Durham E-Theses

The Study of Teacher Written Feedback: The Effectiveness of Electronic Feedback on Student Writing Revisions

MA, BRUCE,WAI,LEUNG

How to cite:

MA, BRUCE,WAI,LEUNG (2017) *The Study of Teacher Written Feedback: The Effectiveness of Electronic Feedback on Student Writing Revisions* , Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/12018/

Use policy

Durham University
School of Education

Doctor of Education Thesis

The Study of Teacher Written Feedback: The
Effectiveness of Electronic Feedback
on Student Writing Revisions

Submitted by

MA Wai Leung Bruce
March 2017

Content	
	Page
Chapter 1 **Abstract**	7
Chapter 2 **Introduction**	9
2.1 ***Feedback and its Significance on Second Language Writing***	9
2.2 ***Background of Teacher Feedback on Writing***	13
2.3 ***Purpose of the Study***	16
Chapter 3 **Literature Review**	19
3.1 ***Theoretical Framework and Language Theories***	19
3.1.1 Theoretical Framework	20
A. Definition of Autonomy	20
B. Role of Autonomy in Language Teaching and Learning	21
C. Nunan's Model of Framework for Developing Learner Autonomy	22
D. Implications for Developing Learner Autonomy in Writing Feedback: The Three-Step Approach to Error Correction within the Model of Framework for Developing Learner Autonomy	24
3.1.2 Language Theories	28
A. The Role of Error Correction	28
B. Krashen's Input Hypothesis	31
C. Implications for the Electronic Feedback ('Mark My Words')	34
3.2 ***Major Issues on Teacher Written Feedback***	39
3.2.1 Focus of Teacher Written Feedback	41
3.2.2 Teacher's centeredness of Teacher Written Feedback	41
3.2.3 Effectiveness of Teacher Written Feedback	43
A. Error Feedback & Misinterpretation of Findings in other Studies	43
B. Teacher Commentary	53
C. Comprehensibility of Teacher Written Feedback	54
D. Treatment of Different Error Types	57
E. Alternative Feedback Treatments	58
F. Assessment for Learning or Assessment of Learning	61
- Implications for Teacher Written Feedback	64
G. Students' Perceptions of Teacher Written Feedback	67
H. Contextualization of Feedback Studies	71
3.3 ***Linguistic Differences between English and Chinese***	80
3.3.1 Theory of Language Typology	81
3.3.2 Theory of Markedness	84

Acknowledgements

I am very thankful to my family members who gave me tremendous support in my pursuit of the Doctorate in Education degree with the University of Durham. Their love and caring on me are beyond any words to describe.

Chapter 1   Abstract

The effectiveness of teacher written feedback has been a subject of debate in second language writing for decades. The most basic debate in this area among ESL writing researchers is whether teacher written feedback in various forms has any positive effects on student writing revisions. Among other researchers, Ferris, Lee, Ene & Upton and Stevenson & Phakiti argued that while the effectiveness of error feedback in the traditional paper-and-pen form (Ferris, 1999, 2002, 2003, 2004, 2006 and Lee, 2008a, 2008b), computer-facilitated form (Ene & Upton, 2014) or computer-generated form (Stevenson & Phakiti, 2013) was not conclusive, more research should be done to explore in what ways error feedback can be improved.

Indeed, the heterogeneity of these studies characterized by different focus, research designs, institutional and instructional contexts, and participant backgrounds, alongside some methodological flaws and misinterpretation of findings identified in my critical review has possibly undermined the validity and reliability of the studies, giving rise to these mixed results for both paper-and-pen feedback and computer-based feedback. As such, the causality between different forms of feedback treatment and their outcomes of error reduction is questioned.

With the primary interest in improving the effectiveness of teacher written feedback in error correction, 'Mark My Words' ('MMWs'), the interactive-based electronic feedback system, was designed in such a way to accommodate individual learners' language needs and to be more responsive to various error types. This study focused on examining on the effectiveness of 'Mark

My Words' ('MMWs'), as a kind of computer-facilitated feedback (i.e. electronic feedback), in improving students' error reduction in their writing revisions, under a controlled condition.

The mixed methods approach was adopted, namely the 'error count' method and 'questionnaire', in this study. The participants were 62 second-year engineering students enrolled in an English for Specific Purposes course in a Hong Kong University. Efforts were made to avoid the impact of extraneous variables on the validity and reliability of the research outcomes under such controlled condition.

The positