focused on the modification of one such method: dynamic corrective feedback (dynamic CF).

Pedagogically, dynamic CF is both an instructional process and a corrective approach. Students produce a small sample of language. The teacher identifies the locations and types of inaccuracies present in the sample using a code of error types. Students correct these inaccuracies and resubmit the sample—receiving additional feedback until the sample reaches an error-free state. In its traditional format, dynamic CF has been shown to improve the accuracy of L2 student writing by providing error feedback that is meaningful, manageable, timely, and constant (Evans, Hartshorn, McCollum,Wolfersberger, 2010; Hartshorn, Evans, Sudweeks, Strong-Krause, Anderson, 2010; Hartshorn, 2008; Lee, 2009).

A short definition of these guiding principles of dynamic CF will be included here as they will be referred to throughout this document. Meaningful feedback engages the learner cognitively. Feedback is meaningful when it is relevant and specific to a student's language needs and leads the learner towards more accurate production. Feedback is manageable when the process of giving it does not overwhelm the teacher and the process of receiving and responding to it does not overwhelm the student. Timely feedback is feedback that is given and received while the initial production and subsequent modifications remain fresh in the students' memory. Constant feedback refers to both opportunities to receive and apply feedback occurring frequently.

As it has been previously applied, DCF dynamic CF has responded only to students' written English production. The research in this study explored the possibility and implications

of modifying the traditional dynamic CF instructional strategy to direct feedback toward students' spoken English production rather than only students' written English production. The eventual goal of this research is to analyze whether such an adaptation provides additional spoken language benefits of increased accuracy while maintaining levels of spoken fluency and complexity. However, this initial research measured the extent to which the modification of dynamic CF leads to the established gains in writing accuracy achieved by the traditional method while sustaining other writing sub-skill performances including fluency, complexity and lexical development. This study also attempted to show that the modification is true to the guiding principles of the original instructional strategy. Irrespective of practice mode, dynamic CF instructional strategy can provide feedback that is manageable, meaningful, timely, and constant, and such feedback can consequently help learners become better producers of the language.

Dynamic CF has the potential to address the limitations facing TESOL practitioners in providing their students with the feedback and instruction these learners need to move towards greater accuracy. In order for focus on form instruction and error feedback to be effective, it must be sustainable so that students eventually receive the critical mass of both intensive and extensive exposure and practice necessary for actual accuracy improvement (Evans et al., 2010; Hartshorn, 2008; Hartshorn et al., 2010). As presented in this study, in both of its variations, dynamic CF promotes manageability as the keystone to providing feedback that has the necessary components of timeliness and constancy while being meaningful and tailored to students' demonstrated needs.

Dynamic CF provides a curricular core for a Linguistic Accuracy class: informing all phases of the present practice and perform stages typically found in an approach to grammar instruction. Extracting accuracy objectives from traditional productive skills courses and

addressing them in their own course helps to avoid conflicts with the communicative instruction trends prevalent in current writing and speaking pedagogy. Research has suggested that when accuracy instruction and feedback are only one of several objectives, it may be significantly less effective (Evans et al., 2011). Consequently, language programs geared towards preparing learners to perform in accuracy sensitive contexts should consider addressing production accuracy objectives in isolation.

Finally, the affective investment of students in any instructional strategy is an important concern. A student disinterested in the goals of accuracy instruction or the methods used to achieve them will not exert the sustained effort they require. Dynamic CF assures students that their individual accuracy needs are being met by tailoring feedback and instruction directly to these individual student's own language production. The instructional strategy also employs several components that provide learners with a way to chart their progress in becoming more accurate producers of the language. Thus, having concrete or tangible evidence of students' personal improvement resulting from their investment of time and energy can help students through the ebbs and flows of personal motivation (Dornyei, 2000).

A New Approach to Dynamic Corrective Feedback

These sound pedagogical principles and the reported needs of students and teachers provide adequate justification of the need for wider promulgation of dynamic CF as an instructional strategy, as it provides a cohesive curriculum that addresses L2 language learners' accuracy needs by providing individualized and meaningful feedback that is timely, consistent, and manageable. The effectiveness of this instructional strategy, in multiple contexts and with a range of language proficiencies, has been substantiated by repeated quantitative analyses (Evans et al., 2010; Evans et al., 2011; Hartshorn, 2008; Hartshorn et al., 2010; Lee, 2009). The results

of these analyses have shown that students prefer dynamic CF over traditional grammar instruction approaches (Lee, 2009). The results have also shown that students' written accuracy significantly increases more through dynamic CF than through traditional process-writing instruction and feedback (Evans et al., 2011; Hartshorn et al., 2010).

While students' accuracy needs related to writing are well-supported, writing is only one of language's two productive modes (Bates, Lane, & Lange, 1993; Ferris, 1995; Leki, 1990). Students in many advanced language production contexts need to be accurate speakers as well as accurate writers of English. Because of the intensive investment required for accuracy improvement, second language practitioners should be interested in ways to boost accuracy levels in both productive skills simultaneously or with only slight modifications of practice modes. However, much of the recent research has looked at the impact of a particular method on accuracy within a productive skill rather than widening their scope to investigate accuracy impact of particular practices across and between skills.

While the dynamic CF instructional strategy has been show effective with university matriculated ESL learners, the study presented here measured the instructional strategy's impact in an intensive English language program in Provo, UT: Brigham Young University's English Language Center (ELC). Although the ELC has a program addressing basic life skills English, the student participants in this study were enrolled in the Academic program which emphasizes the language skills necessary for eventual matriculation in a university. The treatment was applied in a Linguistic Accuracy class which has the primary objective of increasing students' productive language accuracy. The students attend classes for four and a half hours a day, with instructional schedules consisting of Reading, Writing and Listening/Speaking skills classes in

addition to the Lingusitic Accuracy class. The average class size is 15 students from a wide variety of L1 and cultural backgrounds.

The English Language Center is a lab school that serves as part of the educational training of Brigham Young University's undergraduate and graduate TESOL courses. While all of the teachers at the ELC have at least some graduate education in TESOL, levels of expertise and experience do vary widely. However because of the rigor of the objectives of the Linguistic Accuracy curriculum, administrators are selective in making teaching assignments to these particular courses. Also these teachers are compensated for the extra time and preparation that maintaining the dynamic CF curriculum demand.

This study focused on answering the following research questions as they apply to two forms of dynamic CF: one of these forms is similar to previous applications of dynamic CF and informs a corrective dialogue between student and teacher initiated by the collection of a student writing sample; the other form of dynamic CF has been modified to respond to student's spoken production that has been transcribed and then informs the corrective dialogue. These questions will be stated here and reiterated throughout the rest of this thesis:

1. Can the dynamic corrective feedback instructional strategy be altered to respond directly to students' speech?
2. Are there statistically significant differences in changes to written accuracy between the two groups that result from differences in the practice mode that receives feedback?
3. Are there other statistically significant differences including gains or losses in written fluency, complexity, and lexical development that result from differences in the practice mode that receives feedback?

Chapter 2 Review of Literature

This chapter presents the research foundation for the study presented in this thesis. The chapter begins with the assertion that accuracy in form is important for some English language learners. A particularly controversial aspect of grammar instruction, corrective feedback, is then explored. Following the presentation of both sides of this debate, the relationship between error correction and theories of second language acquisition (SLA) is discussed. Two components of current views of SLA, connectionism and skill acquisition theory, are explored as they relate to the relatively new approach of dynamic corrective feedback (dynamic CF).

The evolution of this approach as it has been applied at the English Language Center (ELC) at Brigham Young University is then discussed in depth through examining several research studies on the effect of the approach on students' accuracy when writing. Consequences for the current use of dynamic written corrective feedback in the ELC's linguistic accuracy curriculum are discussed including the need for more balanced practice opportunities in both modes of language production: speaking and writing. Rationalization of this suggestion is provided through key points identified by the instructional strategy's founding practitioners as they relate to skill acquisition theory.

Finally, the research validation needed to justify the modification of the currently practiced form of dynamic corrective feedback is discussed. The chapter concludes with the restatement of the objectives of this study and research support for the primary modification proposed in the adaptation of the method necessary to be responsive to students' speech.

Accuracy in form is important for some English language learners

Accuracy in form is important for some English language learners in certain language contexts (Evans et al., 2010). As it relates to form, accuracy can refer to the extent to which learners' production of the language approximates the rule system of the target language (Skehan, 1998). However, this rule system should not be seen as a static quality of the language, rather, as a dynamic variable that becomes both alternatively forgiving and increasingly exacting depending on the context of language production (Celce-Murcia, 1991). While in the company of friends at a dinner party, if an English language learner intermittently and erroneously exchanges his gender pronouns, the slight error would be of little consequence; however, place the same speaker in a courtroom, and suddenly accuracy in the use of gender pronouns matter a great deal.

Characteristics of accuracy will also vary depending on the skill lens through which communicative ability is viewed. While there is a great deal of overlap between the productive skills, conditions of accurate speaking will not be the same as conditions for accurate writing. Sufficient accuracy in form, in both productive skills, is a critical language component that must be demonstrated by ESL students in order to gain access to particular language contexts such as academia and some professional circles (Eskey, 1983).

The gatekeeper effect of accuracy to these two types of language communities is evidenced by accuracy components to several standardized proficiency measurements. The American Council of Teachers of Foreign Languages (ACTFL) proficiency guidelines, a widely accepted set of proficiency measurements, articulate accuracy expectations when they describe speakers who qualify for their highest marks saying: "They demonstrate virtually no pattern of error in the use of basic structures" (Breiner-Sanders, Lowe, Miles, & Swender, 1999, p. 3). For writers, the ACTFL guidelines are similarly exacting and express that: "Writers at the superior

level demonstrate a high degree of control of grammar and syntax" (Breiner-Sanders, Swender, & Terry, 2001, p. 3). The Test of English as a Foreign Language (TOEFL) standards that are often used for university and professional admissions also express the need for language learners to demonstrate effective use of grammar and vocabulary and exhibit a high degree of automaticity and good control of basic and complex structures (Educational Testing Service, 2004, p. 1)

Determining which language learners have a high need for accuracy focused instruction is an important part of the needs analysis that should be performed by the language teacher and language program administrator. Celce-Mucia (1991), in a prominent article on trends of grammar instruction, included a very helpful way of proceduralizing this needs analysis by evaluating any given student's position on several discrete continuums. These variables are divided into two categories: learner variables, which include age, proficiency, L1 educational background; and instructional variables, which include skill, register, and need/use. The continuum for each of these variables and subset variables spans from grammar instruction being of high importance to grammar instruction being of little importance. Celce-Murcia's figure has been recreated in Table 1.

Table 1

Determining Students' Needs for Focus on Form Instruction

	Focus on Form Instruction Less Important		Focus on Form Instruction More Important
Learner variables			
Age	Children	Adolescents	Adults
Language proficiency	Beginning	Intermediate	Advanced
Educational background	No formal education	Some formal education	Well educated
Instructional variables			
Skill Emphasis Needed	Listening, Reading	Speaking	Writing
Register	Informal	Consultative	Formal
Need/Use	Survival	Vocational	Professional

Placing the students featured in this study on Celce-Murcia's continuum provided justification for the tenets of the focus on form pedagogy presented throughout this study. The population of the Academic program of the ELC aligned with few exceptions against the right side of the continuum for each variable including age, proficiency, and literacy development. Even the youngest student in the study groups from the Academic program met an approximate age distinction of adulthood. The proficiency of students in the Academic program has been determined to be at minimum intermediate low. Few students come to the Academic program of the ELC without extensive L1 literacy development; most arrive with at least a high school diploma and many with some completion of secondary education.

As recommended by the instructional variables in the figure, instructors should look at skills of emphasis, register (or the degree of closeness to the majority of interlocutors a student will or does regularly interact with), and the consequences of inaccuracy within the contexts where the language will be applied (Celce-Murcia, 1991). By virtue of their enrollment in a program geared toward future application of English in academic contexts, students featured in this study have demonstrated their demands for instruction that facilitates more accurate production of the language. These students, with ambitions to engage in university study, also have a high need for writing skills. Many others are learning English because of professional ambitions and applications that will require spoken accuracy levels exceeding those needed for basic life-skills English.

Controversy over Corrective Feedback in Accuracy Instruction

The question of whether or not teacher feedback, and more specifically error correction, should be part of the process through which learners become more accurate has been one involving considerable debate. Some second language researchers and language instructors have voiced concerns of both the ineffectiveness of error correction in increasing accurate production and even raised declarations of potential harm that can result from error correction (Hendericksen, 1978; Lalande, 1982; Semke, 1984; Truscott, 1996). Proponents of correction have provided research supporting the essential role of feedback, including error correction, in facilitating accuracy acquisition (Bitchener, 2008; Ferris, 1999; Hyland, 2003; Polio et al., 1998). As French essayist Joseph Joubert (1867) concluded "It is better to stir a question without deciding it than to decide a question without stirring it" (p. 84). Controversy in scholarship often leads to centering in pedagogy; for this reason the summarized debate that follows has ultimately led to better reasoned practices in relation to effective feedback instructional strategies.

While entire literature reviews have documented the often highly impassioned and research-foddered exchanges between the two camps divided on the issue of feedback, this section will highlight two exchanges of particular significance, which can be used to summarize the arguments and counterarguments in the case for and against error correction. The argument against error correction will be presented by Truscott (1996; 1998) in two articles condemning corrective feedback in a written skills context and an oral skills context. The argument in favor of correction will be presented through two responses to Truscott's initial attacks. One rebuttal is offered by L2 writing specialist, Ferris (1999), and the other from oral skills specialists, Lyster, Lightbown, and Spada (1999).

Truscott (1996) struck first with his criticism of corrective feedback practices in L2 writing pedagogy. As mentioned above, Truscott declared that overt grammar correction was not

only ineffective, it was harmful. He further condemned the practice stating that time and resources directed towards grammar correction are time and resources wasted, arguing that despite the storied history of its practice, grammar correction does not lead students to more accurate production of written language. Truscott strung together several examples of research that failed to show grammar correction's efficacy and even some that insinuated correction contributes to accuracy decline, bolstering his identification of specific reasons that grammar correction fails to realize the intentions of the instructor and the aspirations of their students (e.g.; Knoblauch & Brannon, 1981; Krashen, 1992; Long, 1977).

For language instructors, the most significant of Truscott's conclusions included first, the claim that correction fails because it is irregular and mis-timed as instructors often failed to notice errored production or respond to it within the short window that the learner was sufficiently cognizant to use the feedback (Cohen & Cavalcanti, 1990, Cohen & Robbins, 1976; Zamel, 1975). Second, teachers who notice the error may not have known how the error should be corrected (Cohen & Robbins, 1976; Zamel, 1975). Also, instructors face a complex decision on the extent to which errors should be targeted. That is, whether every error should be addressed (comprehensive correction) or just errors of a particular type (selective correction). Truscott cited research that led to his claim that results for these varying degrees are the same: the only difference between the two is whether the instructor desires to waste their efforts comprehensively or selectively (Knoblauch, 1981; Hillocks, 1986).

Truscott concluded that correction is harmful for teachers because it consumes time that would be better allocated elsewhere; correction is laborious, tedious, and ineffectual. Truscott also examined correction from the vantage of students in identifying why correction fails to lead to accuracy. His conclusion was that students who committed the error likely did not

understand the correction of the error, something that has been identified repeatedly in research and most instructors' personal experiences (Cohen, 1987; p. 350; Hayes & Daiker, 1984; Leki, 1990). Truscott also cited research that showed no direct link between increased mastery of grammar concepts and increased accuracy when applying these concepts in language production (Gass, 1983; Green & Hect, 1992; Sorace, 1985). Even if students understood and corrected the error, this reclaimed accuracy was likely to be fleeting, and due to inconsistency in feedback, errors may not be identified in subsequent breakdowns (Cohen & Robbins, 1976; Zamel 1985).

The need for feedback also assumes that students care to be corrected or even care to be correct—both erroneous assumptions according to Truscott. He cited research that observed that following feedback students did nothing but mentally review the feedback—never engaging in the actual corrections the feedback intended to direct (Cohen, 1987). Other students also expressed their perception of revision after receiving feedback as more punishment than reward (Cohen & Calvacanti, 1990; Radeki & Swales, 1988). Other researchers have claimed that students' apathy toward correction reflects a general apathy toward correctness present everywhere other than language classrooms (Santos, 1988). Students receive some validation for errored production and it may dampen their motivation to endure the pain of error identification, correction, and further attention to accurate production (Leki, 1991).

Truscott (1996) further extended his case against correction with the claim that corrective practices would be tolerable if they were merely benign, but sufficient research shows that correction as a practice is not only impotent—it is also harmful. According to these sources, correction undermines the relaxed and confident learning atmosphere that shields important affective processes and enables language acquisition including accuracy development (Gardner & MacIntyre, 1993a; 1993b). Included in this attribution of harmful effects, Truscott used

research, mainly on L1 correction, which showed that overt correction led to greater levels of dissatisfaction with the particular skill area being targeted (Hillocks, 1986; Knoblauch & Brannon, 1981). Further, L2 writers in correction-free contexts have been shown to write more and enjoy it to a greater degree than their corrected peers (Semke, 1984). These researchers concluded that the accuracy levels attained between corrected writers and non-corrected writer were not significantly different. However, in terms of ultimate accuracy attainment as Truscott suggests, the depressed affection for writing among corrected writers could have long-term consequences for accuracy stagnation or even a decline in accuracy (Truscott, 1996).

Several years following the release of his case against error correction and with the firestorm that followed in full blaze, Truscott (1999) expanded his attack against correction beyond writing curriculums, identifying what's wrong with oral grammar correction. While the fundamentals of this new attack were repetitious, Truscott expressed that in an oral speaking context the problems for both teachers and students in allocating attention resources to identification, prescription, and correction of errors are only exacerbated by the dynamic nature of the exchange. Research has shown the potential for failure at each of these junctures in the transformation and exchange of meaning to be high (Roberts, 1995). Truscott expressed validated doubt that even if each of the required stages of correction was successfully executed, the effect of the correction impacting future production was tenuous.

In the case of both publications, Truscott received strong reverberation from supporters of various applications of corrective feedback instructional strategy. The well-reasoned responses of the published rebuttals presented below show that the issue is far from being as cut-and-dry as Truscott's tenor would lead one to believe. In her first response to Truscott's first publication, Ferris (1999) expressed her initial wishful desire to prove Truscott correct and

liberate herself and colleagues from the tedium and relentless duty to provide error correction. However, she concluded that Truscott's cry that grammar correction be abandoned was premature and overly drastic. In reaching this conclusion, she identified several instances where the research that he cited in support of his central argument is overstated or where he minimizes or obfuscates research findings that would run contrary to his position.

While there were other concerns identified by Ferris (1999) about the research foundation of Truscott's argumentation, most interesting for the purposes of this study are both the contradictions and the contemplations she develops surrounding the practical implications of Truscott's case. Central to the contradictions noted by Ferris is the rationale that while poorly done correction will be of little to no benefit and may even be of detriment to learners, not all correction deserves to be labeled poor. Contrarily, there is a growing body of evidence that effective correction—correction that is selective, prioritized, and clear—does facilitate accuracy improvements (Bates, Lane & Lange, 1993; Ellis, 1998; Ferris, 1995; Reid, 1997). Truscott (1996) may have been too quick to equalize all forms of error treatment in light of this research that shows that all feedback is not created equal.

Ferris (1999) also highlighted Truscott's minimization of the impact of context on the effectiveness of error correction: the *how, when, where,* and *who* of the corrective dialogue matters. Ferris professionally conceded to some points in Truscott's argument that she declared needed further attention. She noted that the problems in terms of limited resources, including time, knowledge and attention, were valid limitations for effective error correction with implications for both teachers and students. She also suggested that there was a need for selectivity in error feedback and the development of a systematized approach for teachers in identifying treatable errors and educating students in using feedback to correct instances of error

and inform future production. Better teacher training and feedback pedagogy can help teachers' feedback be consistent, correct, and clear in their identification of errors and explanations of corrections (Ferris, Harvey, & Nuttall, 1998).

Ferris acknowledged that students who are undermotivated toward accuracy can render the feedback dialogue ineffective. However, motivation can increase when students have a clear vision of how to use error feedback and see that error feedback addresses their specific language needs by adjusting to their proficiency, language background, and prior corrective feedback experience (Ferris, 1999). Also, students can be convinced, not duped as Truscott insinuated, that accuracy matters--not just in the classroom but also the real world language contexts that many are trying to access (Bates, Lane, & Lange, 1993; Reid, 1997).

In response to Truscott's case against oral grammar correction, Lyster, Lightbrown, and Spada (1998) constructed a similarly well-reasoned refusal of Truscott's central cry for abandonment of correction. They, like Ferris, claimed that although at times difficult and with delayed benefits, improvement through corrective feedback is feasible, effective, and often necessary. Primary to the development of this research study was their research-based conclusion that correction is not inherently disruptive, traumatizing, or overwhelming (Doughty & Varela, 1998; Lyster & Ranta, 1997; Spada & Lightbown, 1993). Also, they addressed the reality that correctional effects of feedback should be seen as gradual and not necessarily constrained by developmental sequences (Lightbown, 1998).

Truscott has continued to be prolifically published in his stance against corrective feedback and, while there has been some evolution of his argument, particularly in his attempts to ground its tenets in theories of SLA, he remains strongly opposed to even the most evolved methods of correction currently being practiced (Truscott, 2007; Truscott & Hsu, 2008; Truscott,

2010). However, there is significant evidence that supports a move toward center by the majority of researchers and practitioners, many of whom claim that while the correction debate of the end of the 20th century spurred introspection, healthy dialogue, and a move toward better research practices, the time to move forward from bickering has arrived (Bruton, 2010).

Bruton (2010) expressed this conclusion very well when he identified that while the issue of aiding learners' accuracy achievements is more relevant than ever, the debate for and against correction has become "tedious, sterile and academic" as the arguments for correction nihilism have become redundant, dated, and divorced from most L2 instructional realities (p. 491). Others have added that error correction is, and likely always will be, an expectation students hold for their teachers (Ferris 1995; Leki, 1991; Radeki & Swales, 1988). The time to question whether or not to incorporate corrective feedback into the language classroom has passed; rather, researcher and language instructors should be asking how to best meet their students' reasonable expectations for corrective feedback and maximize its potential value to students (Evans et al., 2010; Hartshorn et al., 2010).

Influence of SLA Theories on Instructional Practices

Views on the significance of error correction and feedback are tied to theories of SLA. A historical overview of these theories can account for much of the evolution of attitudes toward error correction that have been discussed. Early researchers of language acquisition, looking through the behaviorist view of language learning, saw feedback as the critical reinforcement necessary in the formation of accurate language habits and the correction of inaccurate habits. Chomsky (1959) shifted the focus of language acquisition with his generativist view of language acquisition away from patterns of conditioning. Innateness of language parameters, biologically set, elicited notions that corrective feedback was of little aid to the acquisition of a first language.

These notions would be extended to theories of SLA research which resulted in a reduction error correction in second language practice and instruction (Leeman, 2007).

Recent turns in SLA theory are guiding new applications of feedback that are resulting in measureable and significant impact on learners' accuracy performance. These trends toward connectionist models of SLA have caused researcher and language practitioners alike to reexamine the significance of the linguistic environment, including the element of feedback, in shaping the development of accurate production (Leeman, 2007). One of the more important tenets of connectionism as it relates to error correction is the concept that linguistic knowledge is represented as a bank of associations rather than a book of formal rules (Plunkett, 1995). Repeated exposure through various cognitive channels strengthens these associations and facilitates access of the information for production.

For this reason, frequency and statistical prominence of input are seen as major factors of acquisition (Ellis, 2002). Feedback is the mechanism by which learners adjust the strength of associations as they filter out input that is misleading and correct output that is incorrect. The learner relies more on the association if it produces a correct result and less when the association is proven flawed (Plunkett, 1995). Related to the evolution of the connectionist model are the relatively modern views that SLA shares much more in common with acquisition of other complex cognitive skills than previously thought. Language learning is governed by a common set of general learning mechanisms (Leeman, 2007).

This view, commonly referred to as skill acquisition theory, describes three cognitive stages that a learner of skill, such as L2 acquisition, passes through (Dekeyser, 1997). The first stage is the acquisition of new declarative knowledge. The retention and application of this knowledge requires high demands on attention, processing, and memory requirements. As the

cognitive structure is presented with sufficient reason for increased efficiency in accessing and applying this general knowledge, the second stage of proceduralization reduces the cognitive load. Proceduralization occurs as the mind maps out more efficient ways to access and retrieve stored knowledge. In the final stage of automatization, the cognitive demands of access and application becomes minimal, requiring little allocation of attention, processing and memory resources which allows for several skills to be simultaneously and efficiently executed, seemingly unconsciously (Anderson, 1983). Feedback is seen as valuable for each of these cognitive stages, although the reliance on feedback becomes less important as learners advance toward automatization (Leeman, 2007).

Theoretical Foundation of Dynamic Written Corrective Feedback

These two views of language acquisition have formed the theoretical basis for the development of a new approach to feedback pedagogy known as dynamic written corrective feedback (dynamic WCF) that has been applied with marked success at improving students accurate production of written English (Evans et al., 2010; Evans et al., 2011; Hartshorn et al., 2010). These researchers connect this instructional strategy and its goal of improving writing accuracy to tenets of the skill acquisition theory described by DeKeyser (2007), stating that declarative knowledge is requisite for the development of procedural knowledge and proceduralization requires extensive and deliberate practice before it becomes automatized.

The dynamic WCF as an instructional strategy corrects several gaps these researchers have seen in previous methods aimed at increasing accurate production. First, they identify that opportunities for feedback should be frequent and authentic–characteristics that have been lacking in accuracy-focused writing tasks. Second, feedback and instruction should be informed by actual errors produced by students engaged in these frequent and authentic opportunities for

production. This approach is in contrast to methodologies that attempt to improve student performance on stock-selected error types and learner generalities (Bitchener & Knoch, 2009; Sheen et. al., 2009) Finally, both feedback and productive tasks should be meaningful, timely, constant, and manageable (Evans et al., 2010; Hartshorn et al., 2010).

In further explication of these principles of dynamic WCF, meaningful feedback is achieved through indirect correction that engages the learner cognitively. Timeliness and constancy are achieved through shortening the time between production and feedback; in well-executed applications of dynamic WCF pedagogy, instructors return feedback on students' work by the next instructional hour. Manageability is achieved by limiting the amount of correction through limiting the amount of production submitted for feedback. Current applications of the instructional strategy include the daily production of a ten-minute writing sample that is the genesis of a cyclical process of feedback and correction that ends with production of an error-free product and a better-informed and practiced producer of the language (Evans et al., 2010; Hartshorn et al., 2010).

Research Analysis of Dynamic Written Corrective Feedback Instructional strategy

Initial action research

One of the first of these studies as reported in *Language Teaching Research* consisted of two applied grammar classes (n=10 and n=12) over two separate semesters that participated in the dynamic WCF instructional strategy. In addition to the indirect marking of errors, each of the approximately 30 ten-minute paragraphs produced by the students were given a holistic score that was a composition of instructor perceived linguistic accuracy (75%) and content development (25%). At the end of each semester, the students' performance trends showed a consistent and significant increase in the holistic score assigned each draft (Evans et al., 2010).

Because these holistic grades were feared to have some flaws in terms of both reliability and validity, the researcher bolstered the accuracy of this conclusion with the addition of quantitative analysis of error-free clauses of the first quarter of the tasks of the semester and the last quarter of tasks in the semesters. This significantly more objective and reliable analysis confirmed the holistic conclusions of the first and showed an increase of the error-free clause ratio from an average of 43% to nearly 55% between the two groups (Evans et al., 2010).

While indicating promise for the application of dynamic WCF, this study had several limitations that would be addressed in subsequent studies. First, no control group had been utilized so there was nothing against which to compare the performance of the group receiving the new treatment. Second, because the data used in evaluation of the effect was just that collected through the 10-minute practice paragraphs, the authors concede that the study was left vulnerable to criticisms of external validity. Ten-minute paragraph responses are not representative of a typical length in academic writing and subsequent studies would need to show that gains in accuracy could be extended to longer production tasks. In addition to these concerns, net gains in accuracy needed to be demonstrated as occurring without net losses in rhetorical competence, complexity, and fluency (Evans et al., 2010).

Quantitative analysis of dynamic WCF's impact on accuracy, fluency, complexity.

The second study (chronologically) reported in *TESOL Quarterly* addressed all three of these concerns and also included a much larger sample size (N=47), a wider age sampling (18-45 in the treatment group), and a broader range of proficiencies (advanced low to advanced mid) (Hartshorn et al., 2010). For this study, writing accuracy gains achieved by students participating in the dynamic WCF instructional strategy were contrasted with accuracy gains made by students receiving more traditional writing instruction with typical feedback and productive tasks. This

study showed similar results to the first with the treatment group making considerable gains in accuracy over the contrast group, which actually saw a slight decrease in overall accuracy. The measurement of these gains was quantified through a pre-test/post-test design, which extended the length of the writing task to 30 minutes (Hartshorn et al., 2010). This length had greater external validity since it is used by several standardized exams including the TOEFL and was also more representative of authentic academic tasks (Educational Testing Service, 2004).

This second study also addressed potentially adverse, yet previously unexamined, implications of the instructional strategy through analyzing the text produced by the participants on both the pre-test and post-test in terms of rhetorical competence, complexity, and fluency. Rhetorical competence, as measured by blind evaluation of the essays by trained raters, was reported as essentially equivalent between the two groups; this suggests that the treatment had no adverse (or beneficial) effect on rhetorical competence (Hartshorn et al., 2010). Although both groups showed gains in terms of both fluency and complexity, the effect on writing fluency (operationalized as total words in the writing sample) and complexity (operationalized as mean length of T-unit) were measured as slightly higher in the contrast group. This imbalance was significantly smaller than the difference between the two in terms of accuracy gains. The researchers appropriately note, "One might well ask, 'What is the true value of small gains in writing fluency or complexity when the substance of those gains is laden with linguistic errors that undermine communicative efficacy?'"(Hartshorn et al., 2010, p. 102).

Like the first study, the second also had some room for improvement in its design and execution. First as identified by the researchers, it lacked an element of true experimental design in that it did not randomize participants between the study groups. The classes were intact at the onset of the study and had only been balanced by the institutional administration for proficiency,

L1 background, nationality, and gender. The researchers performed analyses that checked many of the potential inequalities between the two groups, all of which returned no statistical difference but acknowledge that future research should, where possible, include randomly assigned groups (Hartshorn et al., 2010). Second, this study did not advance the understanding of which components of the instructional strategy have the greatest effect on accuracy gains, i.e., consistent output, error identification, indirect correction, error logs, and error focused instruction. Finally, in terms of the potential effects of context (intensive English program), proficiency (advanced low) and length of the treatment (single semester), the study also did not further substantiate the benefits of the instructional strategy to a broader sample of the potential population (Evans et al., 2010; Hartshorn et al., 2010).

A New Instructional Context for Dynamic Written Corrective Feedback

A third study, currently in press, addressed one of these contextual variables by using an experimental design very similar to the previous study with the primary exception being that the study groups consisted of university matriculated students rather than students in an intensive English program (Evans et al., 2011). This study substantiates the robustness of the instructional strategy by demonstrating its ability to survive beyond the context of its creation. Like the previous study, this study also made use of treatment and control groups (N =30). Again the treatment group followed the dynamic WCF instructional strategy and the contrast group received traditional writing instruction with the results further ratifying those of the previous studies (Evans et al., 2011).

Surprisingly enough, the control group of this study again saw a decrease in accuracy (operationalized in this study as error-free clause ratio) while the treatment group saw a significant gain in writing accuracy. The researchers offered an interesting hypothesis for the

decrease in the performance of the contrast group in terms of accuracy, suggesting that the treatment group had a more narrow focus in instructional strategy that strongly favored accuracy. Although the contrast group received accuracy feedback as well, the authors suggested that the ability of the learners in the contrast group to attend to accuracy was hindered by limited attentional resources for other writing skill needs including an abundance of rhetorical considerations, organization and structural decisions, and content concerns (Evans et al., 2011).

A new home for dynamic WCF instructional strategies.

To their credit, the originators of the instructional strategy have welcomed continued review of the practice and they themselves have shown a measured and rational approach to applying the instructional strategy in their own spheres of practice. With each step or extension of the instructional strategy, there has been a solid attempt to justify its use by quantifying the effect when the instructional strategy is trialed in new contexts. One of the most recently completed studies, a MA thesis by Soonyeon Lee, has had some prominent influence in directing the current application of the instructional strategy in the curriculum of the Academic program at the English Language Center (Lee, 2009).

While the previously published studies looked at the instructional strategy in terms of contrast against traditional writing pedagogy and its effect on writing accuracy, Lee (2009) contrasted accuracy gains of students in an adaption of dynamic WCF that was curricularized into a Linguistic Accuracy course relative to gains made by students enrolled in a traditional grammar skills class. The purpose of this research was examining the possibility of replacing a traditional grammar skills class with a course consisting dominantly of the dynamic corrective feedback instructional strategy. In addition to the writing oriented Linguistic Accuracy course, students in both of these study groups also received traditional process writing instruction. Lee

also extended the instructional strategy into lower proficiencies than had previously been tested with her study participants approximating intermediate high. In addition to the quantitative analysis on the impact on accuracy attainment, Lee examined students' preference for the instructional strategy relative to the traditional grammar curriculum. While the change in linguistic accuracy of the treatment group of Lee's study was not statistically more significant than the gains made by the control group, which received traditional instruction, there was a slightly greater gain in overall accuracy favoring the practice of dynamic WCF (2009).

As was mentioned, a notable aspect to this research was the qualitative measurement of students' preferences for dynamic written corrective feedback over more traditional grammar instruction. Student's responses revealed a strong preference for dynamic WCF held by students in contrast to preference for traditional instruction (Lee, 2009). These results validated the direction taken by the ELC in beginning to introduce dynamic corrective feedback across a wider range of proficiencies Because of the rigors on both teacher and student when fully engaging in this instructional strategy, student perceptions of its value further justified the curricular space and resources given the instructional strategy at the ELC.

Current curriculum implications of dynamic WCF instructional strategy.

All four of these studies have informed the development of the linguistic accuracy curriculum currently utilized by the Academic program of the English Language and the students are benefiting from the focused attention to writing accuracy. However previous research has only concluded that the instructional strategy positively affects writing accuracy (Evans et al., 2010; Evans et al., 2011; Hartshorn et al., 2010). Writing accurately is one side of a two-sided need for accurate language production for students enrolled in academic preparatory intensive

English programs. The other side of similar significance that should be addressed by such language programs is the production of accurate speech in a variety of communicative settings.

Limiting the Linguistic Accuracy curriculum to a single practice modality, writing, is done with the hope that accuracy gains made through dynamic WCF will spill over from the context of writing into the practice of speaking. Also students receive some accuracy instruction in their speaking skills class. In unpublished research, Hartshorn has noted that there also has been some indication that a particular dosage of the instructional strategy in either modality must be applied before impact on accuracy performance is achieved (personal communication, May 3, 2011). Thus research should show that the quantity of practice in the written mode can be reduced to accommodate speaking practice without impacting the established benefits of the treatment before changes to redirect practice and feedback towards speaking accuracy occur in the established pedagogical application of the instructional strategy. Addressing the first of these two hesitations towards adopting a dual practice model may be best accomplished by returning to the theoretical foundation of dynamic written corrective feedback: skill acquisition theory.

An argument could be made that practice modality makes little difference in the first stage of acquiring declarative knowledge. Consequently feedback received on writing could contribute equally well to the acquisition of declarative knowledge related to speaking accurately. However, practice modality does become a critical component in the second stage of skill acquisition theory: proceduralization (Anderson, 1983). As Hartshorn et al. (2010) emphasized: "The theory predicts that accuracy is a function of practice and…that procedural knowledge does not transfer well. Thus, if students are to learn to produce accurate writing, practice tasks and activities must be authentic" (p. 87). Adapting this conclusion to speaking, if students are to produce accurate speech, practice tasks and activities must be authentic to

speaking. Skill acquisition theory limits the believability that practice in writing accurately alone could lead to the proceduralization of the cognitive and linguistic skills necessary for accuracy in speech. Simply put, practice modality matters.

The second assumption that accuracy instruction received in the context of a listening/speaking class will be sufficient can be refuted by the findings of two of the three studies that showed students in a traditional skills class experienced a decrease of general accuracy even when accuracy was one of the objectives in the course curriculum (Evans et al., 2011; Hartshorn et al., 2010). The researchers themselves hypothesized that this could be the result of an overextension of a learner's attention resources (Evans et al., 2011; Hartshorn et al., 2010). Speaking pedagogy, in a traditional communicative class structure similar to the ELC Listening/Speaking curriculum, results in a similar division of these resources between rhetorical considerations, organization and structural decisions, and content concerns. Consequently this fracturing of attention could lead to a similar inability for students to really digest and apply accuracy-focused instruction and feedback received in this type of course.

Recommendations for a Shift Towards a Dual Skill Practice Modality

The assumptions promoted by the creators of dynamic WCF give adequate rationale for the conclusion: 1) Students desire to improve their linguistic accuracy (both in writing and speaking); 2) Students expect to receive error feedback (on both written and spoken production; 3) Students can improve their linguistic accuracy with appropriate error correction and 4) Error correction can be consequential when it is manageable, meaningful, timely and constant (Evans et al., 2011).

While there were some modifications to the dynamic WCF model that were needed to establish a sister instructional strategy that was responsive to a student's spoken production,

these changes seemed minor and manageable. It was assumed that the instructional strategy was ready to be manipulated in order to examine the benefits of including a dual skill practice mode into the dynamic corrective feedback framework. The robustness of the instructional strategy established through the research based manipulations mentioned above suggested that the benefits of dynamic corrective feedback could be maintained and perhaps even enhanced by the addition of speech accuracy practice.

Justification of Using Speech Transcriptions as Student Input for Dynamic CF Model

The primary variation that would have to occur to include a spoken accuracy component to the linguistic accuracy curriculum is transferring spoken production into a form that could enter the dynamic corrective feedback drafting cycle. Obviously there are few if any alternatives, other than transcribing students' speech. These transcriptions could then be marked with the indirect marking system and initiate the corrective exchange between students and their Linguistic Accuracy teacher. This section will present some of the most current literature that supports this adaptation and suggests that having students transcribe part or all of their speech sample, there could bring benefits for students accuracy including facilitating their own grammatical awareness and benefiting complexity and fluency as well (Lynch 2001; Lynch, 2007; Stillwell, Curabba, Alexander, Kidd, Kim, Stone, & Wyle, 2010; Sheppard, 2011).

One of the more prominent researchers publishing studies on the benefits of transcription activities in the English language classroom is Lynch (2001; 2007), who noticed anecdotal benefits to students in a communicative skills class that resulted from transcribing their speech. Students then engaged in self, peer, and instructor corrective dialogue. Lynch conducted research with both quantitative and qualitative components that can inform the adaptation of dynamic corrective feedback proposed in this study. Lynch highlighted the need for learners to notice,

quoting Batstone (1996) who observed "the intake of grammar is a result of learners paying conscious attention to input" (p.125). He went on to express that there is no reason as long as students have adequate proficiency and metalinquistic awareness that this input could not be their own output, particularly when refined through self-correction, peer-correction, and finally instructor correction (Lynch, 2001; Swain, 1995; Swain & Lapkin, 1998). Other researchers have established benefits to students in transcribing other's speech in the forms of dictations, dictoglosses, and even transcribing interviews (Clennell, 1999; Johnson, 1996). It stands to reason that if there is benefit through general attention to the output of others there could be added benefit to specific attention to one's own output.

Lynch's (2001; 2007) studies had students engage in an unscripted conflict of interest role-play with a partner while being recorded. Both partners then transcribed a 90 to 120-second portion of the dialogue generated in this recorded performance. Once they reached agreement on the accuracy of the transcript to the original recording, the dialogue-partners began editing and revising the dialogue until they reached a level of general satisfaction with the English. At this point, they typed the corrected transcription and submitted it along with the original transcription to the instructor. The instructor, using the original and corrected transcripts, produced an instructor corrected version before the next class. Students were asked to compare the second (their own revisions) and third (the instructor's revisions) drafts and notice the differences between the two (Lynch, 2001; 2007).

Lynch's study (2007) showed that students were not bored or frustrated by the process of transcription, a conclusion corroborated by several additional studies (Stillwell et al., 2010; Sheppard, 2011). His research also showed that while students were able to notice and self-correct many of the mistakes in their transcription, teacher intervention was needed to elevate

awareness of particular types of errors, under-addressed when self-correcting, including lexical errors and formally correct phrases that could be more efficiently or appropriately expressed. Finally this pedagogical application showed that error-focused feedback can sequence harmoniously within the development of a communicative activity, which has also been established by others (Doughty & Varela, 1998; Labkin & Swain, 1996;Lightbown, 1991; Lyster, 1994, 1998). The post-scenario transcription allowed for students to critically engage with their language production in a way that was far less inhibiting or disruptive than direct correction would have been.

Further research by Lynch (2007) have shown that a larger scale application of this type of activity maintains the benefits established in the first while still being feasible and manageable for students and instructors. Lynch has also shown that students who engage in transcription of their own speech retain higher levels of accuracy related to forms targeted through feedback than do students who have the transcription of their speech done by the instructor (Lynch, 2007; Mennim, 2003). This conclusion has implications for the research initiated here that justify the use of student-produced transcripts allowing for the instructor's focus to remain on providing timely, meaningful, constant, and frequent feedback.

Lynch's study has inspired others not only to integrate transcription exercises into their classroom but then also attempt to quantify the effect of such activities (Stillwell et al., 2010; Sheppard, 2011). Prominent conclusions to one of these are worth noting here. Stillwill et al. (2010) noticed that in recycled activities that followed transcription and self, peer, and teacher feedback, students were more likely to attempt to integrate teacher correction than self- or peer-correction into the second attempt, although the researchers noted that only about 55 percent of these attempts were successful. The types of self-corrections where students were most

successful included speech editing for dysfluencies including false starts, pauses, and fillers. Lexical errors was an area where the students relied most heavily on teacher feedback.

Finally, students' perceptions of the usefulness of the activity were surveyed. Students conclusively felt that, while all of the parts of the process were useful to a degree, the most useful elements included receiving teacher corrections, engaging in self-correction, and recycling the speaking activity. All of these elements could benefit students involved in an adaptation of dynamic corrective feedback that utilizes speech transcriptions (Stillwell et al., 2010). Table 2 summarizes the findings of this survey.

Table 2

Students' Perceptions of Transcription Task Usefulness

Elements of Transcription Exercise	Useless				Useful
	1	2	3	4	5
Transcribing own speech	—	—	.04	.52	.44
Transcribing partner's speech	—	.08	.48	.28	.16
Correcting own mistakes	—	—	.04	.12	.84
Correcting partner's mistakes	—	.04	.32	.32	.32
Receiving teacher corrections	—	—	—	.08	.92
Repeating activity a second time	—	—	.12	.24	.64
Filling out usefulness survey	.04	—	.28	.36	.32

Proposed Research on the Effects of a Modified Form of Dynamic CF

The evolution of dynamic corrective feedback instructional strategy has been justified by quantitative and qualitative measurements with each new adaptation and application. Integrating a mode of speaking practice into the dynamic corrective feedback instructional strategy will need to be similarly validated. In addition to showing that such an adaptation adds value to the curriculum in advancing students' spoken production, the adaptation needs to show that any such modification maintains the proven benefits of the current written production modality in benefiting students' writing across skill components. In doing so, an adaptation of the

instructional strategy must be true to the tenets of dynamic written corrective feedback: that of practice being sustainably frequent and authentic, resulting in feedback that is student specific while remaining meaningful, timely, constant, and manageable.

Informing decisions about implementing a dual practice mode approach is the summative goal of this research agenda. However in order to best understand any significant difference in outcome that is dependent on practice mode, it was necessary to first view the two modes, writing and speaking, in isolation. Consequently, the following study consisting of two sets of students enrolled in the ELC's Academic program and participating in the Linguistic Accuracy curriculum was devised. One half of the students received a treatment of the existing dynamic corrective feedback instructional strategy that used student written ten-minute paragraphs. The other half of the students in the study would receive a modified treatment that substituted spoken transcripts for the ten-minute paragraphs in informing the corrective dialogue. The variability of gains for each of the groups between the productive skills of speaking and writing, considering the subskill characteristics of fluency, complexity and accuracy, would then be analyzed.

The research objectives of this study can be summarized as follows:

When students in a Linguistic Accuracy class where the individual student output mode that informs dynamic corrective feedback is speaking are compared to students in a linguistic accuracy class where the individual student output mode that informs dynamic corrective feedback is writing:

1. Can the dynamic corrective feedback instructional strategy be altered to responsive to students' speech?

2. Are there statistically significant differences in written accuracy gains between the two groups that result from differences in the practice mode that receives feedback?
3. Are there other statistically significant differences including gains or losses in written fluency, complexity, and lexical development that result from the practice mode that receives feedback?

The study conducted and described within this thesis consisted of elements that addressed these questions. These objectives intended to show how the benefits of the adapted instructional strategy compare to the established benefits of the traditional instructional strategy as it affects students' written English proficiency.

Chapter 3 Methods

The purpose of this chapter is to describe the research method used to answer the questions described previously including:

1. Can the dynamic corrective feedback instructional strategy be altered to responsive to students' speech?
2. Are there statistically significant differences in written accuracy between the two groups that result from the differences in the practice mode that receives feedback?
3. Are there other statistically significant differences including gains or losses in written fluency, complexity, and lexical development that result from differences in the practice mode that receives feedback?

This chapter will provide a description of the participants in the study including: students who received instruction, practice opportunities, and assessments, and instructors who guided the practice exercises, responded to student production, and gave accuracy instruction. Also, the assignment of participants to the two treatment groups will be described. This chapter will then provide a rationale of the research design employed including a description of the instruments used to elicit student production. Finally, the chapter provides an analysis structure of student performance including: operationalized versions of the research questions; description of raters of pre- and post- test writing samples; and those steps taken to establish and maintain reliability of the results gathered to answer the research question will be presented.

Participants

Students

The treatment and contrast groups consisted of students in four of the five sections of Linguistic Accuracy courses in the Academic program at Brigham Young University's English

Language Center, an intensive English program in Provo, UT, USA. Two of the sections consisted of students in Academic A, the entry level of the Academic program, and the other two sections consisted of students in Academic B, the middle of three levels in the Academic program. Using the standards established by the American Council of Foreign Language Teachers, the English proficiency for most of these students was estimated to be between intermediate mid to advanced low during the 15 week Fall semester of 2010. If a participant was a returning student, they were placed in these levels of the Academic program based on their performance in the previous semester of study and on an end of semester assessment. New students were placed in their respective level by their performance on a placement test that is very similar to the end of semester tests in content, rigor, and structure.

Both returning and new students were placed in sections of their respective level randomly. The following process was used twice: once for the students who had been placed in Academic A and then again for students in Academic B. All of the ID numbers of students placed in a level were entered into a list randomizer. After the list was randomized, each student was assigned a number that corresponded with their rank order generated by the randomizer. Students with an odd number were placed in the first section; students with an even number were placed in the second section.

For each level of Academic A and Academic B, three sets of two sections were generated randomly. For each of these randomized sets of students, the distribution of L1, age, gender, and number of semesters studied at the ELC was then determined. Of the three sets generated, the set with the most balanced sections in terms of these variables was selected to be the section assignments for this semester. This balancing of the randomization was important because the sections assigned for the study would form the intact classes that would rotate together through

three additional classes during a day of instruction at the ELC. The sections of Academic A and the two sections of Academic B were then randomly assigned to participate in either the contrast group or the treatment group.

The breakdown of notable characteristics of each of these groups is described in Table 3 including native language and gender. The age span of the students in the sections of the contrast group was 18-34 (A) and 17-38 (B) with mean ages of 26.38 and 23.25 respectively. The age span of the students in the sections of the treatment group was 19-44 (A) and 18-29 (B) with mean averages of 23.16 and 23.54 respectively. The randomization of students between the sections, improved on previous similar studies of the instructional strategy that used intact classes (Evans et al., 2011; Hartshorn et al., 2010; Lee, 2009). Although these previous studies indicated that these demographic characteristics, such as gender and L1 background seemed to be of negligible effect, they also made the recommendation that future research randomly assign participants where possible (Hartshorn et al., 2010; Lee, 2009).

Table 3

Experimental Group Participants by Native Language and Gender

	Study Groups					
	Contrast Group			Treatment Group		
Native Language	Male	Female	Total	Male	Female	Total
Spanish	3	9	11	6	4	11
Korean	2	-	2	1	4	5
Portuguese	3	2	5	2	3	5
Japanese	1	1	2	1	2	3
French	-	-	0	-	1	1
Vietnamese	-	1	1	-	-	-
Russian	-	1	1	1	1	2
Mandarin	-	2	2		1	1
Ukrainian	-	-	0	-	1	1
Totals	9	16	25	11	17	28

Instructors of Study Groups

While the students were randomly assigned to two groups of the study, teachers had already been assigned to the levels and skill areas of the Academic program when this research initiated. Because of the rigor of the instructional strategy, the administration made careful consideration related to the staffing of this particular course and also compensated the instructors for the added time required to provide the intensive corrective feedback. During the semester of the study, a different instructor was assigned to each of the four sections. Two of the teachers had taught the dynamic corrective feedback instructional strategy in the Academic program previously and two had not. Because the contrast group instructional strategy was more closely aligned to the teaching expectations of their previous experience, the instructors who had taught in Linguistic Accuracy before were assigned sections of the contrast group. The teachers of the treatment group were also highly competent; however, this was their first semester teaching Linguistic Accuracy. This was seen as a potential advantage. The treatment variation of the instructional strategy was the only version with which they were familiar; consequently, this allowed them to remain consistent with the unique aspects of the treatment. The experience of these teaching professionals is outlined in Table 4.

Table 4

Experimental Group Participants by Treatment, Level, Teacher and Teacher's Experience

Group	Level	Teacher	Experience Level	Semesters of DCF Experience	Number of Students
Contrast	Academic A	W	Experienced	3	12
	Academic B	X	Novice	2	13
		Total			25
Treatment	Academic A	Y	Experienced	1	14
	Academic B	Z	Novice	1	14
		Total			28

Research Design

As was made evident in Chapter 2, this study intended to build upon the previously completed research into the effect of dynamic corrective feedback. As a result, the research design was modeled after the design employed by previous researchers of the traditional instructional strategy. While previous studies on DCF had used a pre-test, post-test nonequivalent control group design because they employed intact classes, this study consisted of randomly assigned groups and, consequently, met the key expectations for a randomized control-group pre-test, post-test design. This research design is summarized in Table 5.

Table 5

Pretest, Posttest Equivalent Control Group Design

Experimental Group	Pretest	Treatment	Posttest
Treatment (28 Students)	O_1	X_S	O_2
Contrast (25 Students)	O_1	X_W	O_2

Note: O = Testing Occasion, X= Experimental Treatment $_{S = \text{Speaking Emphasis W=Writing Emphasis}}$

While the study's research design varied slightly from previous studies, it employed a similar analysis of students' performance by using a mixed model, repeated measure Analysis of Variance (ANOVA). This measurement compared the mean performance of the students in the contrast group with the mean performance of the students in the treatment group, and the mean performance of the students before the treatment began to the mean performance of the students after the treatment was completed.

The two 30 minute essays produced by the students during the pre- and post-tests were analyzed in several different ways in order to answer the research questions. These analyses included a measurement of the student participants' written complexity, fluency, and accuracy. Also each writing sample's lexical content was analyzed in two ways to determine the potential

effect of either form of dynamic CF on lexical development. To complete the mixed model ANOVA necessary to contrast the performance of the treatment and contrast group, the Statistical Package for the Social Sciences (SPSS) was used. Each of these variables was evaluated within levels of statistical significance well established for this type of research and the standard used by previous research (p=.05). The factor used to contrast each group's performance before and after the treatment was labeled "time" and included a level for both pre- and post-test performance. The factor used to contrast between the two groups was labeled "group" and also had two levels: contrast and treatment.

Writing samples were assessed for (1) accuracy, (2) fluency, (3) complexity, and (3) lexical development in order to address the research questions of the study. These variables and the method of measurement used in their analyses are listed in Table 6; each will be described in depth in the following section. Also, the reliability procedures for each step of analysis will be described.

Table 6

Dependent Variables and Their Methods of Measurement

Dependent Variables	Method of Measurement
1. Accuracy of writing	(error-free clauses/total clauses)
2. Complexity of writing	(number of words/number of T-units)
3. Fluency of writing	(number of words written in thirty minutes)
4. Lexical Development	(number of types/number of tokens) (number of words derived from the Academic Word List word families/number of total words)

Fluency

The in-house computer software used during the pre- and post-test collection of writing samples was programmed to tag each writing sample with the number of words they contained. As noted in Table 6, writing fluency was set at the number of words in the writing sample. Because time was held constant for each student (30 minutes) differences in word count can be

used as a measurement of rate and consequently fluency. The reliability of word counts should be considered high.

Lexical development

With a similarly high reliability level, the writing samples were analyzed for lexical development by a software application called AntWord Profiler developed by Anthony (2009) at the school of Science and Engineering at Waseda University, Japan. This software analysis is comparable to the Range software created by Nation (1994) but runs on the Mac OS X operating system. AntWord Profiler analyzed each writing sample, calculating the number of types (total number of discrete words used) and tokens (the total number of words in the text). AntWord Profiler also calculated the percentage of words in the entire sample that are derived from any of the word families of the ten sublists of the Academic Word list. These percentages, along with the type token ratio, were calculated by the software and included in the analysis of this variable. As this data was calculated by the computer program, again high levels of reliability can be assumed.

Complexity

Analysis of writing complexity was not afforded the luxury of computer calculation and consequently was subject to the potential for human error. However, as will be discussed, several steps were taken to ensure the reliability of the data used to calculate the complexity of a student's writing. In order to perform the analysis described for complexity, the writing samples needed to be broken down into clauses and then T-units; the latter being defined as an independent clause and any subordinate clauses attached to it (Hunt, 1965). Every researcher who has examined DCF has used one or both of these two structural units in their analysis. For the analysis of writing complexity, the data analysis included the same measure as Hartshorn

(2008; Hartshorn et al., 2010), defining complexity as mean length of T-unit. An example of a writing sample that has been broken down into clauses and T-units is included in Appendix A.

In order to maintain a high level of reliability in breaking the writing samples into these structural units, the principal investigator of the study used the following process. First, the main researcher calibrated with a highly-experienced rater that has been part of several similar studies. Relying on a rubric for defining a clause developed by Lee (2009), the main researcher and the highly-experienced rater each broke down the same six sample essays. The results were then compared and the two raters negotiated any discrepancies until they reached a high level of consistency (Pearson correlation coefficient of .97). Because the breakdown process spanned several days, the primary researcher then randomized the list of all the tagged writing samples and this random ordering was the order in which all of the writing samples were analyzed. After the principal investigator had completed this task, a random sampling of 10% of all the essays was taken and independently broken down by the experienced rater. A Pearson correlation coefficient of the same breakdowns by both the main researcher and the experienced rater showed a .98 agreement on the number of clauses and .96 agreement on the number of T-units as indicated in Table 7 in the next section.

Accuracy

Before the writing samples could be assessed in terms of writing accuracy, the samples were broken into clauses and T-units as described previously. When the consistency of this part of the process had been verified as described above, the essays were then evaluated in terms of error-free clause ratios. Each clause contained in an essay's breakdown sheet was determined as either errored or error-free. Error-free clauses were highlighted and counted. The number of error

free clauses was then divided by the total number of clauses contained in the sample to calculate the error-free clause ratio.

Again the researcher followed a specific pattern in evaluating the accuracy of the writing samples produced by students in each group. Following the evaluation of error-free clauses rubric developed previously by Lee (2009), the main researcher calibrated with the same highly experienced rater on six randomly sampled essays. After a sufficiently high level of confidence was obtained as reported in Table 7, the main researcher identified the error-free clause ratio for the remaining essays in an order that was randomly assigned.

Table 7

Pearson Correlation Coefficients for Each Set of Ratings

Rating Types	Scorers	*R*
Total Clauses	Scorer 1 vs. Scorer 2	.99
Total T-units	Scorer 1 vs. Scorer 2	.96
Error-free Clauses (1st Round)	Scorer 1 vs. Scorer 2	.95
Error-free Clauses (3rd Round)	Scorer 1 vs. Scorer 2	.98

In order to add an extra level of reliability to the accuracy analysis, each sample was blind rated by second raters. The second raters were all TESOL professionals with extensive teaching experience and experience in teaching the dynamic corrective feedback instructional strategy. Their experience is further described in Table 8.

Table 8

Second Rater Teaching Experience & Dynamic Corrective Feedback Experience

Rater	Graduate degree in TESOL or related study	Years of ESL Instruction Experience	DCF Experience.
Rater 1	*Yes*	20	*Yes*
Rater 2	*Yes*	6	*Yes*
Rater 3	*Yes*	2	*Yes*
Rater 4	*Yes*	10	*Yes*
Rater 5	*Yes*	10	*Yes*

The essays were randomly divided between the raters. Using their experience and the rubric described above when needed, the raters made error judgments for each of the writing samples that had been broken down into clauses. The judgments made by the main researcher and the second raters were compared. Clauses without agreement between the two were then subjected to a third review. After this review process, the error-free clause ratio was calculated. As with the complexity measurements, a ten percent sampling of all of the essays was randomly selected and verified by the highly-experienced rater. A Pearson correlation coefficient of the error-free clause ratios by both the main researcher and the experienced rater showed a .98 agreement as reported in Table 7.

Instrument

This section will describe the instrument that was used to elicit written samples on both the pretest and the posttest. Similar to most of the studies done on dynamic written corrective feedback, a 30-minute essay test was used to measure gains in students' abilities to write accurately, fluently, and with a degree of complexity and also to note changes in their lexical advancement.

As was mentioned previously, before both variations of dynamic corrective feedback instructional strategy were initiated and after the treatment period had ended, students from both

groups took an in-house developed computerized test. The computerized test consisted of a 30-minute essay and two 3-minute spoken response questions that will be used in related research but are not addressed here. For the writing portion of the exam, the software allowed students to copy, cut and paste sections of text from their written response, but other than these basic functions, there were no formatting or spell-check features available during the students' practice and assessment. During the pre- and post-test, students were not allowed to write for longer than 30 minutes. After 30 minutes had passed, the program saved students' results and exited the program. Table 10 in the section on the elicitation tool presents the writing prompts inserted into the computerized test for both the pretest and the posttest.

Instructional methods

This section will outline the instructional methods that guided student learning in both the treatment and contrast groups between the pre- and post-tests. The section will first discuss the instructional and learning tasks that were the same for both groups, and then the adaptation of the existing form of dynamic corrective feedback instructional strategy practiced by the treatment group will be addressed. An outline of the instructional sequence that will be described in detail is contained in Table 9.

Table 9

Outline of Instruction Hour Monday – Wednesday

Time	Treatment Group	Contrast group	Time
5 minutes	Announcements, Roll, Class Overview		5 minutes
15 minutes	Written and Spoken Production Practice		15 minutes
15 minutes	Self-transcription		45 minutes
30 minutes	Error Responsive Instruction		

While the instructional strategy was adapted to include a daily oral skills emphasis, preserving the tenets of the dynamic CF instructional strategy was essential. These tenets of the instructional strategy require that feedback be manageable for both the instructor and the students, and that feedback be frequent, timely, and constant. Also highly important, feedback should be meaningful in that it engages students' cognitive awareness of their own errors and leads them toward self-discovery and use of the correct form all while the productive activities that initiate this feedback cycle be authentic and reflect actual academic and communicative tasks.

The ELC curriculum consists of four classes that meet Monday through Thursday with each class lasting 65 minutes. Students in a particular section of a particular level study with the same group of students for the four skills classes of the day. Of these four classes, one is the linguistic accuracy course that was included in the study. Since the effects of dynamic CF have been substantiated, instructors of this course have utilized dynamic CF instructional strategy as the core of their class syllabus. It was important that practice opportunities were as equal as possible for students irrespective of teacher, class, and even treatment condition. Thus the courses were structured in a way that the only significant difference in the scope or sequence of the course was the contrast or treatment form of dynamic CF.

On Monday through Wednesday during the treatment period, linguistic accuracy classes began in the computer lab. The production activities that began class were collected through in-house computer software created specifically for this study. These production activities included

a written response and two oral responses that were elicited through two different prompts. Each day the program was updated with the day's prompts which were the same for each section of the study. These prompts elicited an authentic sample of students' oral and written production. Standardizing the prompts for the sections enabled confidence that the productive opportunities for the study sections were equivalent in terms of difficulty and lexical rigor.

While some previous courses of the curriculum had used computers to write the initial draft of their ten minute paragraph, this was the first semester that all of the classes were required to use the computers for the initial production exercises. Requiring computer use for the production exercises was part of the control for practice effect; students from both treatments needed to have relatively the same practice opportunities. The research questions could be most definitively answered if the only variation between the practice opportunities of the two groups was in the type of production, oral or written, that was used to initiate the dynamic CF process.

The computer program used for the daily productive activities was designed to minimize variations in this stage of the process that could introduce confounding variables into the gains measured in the post-test. Students in both sections had a practice task of both writing (a 10 minute paragraph) and speaking (two 2 minute oral responses). Although students in the contrast group recorded oral responses during the production phase, they did not access these recordings nor did they receive corrective feedback on this oral production. Likewise, students in the treatment group typed 10 minute essay responses; however, they did not see their written work later nor did their teachers' give them feedback on this task. This control would allow interpretation of the results to conclude that differences in gains in a particular skill were the result of where the two variations were different—that is, mode of output receiving feedback—and not differences in practice conditions.

While the steps between production and feedback varied slightly as will be discussed shortly, the principles governing feedback were the same; students were moved toward more accurate production through receiving prompt indirect feedback from the instructor. This feedback occurred between the end of a day's instruction and the following class period. Instructors collected the production sample, which consisted of either a typed 10 minute paragraph or a transcribed oral response. The production sample was marked using a codified set of indirect error identification markings that would identify both error location and error type. These marks are included in Figure 1.

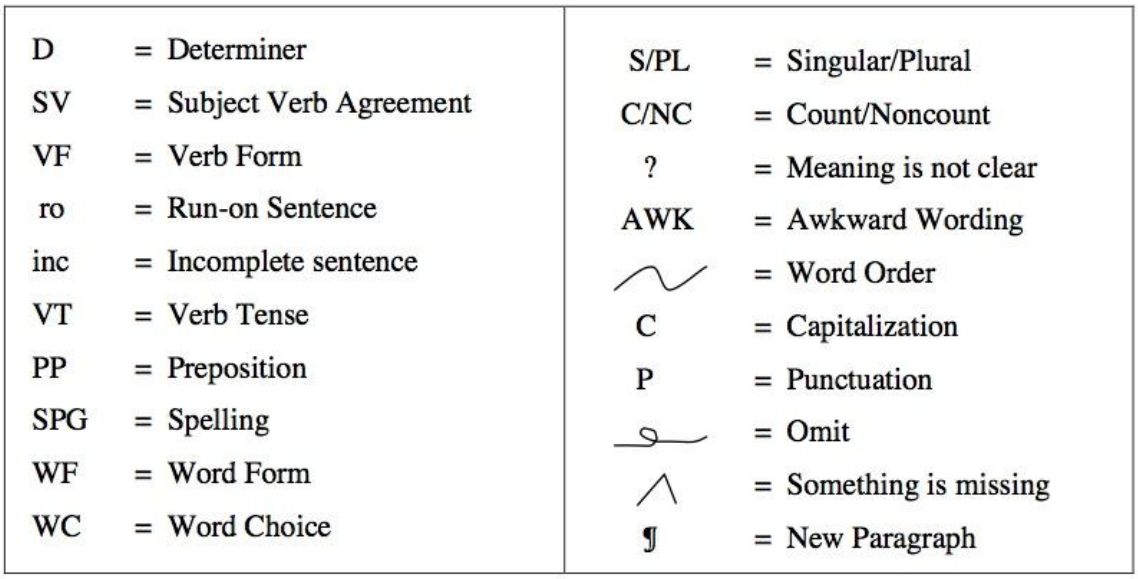

Figure 1. Indirect error coding symbols used to mark ten minute paragraphs and speech transcripts (Hartshorn et al., 2010, p. 107).

The following class period the teachers returned the marked drafts or transcripts to the students. The students were responsible for making the corrections necessary to bring the sample to a near error-free state. To do this, students relied on the error list highlighted in previous dynamic CF studies to help them (Lee, 2009, Appendix F). Students were allowed as many

submissions as needed to correct their draft up until the deadline of one full week after the initial production. If errors persisted or if students introduced new errors in their attempt to correct the initial errors, the sample was again marked and returned to the students the following day.

This drafting process was initiated daily Monday through Wednesday. On any given day students would resubmit and receive a number of drafts back from the teacher. Some of these drafts would be removed from the drafting cycle as they reached an error-free state; others would be cycled several more times. The productive activity required approximately 15 minutes of participation. After the productive activity was finished, students in the treatment group would have a step that was unique to their form of the process before returning to the classroom for the day's instruction. Students in the control group left the lab after the productive activities to return to the classroom for regular instruction.

Also similar to both the contrast group and the treatment group was the use of a text series, *Grammar Dimensions* (Larsen-Freeman & Thewlis, 2007). Teachers also used student-generated errored sentences from the previous day's ten-minute paragraphs or speech transcriptions to inform the instructional aspect of the class. In order to keep the course centered on students' accuracy needs, teachers had the autonomy to address those errors that were most prominent in the previous day's writing samples or transcriptions and were free to use the text to supplement, enhance or reinforce the feedback received through the drafting process.

Students in both the treatment group and the contrast group kept updated error tally sheets where they marked the number or incidence of a particular error type in the corrected written paragraph or speech transcriptions they received. The students would also log errored sentences and the corrections necessary to reach an acceptable correction in an error log. These two forms, error tally sheets and error logs were submitted and checked intermittently by the

instructor. These two forms of tracking mirrored those used in previous applications of dynamic CF (Lee, 2009, Appendix H & I).

As was mentioned above, the sequence of the activities had little variation from Monday to Wednesday; for Thursday instruction, the schedule did not include a mandatory production activity and most often the instructor would use the day to address material that had not fit into the week's schedule. The teacher could also spend the instructional hour doing grammar activities or assessments from the text and allowing both students and the teacher to get caught up on work in the drafting cycle. The only stipulation mandated by the study for Thursday's schedule was that the sections remain solely responsive to the production mode being targeted: writing for the contrast group and speaking for the treatment group.

All of the students in the Linguistic Accuracy courses are required to take five 30-minute essay tests throughout the semester. The pre-tests and post-tests of the study counted for the first and last test of the semester. The three interim tests were identical to the pre-tests and post-tests in format though the prompts for each were different. For these interim tests, there was some variation in the type and degree of feedback instructors provided for the students; again, the only restriction prescribed by the study was that feedback be limited to the production mode emphasis of their particular treatment.

The contrast group of the study followed the established written dynamic CF instructional strategy with a few variations noted above, i.e., three weekly production tasks (one less the number of days in class per week) and two spoken responses were elicited but not incorporated into the drafting process. While students in the contrast group had the same opportunity for spoken English practice, they did not receive any feedback on the practice they completed. The treatment group of the study followed a modification of the established dynamic CF instructional

strategy using the same program and the same prompts as the contrast group, meaning they typed the ten-minute paragraph and recorded two minutes of spoken responses. After the production activity was completed but before leaving the lab, the students in the treatment section were given 15-20 minutes to transcribe their oral response recordings. These response transcripts were used in place of the ten-minute paragraphs for feedback.

In transcribing their speech, students in the treatment sections were instructed to initially include in their transcript all prominent features in their recorded sample. Consequently, as they were transcribing the meaningful items of their recording, students also indicated pauses, non-lexical fillers, and false starts. Students were also instructed to include errored speech as they spoke it even if while transcribing they recognized the error and understood how to fix it. Before submitting the transcript, the students would count the number of false starts, non-lexical fillers and pauses that the speech sample contained and remove them from the transcript. Students in the treatment group tracked these instances of dysfluency on their error tally sheets. Indicating these dysfluencies as an error type was the single difference between the two groups in their use of tally sheets and error logs.

Transcription does require a certain degree of practice. While some of the students had previously transcribed their own speech, other students were new to the process. As a result, the treatment sections were initially only required to transcribe half of each of the two oral responses (approximately two minutes combined). Gradually, the amount of the speech sample that needed to be transcribed and submitted to the drafting process was increased. After the first few weeks, the volume of language entering the drafting cycle for both groups was nearly identical (i.e., the number of words, between the ten-minute paragraphs and transcribed speech samples from both recordings was approximately the same).

Because of the added time to transcribe, sections in the treatment group received 15 – 20 fewer minutes of class instruction time. Instructors and students in both the treatment and contrast group expressed similar concerns desiring more class time—a concern that will be addressed further in Chapter 5. The loss of instructional time for transcription was mitigated by not requiring the initiation of a draft on Thursday for either group, allowing the classes to spend the entire time on instruction or other class work.

Elicitation Procedures

As mentioned previously, the students in the control group and in the treatment group took the pretest before beginning any dynamic CF instruction or the drafting process at the beginning of fall semester 2010. This test occurred the second day of the semester during the regularly scheduled class period. The instructors took the students to the computer lab where the test was administered. The main researcher and the course instructor proctored the test to ensure that it was administered under secure conditions. The pre-test writing task was to write a 30-minute essay in response to a prompt.

These same students took the 30-minute post-test at the end of the treatment period, one week before the end of the same semester. The post-test was also administered during the regular instruction hour and marked the end of any formal dynamic CF instruction or required drafting process assignments. Again the main researcher and the course instructor proctored the test to ensure that it was administered under secure conditions. Table 10 presents the pretest and post-test prompts.

Table 10

Prompts Used to Elicit Written Production Before and After Treatment

Pretest	When people move to another country, some of them decide to follow the customs of the new country. Others prefer to keep their own customs. Compare these two choices. Which one do you prefer? Support your answer with specific details.
Posttest	Every generation of people is different in important ways. How is your generation different from your parents' generation? Use specific reasons and examples to explain your answer.

For both examinations, students typed their responses to these prompts into an in-house developed computer program. Students accessed the exam with their nine-digit BYU identification number which tagged all of their results with their name and time of completion. The test then included a sound check for the oral portion of the exam not included in this study. Once the sound check was completed, the 30-minute written response task began as the prompt appeared at the top of the page and the timer at the bottom of the page began to countdown. Next to the timer was a button that the student could press to check their total word count. The student received a ten minute and five minute warning from the software as they typed. Once the timer ran out students could not type any more, nor could they move on before the timer had ended, and the test moved to the spoken response questions.

These results were then saved and catalogued following the same procedure that the ELC follows for all examinations administered using in-house developed software. For the purpose of the study, a copy of the student performance files was moved from this storage location to another secure location where they could be analyzed. At this point the samples were stripped of any names or identifying features and tagged with a study specific ID tag. The ID tag consisted of a randomly generated string of 5 digits and a letter identifying the sample as coming from either the pretest or posttest. The researcher kept a key that identified the 5 digit string belonging to each student but it was not accessed until the final analysis of the data. The files were kept in

folders with the same ID number system and the database was then sorted by ascending number. As the number for each sample was randomly assigned, rank ordering by ascending number randomly mixed samples from the contrast group and treatment group.

Restatement of Research Questions

With this additional background the primary research study questions can now be restated and operationalized.

1. Can the dynamic CF instructional strategy be altered to responsive to students' speech?

   Operationalized: Will a linguistic accuracy course where the treatment form of the dynamic CF instructional strategy informs instruction and practice function as effectively as a linguistic accuracy course that utilizes the traditional form of the dynamic CF instructional strategy?

2. Are there statistically significant differences in changes in written accuracy levels between the two groups that result from the differences in the practice mode that receives feedback?

   Operationalized: Will the change in mean accuracy scores from the pretest writing samples to the post-test writing samples as measured by error-free clause ratios be significantly different for the students in the treatment group?

3. Are there other statistically significant differences including gains or losses in written fluency, complexity, and lexical development that result from differences in the practice mode that receives feedback?

   a) Operationalized for complexity: Will the change in average number of words per clause and the average number of clauses per T-unit from the pre-test writing

samples to the post-test writing samples be significantly different for the students in the treatment group?

b) Operationalized for fluency: Will the change in total number of words written from the pre-test writing samples to the post-test writing samples be significantly different for the students in the treatment group?

c) Operationalized for lexical development:

   i. Type Token Ratio: Will there be a significant difference in changes in the type token ratio from the pretest writing samples to the post-test writing samples for the students in the treatment group?

   ii. Increased academic vocabulary density: Will there be a significant difference in changes in the percent of total words found on the Academic Word List as part of any word family from the pre-test writing samples to the post-test writing samples for the students in the treatment group?

Chapter 4 Results

This chapter will achieve three primary purposes. First, the results of the methods used to estimate reliability will be presented. Second, for each of the variables in the study, the descriptive statistics, and the several repeated measures ANOVA results used to answer the primary research questions will be presented. Last, the overall effectiveness of the two treatments in isolation will be estimated by tests of simple main effects comparing the pretest and posttest performance of the student participants.

Reliability Estimates

The results achieved in the analysis of the data are only valuable if they are derived from reliable findings. Consequently, establishing the reliability of the methods through which the data were gathered is an important step in evaluating the effect of the treatment. The procedures designed to provide evidence for reliability were described in Chapter 3 along with reliability estimates (see Table 7 in Chapter 3).

Effect size

In keeping with the body of research on dynamic CF, this study reports effect size with both significant and non-significant results. Lee (2009) concluded in her analysis that reporting effect size was recommended to compensate for possible deflation of statistical significance due to insufficient sample size. This study used the same measure of effect size, partial eta squared (η^2_p), to establish the magnitude of effect used by Lee and Hartshorn (2008) in their evaluations of the effect of dynamic CF. Both researchers cited in their analysis of this measure of effect size Bakeman and Robinson (2005) who recommended using η^2_p in repeated measure designs because it can extract the effect of a specific variable and be used for comparison within and across studies. In terms of determining the relative size of an effect measurement, the standard

first presented by Cohen (1988) and then promoted by Huck (2008) was used. These researchers suggested that effect sizes greater than .01 should be seen as small, greater than .06 as medium and greater than .14 as large.

ANOVA Results

This section presents the results from the mixed model, repeated measure Analysis of Variance (ANOVA) that was used to compare gains in four primary aspects of writing performance: accuracy, fluency complexity, and lexical development. There is always a risk with multiple analyses that statistically significant results could be the consequence of chance. While those behind the research agenda acknowledge this, achieving a more holistic view of the instructional strategy's impact on all the sub-skills of writing seemed to justify the potential for such statistical risks.

T-tests revealed that for the majority of the variables being examined in the study, equal variance could be assumed, including: for the combined study group the variables of accuracy t (51) = .814, p = .419; complexity, t (51) = 1.369, p = .177; fluency t (51) = .053, p = .419; percent of words derived from AWL word families, t (51) = -.757, p = .453; for the Academic A study sections: accuracy t (24) = 1.14, p = .267; fluency t (24) = 1.55, p = .134; percent of words derived from AWL word families, t (24) = -.446, p = .659; for the Academic B study sections: accuracy t (25) = .032, p = .975; complexity, t (25) = -.273 p = .787; fluency t (25) = 1.174, p = .251; type-token ratio, t (25) = -.819, p = .421; and percent of words derived from the AWL word families, t (25) = -.819, p = .421.

The t-test revealed that in the following instances equal variance could be assumed. This is likely to have no real effect on interpreting the overall gains of the groups being examined. For the combined study sections, only in terms of type token ratio t (50.858) = -2.2, p = .032 could

equal variance not be assumed. Also for the Academic A study sections in terms of complexity, t (21.23) = 2.736, p = .012 and type token ratio, t (23.978) = 23.978, p = .033, equal variance could not be assumed.

A repeated measures ANOVA test was used to measure all of the subcomponents needed to answer the primary research questions. At the center of this study was the question "To what extent will the treatment variation of dynamic CF produce equivalent levels of fluency, complexity, and accuracy on a new piece of writing as the traditional approach used by the contrast group?" Of further interest to research was the impact, if any, of either variation of dynamic CF on lexical development. Operationally, these subcomponents were defined in Chapter 3.

This chapter presents the ANOVA results for each of the subcomponents in the order that they have been described above. The results of each of these subcomponent sections is briefly summarized; a table will then present the descriptive statistics for the subcomponent. The mixed ANOVA results are then be discussed and presented. Finally, a table addresses the simple main effect size for the subcomponent and how the simple main effect data can affect the interpretation of the results is discussed.

Fluency

While the aim of dynamic CF is to help facilitate accuracy gains, as it has been discussed here and in other studies on the instructional strategy, any notable impact on other areas of writing skill production—namely fluency and complexity—are significant (Hartshorn, 2008; Hartshorn et. al, 2010). In this section, the results of this study in regard to fluency are presented. As mentioned in the analysis section of Chapter 3, the measurement of fluency was operationalized as the total words in a participant's writing sample.

First, the descriptive statistics as presented for the three grouping in Tables 11-13 for the combined grouping and Academic A study, the treatment sections increased in the total words of their writing sample while the contrast group decreased in their total words. For the Academic B grouping, both the contrast and treatment group saw an increase in their total word count; however, the gain for the treatment group was larger (3.1% increase for the contrast and 7% increase for the treatment).

Table 11

Descriptive Statistics for Fluency Scores for Combined Study Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast	Mean	397.84	381.20	389.52
(n= 25)	SD	100.68	71.73	86.21
Treatment	Mean	342.86	370.39	356.63
(n= 28)	SD	101.28	82.22	91.75
Total	Mean	368.79	375.49	372.14
(n= 53)	SD	103.79	76.91	90.35

Table 12

Descriptive Statistics for Fluency Scores for Academic A Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast	Mean	405.42	357.42	381.42
(n=12)	SD	113.212	65.56	89.386
Treatment	Mean	332.36	362.64	347.50
(n= 14)	SD	124.99	89.68	107.33
Total	Mean	366.08	360.23	363.16
(n= 26)	SD	123.05	77.97	100.51

Table 13

Descriptive Statistics for Fluency Scores for Academic B Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast	Mean	390.85	403.15	397
(n= 13)	SD	91.72	72.546	82.133
Treatment	Mean	353.36	378.14	365.75
(n= 14)	SD	73.79	76.608	75.199
Total	Mean	371.41	390.19	380.8
(n= 27)	SD	83.484	74.335	78.9095

The mixed ANOVA results presented in Table 14 show that for the combined group this difference in gains of fluency for the treatment sections and decline of fluency for the contrast sections was not significant (p=.054). The mixed ANOVA results for Academic A sections in isolation in Table 15 show that the difference was significant (p=.04). The effect size for the combined groups (η^2_p =.071) shows that the treatment variation had a medium effect on fluency performance. The effect size reported for the Academic A study groups shows that for this proficiency level this effect appears to be large (η^2_p =.162). Although it was noted above that gains in fluency evidenced in the Academic B treatment section presented in Table 16 were larger than those gains noted in the contrast section, these gains were not significant (p=.239), and the treatment appeared to have little effect on fluency (η^2_p =.008) for students at this level of study.

Table 14

Mixed ANOVA Summary Table for Fluency Scores for Combined Study Sections

Source	*SS*	*Df*	*MS*	*F*	*p*	η^2_p
Between Subjects		52				
Group	28583.272	1	28583.272	2.217	.143	.042
Error	657623.105	51	12894.571			
Within Subjects		53				
Time	783.977	1	783.977	.237	.628	.005
Time × Group	12887.223	1	12887.223	3.898	.054	.071
Error	168594.362	51	3305.772			
Total	868471.939	105				

Table 15

Mixed ANOVA Summary Table for Fluency Scores for Academic A Study Sections

Source	*SS*	*Df*	*MS*	*F*	*p*	η^2_p
Between Subjects		25				
Group	14865.936	1	14865.93	.906	.351	.036
Error	393672.833	24	16403.03			
Within Subjects		26				
Time	1013.802	1	1013.802	.238	.630	.010
Time × Group	19800.264	1	19800.26	4.649	.041	.162
Error	102211.429	24	4258.810			
Total	115546.029	51				

Table 16

Mixed ANOVA Summary Table for Fluency Scores for Academic B Study Sections

Source	*SS*	*Df*	*MS*	*F*	*p*	η^2_p
Between Subjects		26				
Group	13165.509	1	13165.509	1.284	.268	.049
Error	256256.750	25	10250.270			
Within Subjects		27				
Time	4637.363	1	4637.363	2.111	.159	
Time × Group	524.770	1	524.770	.239	.629	.009
Error	54929.563	25	2197.183			
Total	329513.955	53				

Tables 17, 18, & 19 below present the simple main effects tests that contrast the pre-test and post-test performance for the study groups analyzed above. While the ANOVA calculation contrasts performances between the groups and identifies where there is a significant difference, the simple main effect estimate provided by these tests indicates whether the treatment or contrast groups independently demonstrated a significant change for the variable being examined. While the ANOVA showed that the contrast between the two groups was significant, the tests for simple main effects show that independently the gains or losses of the sections were not statistically significant.

Table 17

Simple Main Effects for Pretest and Posttest Fluency Scores for Combined Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	2964.500	1	2964.500	.368	.546
Treatment	10615.02	1	10615.018	1.317	.254
Error	822108.11	102	8059.883		

Table 18

Simple Main Effects for Pretest and Posttest Fluency Scores for Academic A Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	13824.000	1	13824.000	1.338	.253
Treatment	6420.571	1	6420.571	.621	.434
Error	495884.262	48	10330.922		

Table 19

Simple Main Effects for Pretest and Posttest Fluency Scores for Academic B Study Groups

Form of dynamic CF	SS	*df*	MS	F	*P*
Contrast	984.615	1	984.615	.158	.693
Treatment	4300.321	1	4300.321	.691	.410
Error	311186.313	50	6223.726		

Complexity

In addition to determining the effects of the two variations of dynamic CF on the fluency of writing, the research examined both form's impact on writing complexity. In this section, the results of the complexity analysis are reported. As mentioned in the analysis section of Chapter 3, measurement of complexity was operationalized as mean length of T-unit or the average number of words per T-unit.

The descriptive statistics presented in Tables 20-22 for the three grouping combinations showed that students in both variations of dynamic CF demonstrated a decline in mean length of T-unit although this decline was less prominent in the treatment sections of each grouping than it was in the contrast sections. For the combined group of students from Academic A and B, the treatment group saw a decline of 5.4 % while the treatment group saw a decline of 18.9 %. For

the Academic A sections in isolation, the treatment group saw a minimal decline (1.9 %) while the contrast group saw a much more significant decline (19.9 %). For the Academic B sections in isolation, students receiving the treatment variation of dynamic CF saw a much larger decline than in Academic A (8.6 %) but a significantly smaller decline than the contrast group in Academic B (18 %).

Table 20

Descriptive Statistics for Complexity Scores for Combined Study Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 25)	Mean	16.10	13.05	14.58
	SD	3.67	2.42	3.05
Treatment (n= 28)	Mean	14.76	13.96	14.36
	SD	3.44	2.72	3.08
Total (n= 53)	Mean	15.39	13.53	14.46
	SD	3.58	2.60	3.09

Table 21

Descriptive Statistics for Complexity Scores for Academic A Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n=12)	Mean	17.12	13.72	15.42
	SD	3.21	2.53	2.87
Treatment (n= 14)	Mean	13.95	13.68	13.82
	SD	2.61	2.58	2.59
Total (n= 26)	Mean	15.41	13.70	14.56
	SD	3.27	2.50	2.88

Table 22

Descriptive Statistics for Complexity Scores for Academic B Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 13)	Mean	15.16	12.43	27.59
	SD	3.94	2.24	6.18
Treatment (n= 14)	Mean	15.58	14.23	29.81
	SD	4.04	2.94	6.98
Total (n= 27)	Mean	15.38	13.36	28.74
	SD	3.92	2.74	6.65

The mixed ANOVA results presented in Tables 23-25 for the complexity measurement showed that while each of these groupings demonstrate a difference in complexity performance favoring the treatment, this difference was statistically significant when viewed in combination (p=.008) and when Academic A was viewed in isolation (p=.015). Both of these groupings also revealed a large effect size (η^2_p =.131 and η^2_p =.222 respectively). The difference of the treatment in favor of the students of Academic B receiving speech responsive dynamic CF toward complexity was not statistically significant (p=.227) but neared a moderate effect size (η^2_p =.058).

Table 23

Mixed ANOVA Summary Table for Complexity Scores for Combined Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		52				
Group	33.223	1	33.223	7.665	.008	.131
Error	221.052	51	4.334			
Within Subjects		53				
Time	98.140	1	98.140	22.642	.000	.307
Time × Group	33.223	1	33.223	7.665	.008	.131
Error	221.052	51	4.334			
Total	606.69	105				

Table 24

Mixed ANOVA Summary Table for Fluency Scores for Academic A Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		25				
Group	33.344	1	33.344	3.242	.084	.119
Error	246.820	24	10.284			
Within Subjects		26				
Time	43.381	1	43.381	9.382	.005	.281
Time × Group	31.685	1	31.685	6.852	.015	.222
Error	110.978	24	4.624			
Total	466.208	51				

Table 25

Mixed ANOVA Summary Table for Complexity Scores for Academic B Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		26				
Group	16.619	1	16.619	.891	.354	.034
Error	466.164	25	18.647			
Within Subjects		27				
Time	55.900	1	55.900	13.361	.001	.348
Time × Group	6.429	1	6.429	1.537	.227	.058
Error	104.591	25	4.184			
Total	649.703	53				

The ANOVA results related to the measurement of complexity revealed a significant difference between participants in the two study groups even though both groups showed some decline in the measurement of complexity. Tables 26, 27, and 28 indicate that viewed in isolation the decrease in complexity for the treatment group was not statistically significant. However the decline in complexity for the contrast group was statistically significant for all of the study groups.

Table 26

Simple Main Effects for Pretest and Posttest Complexity Scores for Combined Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	102.864	1	102.864	10.835	.001
Treatment	9.112	1	9.112	.960	.330
Error	968.387	102	9.494		

Table 27

Simple Main Effects for Pretest and Posttest Complexity Scores for the Academic A Study Groups

Form of DYNAMIC CF	SS	*df*	MS	F	*p*
Contrast	69.279	1	69.279	9.294	.004
Treatment	.497	1	.497	.067	.797
Error	968.387	102	357.797		

Table 28

Simple Main Effects for Pretest and Posttest Complexity Scores for the Academic B Study Groups

Form of DYNAMIC CF	SS	*df*	MS	F	*p*
Contrast	48.331	1	48.331	4.234	.045
Treatment	12.677	1	12.677	1.111	.297
Error	570.755	50	11.415		

Accuracy

The primary objective of the Linguistic Accuracy courses is to improve students' accurate production of English. Dynamic CF's primary established benefit has been shown to be an increase of accurate written production (Evans et al., 2010; Evans et al., 2011; Hartshorn et al., 2010; Hartshorn, 2008; Lee, 2009). As mentioned in the analysis section of Chapter 3, measurement of accuracy was operationalized as the error-free clause ratio or the number of clauses that are without error divided by the total number of clauses.

Perhaps not surprisingly, the results for this particular variable of written performance as presented in Table 29 – 31 showed a much greater variety than the proceeding two variables. All the sections in the study evidenced a decline in accuracy under the measurement employed. For

the treatment sections in the study, this decline seemed to be more consistent. The combined treatment groups saw a 7.9% decline in error-free clause ratio, the Academic A treatment section saw a 9.9% decline and the Academic B section saw a decline of 6.1 %. For the contrast sections of the study groups, there was much more variation in the decline evidenced. For the combined contrast groups the decline was measured to be 4.9 %. The decline for Academic A was very slight (.2 %); however, the decline for Academic B was notable (9.1 %).

Table 29

Descriptive Statistics for Accuracy Scores for Combined Study Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 25)	Mean	0.601	0.571	0.586
	SD	0.120	0.135	0.128
Treatment (n= 28)	Mean	0.572	0.527	0.550
	SD	0.142	0.137	0.140
Total (n= 53)	Mean	0.586	0.548	0.567
	SD	0.132	0.136	0.134

Table 30

Descriptive Statistics for Accuracy Scores for Academic A Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n=12)	Mean	0.580	0.579	0.580
	SD	0.120	0.154	0.137
Treatment (n= 14)	Mean	0.524	0.472	0.498
	SD	0.129	0.114	0.122
Total (n= 26)	Mean	0.550	0.522	0.536
	SD	0.126	0.142	0.134

Table 31

Descriptive Statistics for Complexity Scores for Academic B Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 13)	Mean	0.621	0.564	0.592
	SD	0.122	0.120	0.121
Treatment (n= 14)	Mean	0.619	0.581	0.600
	SD	0.142	0.139	0.140
Total (n= 27)	Mean	0.620	0.573	0.596
	SD	0.130	0.128	0.129

The mixed ANOVA results in Tables 32 – 34 for the accuracy measurement showed that for none of the groupings the difference in accuracy gains (or in this case losses) was of statistical significance. Although not statistically significant, the difference in the Academic A contrast section was notable (p=.071) and evidenced a minimal effect size (η^2_p =.035). Also viewing the groups together or looking at the performance of students in Academic B alone was not of statistical significance nor evidenced an effect size sufficiently large to mention.

Table 32

Mixed ANOVA Summary Table for Accuracy Scores for Combined Study Sections

Source	*SS*	*Df*	*MS*	*F*	*P*	η^2_p
Between Subjects		52				
Group	.036	1	.036	1.325	.255	.025
Error	1.386	51	.027			
Within Subjects		53				
Time	.037	1	.037	4.273	.044	.077
Time × Group	.001	1	.001	.165	.686	.003
Error	.441	51	.009			
Total	3.802	105				

Table 33

Mixed ANOVA Summary Table for Accuracy Scores for Academic A Study Sections

Source	*SS*	*df*	*MS*	*F*	*P*	η^2_p
Between Subjects		25				
Group	.086	1	.086	3.583	.070	.130
Error	.573	24	.024			
Within Subjects		26				
Time	.009	1	.009	.936	.343	.038
Time × Group	.008	1	.008	.872	.360	.035
Error	.231	24	.010			
Total	0.907	51				

Table 34

Mixed ANOVA Summary Table for Accuracy Scores for Academic B Study Sections

Source	*SS*	*df*	*MS*	*F*	*P*	η^2_p
Between Subjects		26				
Group	.001	1	.001	.033	.858	.001
Error	.665	25	.027			
Within Subjects		27				
Time	.030	1	.030	3.777	.063	.131
Time × Group	.001	1	.001	.158	.694	.006
Error	.199	25	.008			
Total	.001	1	.001	.033	.858	.001

The simple main effect estimates for the accuracy variable listed in Table 35-37 temper the apparent decline indicated Tables 32-34. Although both the treatment and contrast sections in all arrangements viewed here showed a decrease in their overall accuracy as measured by error-free clause ratios in the posttest relative to the pretest, the tests for simple main effects presented in Tables 35-37 indicate that these declines were not statistically significant.

Table 35

Simple Main Effects for Pretest and Posttest Accuracy Scores for Combined Study Groups

Form of dynamic CF	SS	*df*	MS	F	*P*
Contrast	.015	1	.015	.815	.369
Treatment	.029	1	.029	1.611	.207
Error	1.842	102	.018		

Table 36

Simple Main Effects for Pretest and Posttest Accuracy Scores for Academic A Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	4.982E-6	1	4.982E-6	.000	.986
Treatment	.019	1	.019	1.125	.294
Error	.804	48	.017		

Table 37

Simple Main Effects for Pretest and Posttest Accuracy Scores for Academic B Study Groups

Form of DYNAMIC CF	SS	*df*	MS	F	*P*
Contrast	.021	1	.021	1.219	.275
Treatment	.010	1	.010	.573	.453
Error	.865	50	.017		

Lexical Development

While extensive measurements of dynamic CF's influence on writing fluency, complexity and accuracy had previously been done, the instructional strategy's impact on lexical development of student's writing had only been of minimal interest in some of the previous research investigations of dynamic CF (Hartshorn, 2008; Hartshorn et al., 2010). In this study,

lexical development was viewed through two measures as explained in Chapter 3. The first of these measurements was a traditional lexical measurement: type token ratio. The second was a contextually significant measurement of the percentage of total words in a writing sample that are derived from Academic Word List word families. First, the results of the type token ratio measurement will be presented and then the results from the Academic Word List measurement will be presented.

As Tables 38-40 for the descriptive statistics indicate, all sections in the study saw an increase in the type token ratio of their post-test writing sample over the pre-test writing sample. While in all three of the groupings, the contrast sections showed a slightly larger increase in type token ratio, the ANOVA summary Table shows that this increase was not statistically significant (p=.16) and the effect size shows that it was not large enough to be of practical significance .

Table 38

Descriptive Statistics for Type-token Ratio for Combined Study Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 25)	Mean	0.376	0.430	0.404
	SD	0.061	0.112	0.086
Treatment (n= 28)	Mean	0.418	0.443	0.430
	SD	0.072	0.056	0.064
Total (n= 53)	Mean	0.399	0.437	0.418
	SD	0.069	0.086	0.078

Table 39

Descriptive Statistics for Type-token Ratio Scores for Academic A Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n=12)	Mean	0.3670	0.4165	0.3917
	SD	0.0611	0.0516	0.0563
Treatment (n= 14)	Mean	0.4269	0.4406	0.4337
	SD	0.0740	0.0409	0.0574
Total (n= 26)	Mean	0.3992	0.4295	0.4143
	SD	0.0736	0.0468	0.0602

Table 40

Descriptive Statistics for Type-token Ratio Scores for Academic B Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 13)	Mean	0.3874	0.4434	0.4154
	SD	0.0611	0.1486	0.1049
Treatment (n= 14)	Mean	0.4083	0.4448	0.4266
	SD	0.0711	0.0689	0.0700
Total (n= 27)	Mean	0.3982	0.4441	0.4212
	SD	0.0661	0.1121	0.0891

Table 41

Mixed ANOVA Summary Table for Type-token Ratio for Combined Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		52				
Group	.018	1	.018	2.423	.126	.045
Error	.114	24	.005			
Within Subjects		53				
Time	.013	1	.013	6.273	.019	.207
Time × Group	.004	1	.004	2.032	.167	.078
Error	.049	24	.002			
Total	0.203	51				

Table 42

Mixed ANOVA Summary Table for Type-Token Ratio Scores for Academic A Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		25				
Group	.023	1	.023	4.813	.038	.167
Error	.114	24	.005			
Within Subjects		26				
Time	.013	1	.013	6.273	.019	.207
Time × Group	.004	1	.004	2.032	.167	.078
Error	.049	24	.002			
Total	0.203	51				

Table 43

Mixed ANOVA Summary Table for Type-Token Ratio Scores Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		26				
Group	.002	1	.002	.164	.689	.007
Error	.258	25	.010			
Within Subjects		27				
Time	.029	1	.029	4.018	.056	.138
Time × Group	.001	1	.001	.179	.676	.007
Error	.179	25	.007			
Total	0.469	53				

The calculations of the simple main effect statistics presented in Table 44-46 for the treatment and contrast sections respectively indicate that participants in both treatment and in the contrast sections made significant gains in increasing the type-token ratio of their writing between the pre- and post-tests. The single section that was an exception to this was the Academic B group in the Contrast section for which the t-test indicates that their increase was not statistically significant.

Table 44

Simple Main Effects for Pretest and Posttest Type-Token Ratio Scores for Combined Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	.033	1	.033	5.587	.020
Treatment	.009	1	.009	1.470	.228
Error	.608	102	.006		

Table 45

Simple Main Effects for Pretest and Posttest Type-Token Ratio Scores for the Academic A Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	.015	1	.015	4.348	.042
Treatment	.001	1	.001	.383	.539
Error	.163	48	.003		

Table 46

Simple Main Effects for Pretest and Posttest Type-Token Ratio Scores for the Academic B Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	.020	1	.020	2.329	.133
Treatment	.009	1	.009	1.064	.307
Error	.438	50	.009		

The second subsection of the lexical development analysis presents the results of the analysis of the two variations of dynamic CF on changes in the percent of total words derived from Academic Word List word families. The contextual significance of this measurement is addressed in Chapter 3. This subsection will also proceed with the presentation of the descriptive statistics in Tables 47-49. After which, a short analysis will bridge the descriptive statistic results with the presentation of the ANOVA results in Tables 50-52 and the presentation of the results from simple main effects tests in Tables 53-55.

Table 47

Descriptive Statistics for Percent of Words Derived from AWL Word Families for Combined Study Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 25)	Mean	3.44	8.26	5.85
	SD	1.32	2.08	1.7
Treatment (n= 28)	Mean	3.87	8.65	6.26
	SD	2.59	3.39	2.99
Total (n= 53)	Mean	3.67	8.47	6.07
	SD	2.08	3.11	2.595

Table 48

Descriptive Statistics for Percent of Words Derived from AWL Word Families for Academic A Study Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n=12)	Mean	3.3	9.11	6.20
	SD	1.41	2.75	2.08
Treatment (n= 14)	Mean	3.67	8.77	6.22
	SD	2.57	3.52	3.04
Total (n= 26)	Mean	3.50	8.93	6.21
	SD	2.08	3.13	2.61

Table 49

Descriptive Statistics for Percent of Words Derived from AWL Word Families for Academic B Study Sections

Form of dynamic CF		Pretest	Posttest	Mean
Contrast (n= 13)	Mean	3.56	7.48	5.52
	SD	1.27	2.76	2.02
Treatment (n= 14)	Mean	4.07	8.53	6.30
	SD	2.70	3.38	3.04
Total (n= 27)	Mean	3.83	8.03	5.93
	SD	2.11	3.08	2.60

As is evidenced, all sections in the group made substantial gains in the percent of total words in the posttest writing sample that are derived from Academic Word List word families. There was little to no difference in the relative gains between the treatment and contrast variation of dynamic CF. This conclusion was ratified by the ANOVA results that show absolutely no significant difference resulting in the contrasting treatments of the two groups as indicated by the *p* and *partial eta* squared values.

Table 50

Mixed ANOVA Summary Table for Percent of Words derived from AWL Word Families for Combined Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		52				
Group	4.456	1	4.456	.491	.487	.010
Error	462.694	51	9.072			
Within Subjects		53				
Time	609.437	1	609.437	118.890	.000	.700
Time × Group	.016	1	.016	.003	.955	.000
Error	261.429	51	5.126			
Total	1338.032	105				

Table 51

Mixed ANOVA Summary Table for Percent of Words derived from AWL Word Families for Academic A Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		25				
Group	.004	1	.004	.000	.983	.000
Error	191.582	24	7.983			
Within Subjects		26				
Time	384.435	1	384.435	57.599	.000	.706
Time × Group	1.621	1	1.621	.243	.627	.010
Error	160.185	24	6.674			
Total	737.827	51				

Table 52

Mixed ANOVA Summary Table for Percent of Words derived from AWL Word Families for Academic B Study Sections

Source	*SS*	*df*	*MS*	*F*	*p*	η^2_p
Between Subjects		26				
Group	8.138	1	8.138	.767	.389	.030
Error	265.236	25	10.609			
Within Subjects		27				
Time	236.695	1	236.695	66.706	.000	.727
Time × Group	.961	1	.961	.271	.607	.011
Error	88.709	25	3.548			
Total	599.739	53				

The simple main effects estimates presented in Tables 49 and 50 further demonstrate increases in relative frequency of academic vocabulary between the pre and post-tests. The sections of the study all demonstrated statistically significant increases in the percent of total words in the writing sample derived from Academic wordlist word families after the treatment period.

Table 53

Simple Main Effects for Pretest and Posttest Percent of Total Words Derived from AWL Word Family Scores for Combined Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	299.635	1	299.635	42.476	.000
Treatment	319.686	1	319.686	45.318	.000
Error	719.539	102	7.054		

Table 54

Simple Main Effects for Pretest and Posttest Percent of Total Words Derived from AWL Word Family Scores for the Academic A Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	202.420	1	202.420	27.621	.000
Treatment	182.070	1	182.070	24.844	.000
Error	351.766	48	7.328		

Table 55

Simple Main Effects for Pretest and Posttest Percent of Total Words Derived from AWL Word Family Scores for the Academic B Study Groups

Form of dynamic CF	SS	*df*	MS	F	*p*
Contrast	100.038	1	100.038	14.132	.000
Treatment	139.063	1	139.063	19.645	.000
Error	353.945	50	7.079		

Chapter 5: Discussion and Conclusions

The purpose of this chapter is to address the results related to the research questions and the practical implications of these findings. In addition to this reflective discussion, this chapter will identify a number of limitations with this study, present several pedagogical implications, and extend some suggestions for further research.

Discussion

Although accurate production is a critical component of language mastery for a particular subset of English language learners, achieving and maintaining these levels of accuracy are difficult challenges. Because of the high-stakes nature of accurate production in specific language production contexts and the reality that much of the investment of time and attention towards accuracy attainment by both student and teacher fails for several reasons, language instructors and researchers should continue to examine the process of facilitating accurate production in both writing and speaking skills. Substantial evidence has demonstrated that while the ultimate accuracy attainment of certain learners may be limited, improvement can be fostered by the right instructional approaches (Lyster & Ranta, 1997; Spada & Lightbown, 1993; Lyster et al., 1999; Ferris, 1999; Ellis, 1998; Bitchener, 2008). Among these best practices, dynamic CF has been shown to facilitate improvements in written production accuracy with intermediate high and advanced low learners engaged in intensive and institutionally supported language instruction (Evans et al., 2011; Evans et al., 2010; Hartshorn et al., 2010; Hartshorn, 2008; Lee, 2009). This research study intended to examine the effect of adapting the existing dynamic CF model to more directly address students' spoken language accuracy needs.

The potential for dynamic CF to benefit L2 learners is the result of engaging students in authentic negotiation while enabling meaningful, manageable, timely, and constant feedback. As

has been addressed by its proponents, however, it is important to understand that there is a degree of tension in accuracy development within language subskills. Increased cognitive and attention resources directed toward any one particular subskill can reduce the cognitive raw materials available to sustain levels of performance in others.

SLA researchers have noted that L2 learners' language skills start as controlled processes before moving to automatic processes. Controlled processes allow for self-regulation, which is required for initial language refinement. However, because of their cognitive and attention demands, controlled processes can create bottlenecks in performance. Thus a learner's intent to demonstrate a particular level of accuracy, before such a level can be maintained by automatic rather than controlled processes, can cause other demands to wait for processing. This limited model of capacity can offer a partial explanation for the observation that an increased attention to accuracy can impede demonstration of fluency or complexity (Ortega, 2009). For this reason, researchers of dynamic CF have conditioned their conclusions into the efficacy of the practice saying that the gains in the subskill of accuracy are justified when they do not create undue losses in the other subskill areas (Hartshorn et al., 2010).

For this reason, in addition to compensating for undesirable reductions in written fluency and complexity, addressing the spoken production components of accuracy, fluency, and complexity should not be overlooked. SLA research has shown that proceduralization—that is, moving from controlled processes that govern these language subskills to automatic processes—is skill specific (DeKeyser, 2007). This is especially important when examining the impact and improvement of the dynamic CF instructional strategy. Adapting the current instructional strategy to create practice procedures that activate the cognitive systems of spoken English

production is important to spread the benefits of dynamic CF more widely and avoid potentially harmful imbalances of strength.

The feasibility of an oral skills adaptation was established with an implementation a variation of dynamic CF that used spoken production in place of writing samples. This variation functioned similarly, both pedagogically and practically, within the instructional context examined here to the more traditional form of the contrast treatment. However, efficacy, not feasibility, will need to be determined in order for the adaptation to be considered a success and more fully integrated into the curriculum. There is the hope that the adaptation will offer unique benefits for spoken language production, particularly in terms of spoken accuracy attainment and spoken fluency and complexity maintenance; however, this modification should not detract from the established benefits to written accuracy of the form of written dynamic CF used by the contrast group.

This research study, while initiating the data collection to answer all of the components of a broad research agenda described further in the section for future research, only provided conclusions to the questions pertaining to the impact of the modification on student participants' written performance. The study described how the gains of written accuracy, fluency, and complexity differed for students engaged in a form of dynamic CF that was responsive to speaking from those attained by students participating in the conventional written production responsive dynamic CF model. The study also sought to add to the understanding of the impact of the dynamic CF on crucial systems of language production by examining its influence on lexical development.

In order to examine these elements of written production by contrasting performance on a pre- and post-treatment, five statistical analyses were devised. Four were determined at the onset

of the study to examine accuracy, fluency, complexity, and lexical development respectively; the fifth was added to clarify and contextualize the results of the findings of the analysis of lexical development. Because the study included many different statistical tests, it may be helpful to first provide a synopsis of their findings.

These analyses were conducted on data gathered from two consecutive levels of proficiency that are currently using the dynamic CF instructional strategy at the English Language Center. The data from these two levels of proficiency were analyzed together and in isolation. Consequently, Table 56 summarizes the findings with these three distinctions: study sections from A and B combined, the study sections from A in isolation, and the study sections from B in isolation. Table 56 includes the relevant dependent variables, the associated *p*-values, the eta statistics that establish effect size, and an interpretation of this effect size on writing performance relative to the subskill being examined. The effect is also labeled as *negligible, small, moderate,* or *large*.

Table 56

A Summary of Findings Used to Answer the Primary Research Questions

Grouping	Dependent Variable	p	η^2_p	Effect Estimate
Combined	Accuracy Scores	.165	.003	negligible
	Complexity Scores	.008	.131	large
	Fluency Scores	.054	.071	moderate
	Lexical Development Scores			
	Type Token Ratio	.292	.022	small
	AWL Derivation	.955	.000	negligible
Academic A	Accuracy Scores	.360	.035	small
	Complexity Scores	.015	.222	large
	Fluency Scores	.041	.162	large
	Lexical Development Scores			
	Type Token Ratio	.167	.078	moderate
	AWL Derivation	.627	.010	small
Academic B	Accuracy Scores	.694	.006	negligible
	Complexity Scores	.227	.058	small
	Fluency Scores	.239	.009	negligible
	Lexical Development Scores			
	Type Token Ratio	.676	.007	negligible
	AWL Derivation	.607	.011	small

Table 56 shows that the treatment seemed to have no real advantage for improved writing accuracy, a finding that is inconsistent with the gains to accuracy observed in earlier studies. This surprising decline will be addressed shortly. Interestingly, there were noted advantages for the treatment group in terms of writing complexity and fluency for Academic A and when the groups were viewed in combination. This finding addresses concerns of previous research that feared there might be a stifling impact of dynamic CF on writing fluency and complexity (Hartshorn, 2008; Hartshorn et al., 2010). Also, the results are clear that one variation of the instructional strategy offers no advantage over the other in terms of lexical development by either measurement in any grouping. Although it may not be impacted by the instructional

strategy, lexical development could have impacted the effects of the instructional strategy particularly in regard to accuracy, which will also be discussed further.

The most critical result to discuss is the decline in accuracy evidenced in both the treatment and control groups, each of which were using a form of dynamic CF. While it perhaps is to be expected that the modification of dynamic CF used by the treatment group would have varied from the results of the instructional strategy established from other studies, that the contrast group deviated so significantly from the expected results is surprising and should be accounted for. To initiate this discussion, the descriptive statistic results from three dynamic CF studies are presented in Table 57. Table 58 then presents a summary of the descriptive statistics from this study.

Table 57

Review of Accuracy Performance Descriptive Statistics from three dynamic CF research studies

	Hartshorn et. al. 2008			Evans et. al 2011			Lee 2009		
Group	Pretest Mean	Posttest Mean	Change	Pretest Mean	Posttest Mean	Change	Pretest Mean	Posttest Mean	Change
Control	.163	.138	-.153	.514	.503	-.021	.179	.268	+.497
Treatment	.140	.242	+.728	.471	.578	+.227	.242	.369	+.527
Total	.149	.200	+.342	.491	.543	+.106	.221	.337	+.522

Table 58

Review of Accuracy Performance Descriptive Statistics from Three Groupings in this Study.

	Combined			Academic A			Academic B		
Group	Pretest Mean	Posttest Mean	Change	Pretest Mean	Posttest Mean	Change	Pretest Mean	Posttest Mean	Change
Control	0.601	0.571	-.050	0.580	0.579	-.002	0.621	0.564	-.092
Treatment	0.572	0.527	-.079	0.524	0.472	-.010	0.619	0.581	-.061
Total	0.586	0.548	-.065	0.550	0.522	-.050	0.620	0.573	-.076

As Table 57 indicates, in each of these three studies, the treatment group, which received a form of written dynamic CF, saw an increase in mean accuracy performance. The differences in the comparisons being made in these studies should be understood. Evans et al. (2011) and Hartshorn et al. (2010) looked at students learning through the dynamic CF instructional strategy in contrast to students learning through regular process writing. Lee (2009) contrasted students using dynamic CF as an alternative to traditional grammar instruction. Also, in Lee's study, students in both study groups also engaged in a traditional process writing curriculum. In the research being addressed here, two groups of students using dynamic CF as part of a Linguistic Accuracy class are compared. The difference being examined was in the student production mode receiving feedback, speaking or writing, which varied between the treatment and contrast group. Like the study groups in Lee (2009), both sections of these students also had a traditional process writing course.

Differences in measurement, context, and proficiency should also be considered. Hartshorn et al. (2010) used error free T-units as their accuracy standard, which likely accounts for the smaller ratios reported because the T-unit encompasses more language and is at increased risk for error. Notable differences in context would include the Evans et al. (2011) study which included matriculated university students while Lee (2009) and Hartshorn et al. (2010) consisted of students enrolled in the same IEP as the participants in this study. Participants in Hartshorn et al. (2010) and Evans et al.'s (2010; 2011) studies would likely have been of higher language proficiency than the majority of students examined in the current study. Lee's study (2009) consisted of students that would have approximated the central 50 % of students in this study in terms of language proficiency.

In accounting for differences in the accuracy results, the differences of context, proficiency, and research design mentioned above potentially contributed to the disparity. Also, for all of the students examined in the previously conducted studies, the semester of treatment was the first semester of treatment with dynamic CF instructional strategy. In the current study however, a good portion of students from both proficiency levels, but particularly Academic B, had already participated in the dynamic CF instructional strategy. Likely this offers some advantages, but it also could have depressed the immediate gains established in the previous studies.

Finally, there is the question of the impact of lexical development on accuracy. It was determined that neither variation of dynamic CF seemed to impact lexical development. However, there are both qualitative and quantitative reasons to suspect that lexical development was depressing the accuracy gains previously established. Although this study was conducted in the same IEP as the studies presented in Hartshorn et al. (2010) and Lee (2009), some notable changes have occurred in the curriculum of the IEP that were not present during the previous examinations. Perhaps the one that has the most direct impact on accuracy performance is a new focus on academic vocabulary acquisition that is being integrated in each of the Academic program's skills classes. Each week the students in the Academic program take a test on sublists from the Academic Word List and each skills class is supposed to include 40 minutes of instruction or activities focused on the list of words for that week.

The data from the lexical development component of this study suggest that this focused attention and instruction on acquiring vocabulary may have the desired effect as the post-tests for both groups contained, on average, a much higher concentration of words derived from the word families on the Academic Word List, something to be examined by future or retrospective

analysis. Students may be becoming familiar enough with the meaning of the words to attempt to use them in their own production but missing out on key grammatical aspects of word knowledge necessary to use the academic words correctly. Thus, the increased lexical complexity of their writing may be leading toward the perceived decline in the accuracy of their writing. This issue will be addressed further in the section of this chapter on suggestions for future research.

Limitations

While this study addressed some of the limitations of previous dynamic CF research primarily through the use of randomized assignment of participants to treatments, evaluation of whole skill performance including accuracy, fluency, complexity, inclusion of additional lexical analyses, and concurrent evaluation of two consecutive proficiency levels, it is not without its own set of limitations that should influence the interpretation of its findings. This section will summarize some of these limitations.

This study altered the daily writing schedule from previous administrations of the dynamic CF treatment. While in previous studies, a new paragraph was scheduled to start four times a week, this administration of the instructional strategy for both variations initiated a new drafting cycle three times weekly. This reduction to a three-day week was done for several reasons. First, because the classes were required to go to the computer lab for the practice activities, four days of productive tasks would have further monopolized the lab resources of the institution. Also, not having a fourth production activity added additional instruction time, which was appreciated by all the teachers and allowed them space to meet other course objectives.

While the contrast section of the study could have potentially initiated a drafting cycle each day of the week, the treatment group, when time for transcription was factored in, would

have been overly burdened by this pace. Thus for the two groups to have equivalent practice opportunities, the initiation of drafting cycles was capped at three per week for a total of 30 initiations over the course of the semester. No drafts were initiated the first week of the semester or the last week of the semester, a mid semester week was also free from new initiations of drafts to accommodate mid semester evaluations and reprieve for both students and teachers. Adaptations to the instructional strategy that may address the limitation of instruction time will be addressed in the section of this chapter on suggestions for future research.

A reduction of drafting cycles could have contributed to the disparity of results between this research and previously concluded studies. Researchers in dynamic CF have promoted that a necessary threshold of treatment is necessary before the benefits of the instructional strategy are realized but there has not been an established quantification of where that threshold occurs. Hartshorn (2008; Hartshorn et al. 2010) and Lee (2009) indicated that paragraphs were initiated nearly every day of instruction for sections receiving dynamic CF instruction. Evans et al. (2011) indicate that drafting cycles began 3 or 4 times a week. While none of these researchers indicate in publication the exact number or an approximation of the number of drafting cycles that occurred in a semester of treatment, through personal communication with Hartshorn and Evans it is estimated that the 30 drafting cycles initiated by the current study was anywhere from 15 to 40 percent less than previous administrations (personal communication, June 1, 2011).

Teacher effect should also be addressed. Teachers were not randomly assigned to the section or the treatment that they taught in. For this administration, there were four teachers assigned to four sections. A more ideal arrangement would have had one teacher for both sections in a particular treatment or even a single teacher for all four sections in the study, thus ensuring that the elements of a typical class experience were far more standardized than they

likely were. The teachers in both study groups were asked to work together with the teacher teaching the same variation of dynamic CF in order to keep their sections as similar as possible. Nevertheless, it is possible that a teacher's personality, rapport with students, experience, or expertise contributed to the observed results.

In previous research applications of the instructional strategy the primary researcher has been directly involved with the classroom instruction administration of either the treatment or contrast groups (Evans et al, 2010; Hartshorn et al., 2010; Lee, 2009). In this study the researcher remained independent of classroom instruction with the exception of proctoring the administration of pre and post-tests. While the study may have gained some insight from this more objective arrangement, this may have also contributed to instances of a lack of precision or cross-section inconsistency. Because of the rigor of the instructional strategy and the necessity to maintain high levels of consistency, future investigations should consider having individuals invested in the research be similarly invested in the treatment and contrast instruction.

Similar to teacher effect, class dynamic could have affected individual participants' experience with the pedagogy. As was mentioned, participants were randomly assigned to the treatment variations after being placed in a particular level. The random assignment that generated the most demographically balanced sections was used. Until section assignment was complete, the researcher remained blind to the names of the students being placed in a section. Thus, although demographic generalities were addressed, personality differences were not included in assigning sections, creating the potential for classes to be balanced in terms of demographic but imbalanced in some of the other factors of influence including personality, motivation, and work ethic.

Differences in section dynamic were evidenced throughout the semester including participation, student/teacher rapport, attendance, and attrition. Unfortunately, by the time these were discovered, intervening would have led to other problems and inconsistencies of data. Because both of only two sections of students in Academic A and Academic B were participating in the study, albeit receiving different forms of treatment, it was not possible to move problematic students into other classes, Future studies should consider how they will address students and instructors with study-related concerns and if possible have a non-study related course to divert these concerns toward.

While in order for student assignment to be random, an established requirement for quantitative research, all risks to classroom dynamic may not be avoidable; however, certain steps could minimize the effects of potential dynamic disparities. First, similar to controlling for teacher effect, limiting the number of teachers over sections included in the analysis could help equalize the overall section dynamic throughout a future study. Also, including known factors of student participation including attendance, diligence in coursework and participation as one aspect of the strata used to confirm a balanced randomization is selected could also alleviate the disparities in dynamic seen in this administration.

As has been mentioned in the limitations of other studies in dynamic CF and in other studies examining accuracy, operationalizing the measurement of accuracy is challenging (Hartshorn, 2008; Hartshorn et al., 2010; Lee, 2009). This study used error-free clauses which perhaps is a little more discriminating than using a more expansive unit like T-units; however, using error-free clauses did not account for varying levels of accuracy within clauses that are not error-free. A clause that contained six errors was treated as equal to clauses that only contained one. Also egregiousness of error was not accounted for by this measurement. Because of

constraints on time and the need for consistency between raters, an error was treated as an error regardless of the level of distration from the intended meaning it created.

As has been recommended by other researchers of dynamic CF in their identification of limitations, different measurements of accuracy could be explored. One of these measurements is errors per hundred words, which would identify shifts toward accuracy that do not quite reach the threshold of error free clauses but are improvements nonetheless (Foster et al., 2000). Also, error type identification and error egregiousness evaluation could be pursued. However, it should be recognized that using a different operationalization of the measurement of accuracy will present added difficulties. These difficulties may include an increased investment of time and expertise to achieve necessary levels of reliability. This added investment could be counterbalanced by a reduction in the number of participants included in a study analysis.

Next, in regard to the instruments used in eliciting student production, added measures could be taken to ensure or verify that prompts used for the pre-test and post-test were equal in both linguistic, cognitive, and experiential demands. Steps were taken to preemptively identify that prompts were similarly demanding. Some of the measures taken included selection of the prompts from the same source, TOEFL like 30-minute essay tasks. The main researcher also showed the prompts to a range of experienced teachers in dynamic CF that expressed confidence that the pre-test and post-test prompts would be of equivalent difficulty. However, particularly in accounting for declines in accuracy from the pre-test to post-test, ruling out a variance in difficulty of prompt that could have contributed to this result is important. Future studies may benefit from using a multiple forms testing schema where students are randomly assigned to respond to one of three or four prompts for the pre-test and then randomly assigned to respond to another of the three or four prompts for the post-test. Results for students responding to a

particular prompt as pre-test or as a post-test could then help identify any influence of prompt on test performance.

An additional limitation worth noting is that the sections of students remained intact for all of their other courses at the ELC. This creates the potential that even if participants' experiences within the sections of Linguistic Accuracy were equivalent, instruction, practice, and assessment differences in one of the other three skills classes could have advantaged or disadvantaged students in a particular section and have had a confounding influence on the results of this study. Where possible, the researcher tried to minimize the potential for a class beyond Linguistic Accuracy to have a direct impact on the results of the study, but it is highly unlikely that the four classes in which all of the participants were enrolled were equivalent. Consequently, there is no guarantee that some of these differences did not impact their performance in Linguistic Accuracy.

Pedagogical Implications

Even considering these limitations, this study does present some interesting pedagogical implications. First, as was the intention of its initiation, this research does provide some justification for the initial development of a dual-skills approach to dynamic CF. It was never intended for the end pedagogical application of this research to be an either/or assertion. Indeed this study did not present conclusive evidence that would validate the modified form of dynamic CF as adequate in raising students' accuracy in writing, which is the primary objective of dynamic CFs application. As discussed above, this should certainly elicit additional examination, particularly as this conclusion is different from the previous research on the instructional strategy (Evans et al., 2011; Evans et al., 2010, Hartshorn et. al, 2010; Hartshorn, 2008; Lee 2008).

However, the data suggest advantages to both forms in ways that could complement each other if integrated.

The founding principles that informed the development of written dynamic CF are no less applicable to the effective feedback of oral production and this application of dynamic CF directly addresses many of the concerns voiced by opponents to oral grammar correction (Truscott, 1999). To the founding principles of manageability, meaningfulness, timely, and constant, from an oral skills perspective it is important that feedback be non-disruptive. Balancing the need to be non-disruptive, feedback should be immediate in giving learners prompt access to the original contextualized error. Student transcription promptly following the original production initiates this immediate feedback and perhaps to a greater degree than the traditional method fosters early introspection. Many learners in the treatment variation reported engaging in self-evaluation of both the form and content of the language production before submitting it to receive feedback from the instructor. This likely injects a necessary element of learner ownership over the corrective dialogue that is less likely with the instructor first-strike nature of the traditional written dynamic CF model.

Suggestions for Further Research

Full validation of this new modification to the dynamic CF model is still not complete. While the data here supports that there is no significant detriment to written accuracy, there was also no demonstrated benefit to written accuracy by the speech responsive dynamic CF treatment used in this study. The benefits to writing complexity and fluency certainly indicate the promise that the modified treatment can complement the traditional instructional strategy but the benefits to students' spoken accuracy, complexity, and fluency also need to be established. It is important to note that the benefits or detriments of written dynamic CF to speaking accuracy, complexity,

and fluency have also not yet been determined through a broad enough examination. The data collected at the onset of this study can be used to answer both sides of this investigation.

As was discussed in Chapter 2, the research study was devised to collect the raw data needed to answer all of the components of the following summarized research questions:

> When students in a linguistic accuracy class where the individual student output mode that informs the dynamic CF process is speaking rather writing, what are the differences in:
>
> A. demonstrated accuracy when writing?
> B. other demonstrated writing skill competencies including fluency, complexity and lexical variety?
> C. demonstrated accuracy when speaking?
> D. other demonstrated speaking skill competencies including fluency, complexity and lexical variety
> E. overall course satisfaction?

The data analysis presented here was only directed toward addressing components A and B. This narrowed focus was in part due to the investment of time and resources that will be required to adequately address components C and D. While the initial stages of this analysis have been completed and the analysis structure presented here provides a framework for the rest, this process will be considerably more intense and require additional innovations in order for it to be successfully completed.

Also, the research study conducted here included a post-semester questionnaire that did explore some aspects of affective impact that resulted from both the treatment and contrast variations of dynamic CF used in this study. As established by Lee (2009), face validity of all

aspects of the instructional strategy for the student and instructor participants is a critical component to the instructional strategy's success. It is unlikely that an unmotivated student will achieve a desirable outcome as the instructional strategy is rigorous and requires consistent engagement. How the instructional strategy can best be adapted to maximize motivation for the largest number of students in a class is a question that has been raised but not sufficiently addressed by previous studies. The rigor of this instructional strategy needs to be respected. Student and teacher burnout are risks that should be minimized through careful analysis on the amount of treatment that is necessary to achieve the desired effect. Complaints of rigor could actually be no more than complaints of repetition. Intersecting speech responsive and writing responsive dynamic CF cycles may maintain student interest for longer than a single skills approach has shown possible.

Among those questions about dynamic CF feedback and its implementation that have yet to be answered but should be a research priority is the question of instructional time. A possible cause of the reduction seen in raising student's accuracy may have been a reduction in class instruction time. While the actual production time of the contrast group was minimally expanded in this study (with the addition of 5 minutes of speech recording to control for practice effect), the treatment group lost as much as 20 minutes to complete their transcription. Also there was some time lost by both classes in moving between the classroom and computer lab. While the loss of time is easier to quantify, there were also some benefits reported by both teachers and students in favor of using computers. One immediate solution to consumption of in-class time by the production exercises would be for both the writing and speaking/transcribing portions of these methods to be done outside of the instructional hour. Future research could look to develop

ways to facilitate this and examine the impact of additional instructional time on the efficacy of the treatment in addressing accuracy needs.

Another suggestion for future research would be more in-depth examination into the relationship between lexical development and success with dynamic CF instructional strategy. This research could start with doing retroactive analysis on the previously gathered data for other dynamic CF studies using the same lexical analyses of this study to compare the lexical development of the participants in the previous studies during the course of the treatment. If these studies show that the students did not experience the same rapid fluctuations in the lexical content of their written production that was evidenced by the participants in this study, there would be some support for the hypothesis that the accuracy decline in these students' post-test performance was in part the result of being in a state of lexical flux. Gaining research-based support for this hypothesis would then motivate examinations into how lexical development and accuracy development could be better achieved simultaneously.

Finally, previously conducted studies have looked at the impact of dynamic CF instructional strategy on groups of students. Now that this quantity validated research has been done, there would be value in looking at students which do particularly well under the instructional strategy and students who fail to achieve similar benefits during a semester of study. This microanalysis of the impact of this instructional strategy could help identify characteristics in participation, person and interaction that lead to both success and failure. Once these characteristics are identified how those that lead to success can be magnified and how those that lead to failure can be minimized can be explored. While students who are not willing to put forth the required effort cannot be lead toward appreciable gains in accuracy, it is important that how the instructional strategy can best be tailored to meet individual learner traits

is understood. Thus, those traits that lead a learner toward success can be tapped and those that tend to trip an otherwise well-intentioned learner can be avoided.

Conclusion

The purpose of this study was to determine the effect of modifying a teaching method known as dynamic CF that has been shown to improve L2 writing accuracy to include an emphasis on student spoken production. A treatment group participated in a semester long Linguistic Accuracy course that utilized this modified version of dynamic CF. The treatment groups gains in terms of written accuracy, complexity, fluency, and lexical development were compared to gains in the same writing subskill areas by students in a contrast group which participated in the traditional writing focused version of dynamic CF. Students in the treatment group received feedback only on transcriptions of their speech while students in the contrast group received feedback only on ten-minute written responses to daily prompts.

The pre-test performances and the post-test performances of both students on 30-minute writing samples were analyzed using a mixed model ANOVA analysis. These repeated measures showed that while there was not a significant difference between the groups in terms of changes in overall accuracy, there were statistically significant advantages for students in the treatment group in terms of their writing fluency and complexity particularly for students at the less advanced end of the proficiency continuum included in the study. The analysis also showed no significant advantage to lexical development for students in either group. This study provided evidence that the modified treatment does not result in notable negative consequences to ESL learners' writing when contrasted with a more traditional application of dynamic CF instruction. This is an acceptable base from which to pursue further evaluation of the instructional strategy's impact on speaking accuracy, fluency, complexity and spoken lexical development.

This research initiates further exploration into both variations of dynamic CF and their impact on L2 English learners' spoken accuracy, fluency, complexity, and lexical development. Analysis will also be done that examines students' preferences for the two treatments and will lead to the pursuit of a dual-skills approach for dynamic CF. Achievement of greater levels of accuracy in both written and spoken production should be an important part of some English language learners' individual language learning plans, particularly those trying to access inaccuracy-sensitive language contexts including some professional applications and academics. It is hoped that the understandings achieved in this study will help inform improved instructional practices that can be employed to the benefit of these students and the instructors and institutions that serve them.

References

Anthony, L. (2009). AntWordProfiler 1.2000m Macintosh OSX. [software]. Retrieved from http://www.antlab.sci.waseda.ac.jp/antwordprofiler_index.html

Basturkmen, H. (2002). Learner observation, and reflection on, spoken discourse: An approach for teaching academic speaking. *TESOL Journal, 11*(2), 26-30.

Bates, L., Lane J., Lange E. (1993). *Writing Clearly: Responding to ESL Compositions*. Boston: Heinle & Heinle.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*(17), 409-431.

Breiner-Sanders, K. E., Lowe, P., Miles., & Swender, E. (1999). ACTFL proficiency guidelines speaking. from www.actfl.org

Bruton, A. (2010). Another reply to Truscott on error corrections: Improved designs over statistics. *System, 38,* 491-498.

Celce-Murcia, M. (1991). Grammar Pedagogy in Second and Foreign Language Teaching. *TESOL Quarterly, 25*(3), 459-480.

Chomsky, N. (1959). Review of Verbal Behavior. *Language, 35*(1), 26-58.

Cohen, A., Robbins, M. (1976). Toward assessing interlanguage performance: The relationship between selected errors, learners' characteristics and learners' explanations. *Language Learning*(26), 45-66.

Cohen, A., Calvacanti, M. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), Second language writing: Research insights for the classroom (pp. 155-177). Cambridge: Cambridge University Press.

Cohen, A. D. (1987). Student processing of feedback on their compositions. In J. R. A. Wenden (Ed.),

Learner strategies in language learning. Englewood Cliffs, NJ: Prentice-Hall.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates

Corder, S. P. (1981). *Error analysis and interlanguage*. London: Oxford University Press.

DeKeyser, R. (1997). Beyond explicit rule learning: automatizing second language morphosyntax. *Studies in Second Language Acquisition*. 19. 195-221.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

Dornyei, Z. (2000). Motivation In Action: Toward a Process-oriented Conceptualization of Student Motivation. *The British Journal of educational psychology, 70*(4), 519-538.

Dornyei, Z., Csizer, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*. 2, 203-209.

Doughty C., Varela, E. (1998). Communicative focus on form. In C. Doughty, J. Williams (Ed.), *Focus on form in classroom second language acquisition*. New York: Cambridge University Press.

Educational Testing Services. (2004). iBT/Next generation TOEFL test: Independent speaking rubrics Educational Testing Services. Retrieved from http://ets.org/TOEFL/

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implict and explicit language acquisition. *Studies in Second Language Acquisition, 24*(143-88).

Ellis, R. (1998). Teaching and research: Options in grammar teaching. *TESOL Quarterly, 32,* 39-60.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*(83-107).

Eskey, D. E. (1983). Meanwhile, back in the real world...: Accuracy and fluency in second language teaching. *TESOL Quarterly*(17), 315-323.

Evans, N. W., Hartshorn, K.J., McCollum, R.M., Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research, 14*(4), 445-463.

Evans, N. W., Hartshorn, K.J., Strong-Krause, D. (2011). The efficacy of dynamic CF for university-matriculated ESL learners. *System*.

Ferris, D. R. (1995). Can advanced ESL students be taught to correct their most serious and frequent errors? *CATESOL Journal, 8*, 41-62.

Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*, 33-53.

Ferris, D. R., Harvey, H., Nutall, G. (1998). *Assessing a joint training project: Editing strategies for ESL teachers and students.* Paper presented at the Association for Applied Linguistics, Seattle, WA.

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

Folse, K. S., Brummett, K. (2006). Pedagogical Grammar Courses Offered by MATESOL Programs in Floriday. *Sunshine State TESOL Journal* 1-12.

Folse, K. S. (2011). *Your Options in Teaching English grammar in 2011 and Beyond.* Paper presented at the TESOL, New Orleans. Mar 19, 2011.

Foster, P., Tonkyn, A., Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics, 21*, 345-375.

Gardener, R. C., MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning, 43*, 157-194.

Gardener, R. C., MacIntyre, P. D. (1993). A student's contributions to second language learning: Part II. Affective variables. *Language Teaching, 26*, 1-11.

Gass, S. M. (1983). The development of L2 intuitions. *TESOL Quarterly, 17*, 273-291.

Green, P. S., Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics, 13*, 168-184.

Hartshorn, K. J. (2008). *The effects of manageable error feedback on ESL writing accuracy.* Brigham Young University, Provo.

Hartshorn, K. J., Evans. N.W. Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., Anderson, N. J., (2010). Effects of Dynamic CF on ESL Writing Accuracy. *TESOL Quarterly, 44*(1), 84-109.

Hayes, M. F., Daiker, D. A. (1984). Using protocol analysis in evaluating responses to student writing. *Freshman English News, 13*(2), 1-10.

Hedgecock, J., Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback on L2 writing. *The Modern Language Journal*(80), 287-308.

Hendrickson, J. M. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *Modern Language Journal*(64), 216-221.

Hendrickson, J. M. (1980). The treatment of error in written work. *The Modern Language Journal, 64*, 216-221.

Hillocks, G., Jr. (1986). *Research on written composition: New Directions for teaching.* Paper presented at the ERIC Clearinghouse on Reading and Communication Skills/National Conference on Research in English, Urbana, IL.

Huang, S. (2008). Raising learner-initiated attention to formal aspects of their oral production through transcription and stimulated reflection. *International Review of Applied Linguistics in Language Teaching, 46*, 375-392.

Huck, S. W. (2008) *Reading statistics and research* (5th ed.). Boston: Pearson Education, Inc.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Urbana , IL: The National

Council of Teachers of English.

Joubert, Joseph. (1867) *Some of the "thoughts of Joseph Joubert*. (G.H. Calvert, Trans.). Boston: William V. Spencer.

Knoblauch, C., Brannon, L. (1981). Teacher commentary on student writing: The state of the art. *Freshman English News, 10*(2), 1-4.

Krashen, S. (Ed.). (1992). *Comprehensible input and some competing hypotheses*. Ottawa, Canada: University of Ottawa Press.

Krashen, S., Terrell, T. (1983). *The natural approach*. Hayward, CA: Alemany Press.

Lalande, J., II. (1982). Reducing composition errors: An experiment. *The Modern Language Journal, 32*, 257-261.

Larsen-Freeman, D., Thewlis, S., (2007). *Grammar Dimensions 3: Form, Meaning, And Use*. Boston, MA: Heinle.

Leeman, J. (2007). Feedback in L2 learning: responding to errors during practice. In R. DeKeyser (Ed.), *Practice in a Second Language* (pp. 111-135). New York: Oxford.

Leki, I. (1990). *Coaching from the margins: Issues in written response*. Cambridge: Cambridge University Press.

Lightbown, P. M. (1998). The importance of timing in focus on form. In W. Doughty C., J. (Ed.), *Focus on form in classroom second language acquisition* (pp. 177-196). New York: Cambridge University Press.

Long, M. H. (Ed.). (1977). *Teacher feedback on learner error: Mapping cognitions.*Washington, DC: TESOL.

Lynch, T., Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research, 4*, 221-250.

Lynch, T. (2001). Seeing what they meant: Transcribing as a route to noticing. *ELT Journal, 55*(2), 124-132.

Lyster, R., Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*, 37-66.

Lyster, R., Lightbown, P. M., Spada, N. (1999). A response to Truscott's 'what's wrong with oral grammar correction.' *Canadian Modern Language Review, 55*(4), 11.

Mennim, P. (2003). Rehearsed oral output and reactive focus on form. *ELT Journal, 57*(2), 130-138.

Nation, I. S. P., Heatley, A. (1994). Range: A program for the analysis of vocabulary in texts [software]. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

Ortega, L. (2009). *Understanding second language acquistion*. In B. Comrie & G. Corbett (Eds.) Understanding Language Series. London: Hodder Education.

Plunkett, K. (1995). Connectionist approaches to language acquistion. In P. Fletcher, B. MacWhinney, (Ed.), *The handbook of child language* (pp. 36-72). Oxford: Blackwell.

Polio, C., Fleck, C., Leder, N. (1998). "If I only had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*(7), 43-68.

Radecki, P. M., Swales, J. M. (1988). ESL student reaction to written comments on their written work. *System, 16*, 355-365.

Reid, J. M. (1997). *Responding to ESL student language problems: Error analysis and revision plans*. Boston: Heinle & Heinle.

Roberts, M. A. (1995). *Awareness and the efficacy of error correction*. Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Center.

Santos, T. A. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly, 22*(69-90).

Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals*(17), 195-202.

Sheppard, B. (2011). *Learning to Listen to Ourselves: Techniques for Speaking more accurately.* Paper presented at the TESOL Convention, New Orleans. Mar. 18, 2011.

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** (2009), 556–569.

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

Skehan, P., Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction*. New York: Cambridge University Press.

Spada, N., Lightbown, P.M.(1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition, 15*, 205-224.

Sorace, A. (1985). Metalinguistic knowledge and language use in acquisition poor environments. *Applied Linguistics*(6), 239-234.

Stillwell, C., Curraba, B., Alexander, K., Kidd, A., Kim, E., Stone, P., Wyle, C. (2009). Students transcribing tasks: Noticing, fluency, accuracy, and complexity. *64, 4*(445-455).

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

Truscott, J. (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review, 55*(4), 437-456.

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. H., Hsu A.Y. (2008). Error Correction, revision, and learning. *Journal of Second Language Writing, 16*, 292-305.

Truscott, J. (2010), Some thoughts on Anthony Bruton's critique of the correction debate, *System* **38**. 329–335

Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Wilkins, D. (1076). *Notional syllabuses*. Oxford: Oxford University Press.

Winitz, H. (Ed.). (1981). *The comprehension approach to foreign language instruction*. New York: Newbury House.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly. 19,* 79-101.

Appendix A Sample of Participant Writing After Clause and Error Identification

ID Number	Sample	# of Words	# of Clauses	# of T Units	ERD Clauses	ERF Clauses
76732	E	323	46	27	26	26

#T	#C	Clause
1	1	Life in another country may be painful.
2	2	If people moved to another country,
	3	they'd better follow the local customs
	4	Because that is the best way to get ride off the pain.
3	5	It's easy to feel that
	6	you are all alone
	7	when you are far away from home and come to a new enviornment.
4	8	You would feel that
	9	nothing is right.
5	10	The culture shock is a pain, and
6	1	I still remember the first week
	2	when I got here.
7	3	When I first get to America,
8	4	it was a really hard time
9	5	. I was looking for Chinese restaurants everywhere, and complained that
10	6	the food is not the way
	7	it tasted in China.
1	8	I was depressed all days.
2	9	Luckly, my host family is very nice to me,
3	20	they tried to cook different meals everyday
	1	so that I could find some food that
	2	I like here. And
4	3	I was surprised that
	4	some food here dose tasted good.
5	5	I didn't know how to deal with people, either
6	6	Because the way---are so different.
	7	we do things
7	8	While, you can't change all people around you,
	9	then you must change yourself.
8	30	So I tried to learn the customs through the way
	1	they do it.
9	2	I went to church with them, and tried to do my lundary once a week, which
	3	I used to do everyday with my hands.
20	4	And a week later, I am able to enjoy the life here.
1	5	One of my friends is also studying in foreign country, and
2	6	he always told me
	7	how hard the life is.
3	8	I do believe him,

	9	because he's trying to be a complete Chinese in Canada.
4	40	When you feel bad, --- means that you need do some changes.
	1	--that means you are not doing it right---
5	2	To Change a country or a culture maybe need millions of years,
	3	but to change a person maybe just need 10 seconds.
6	4	Change yourself to follow the local customs, and
7	5	you will find
	6	the life is much easier.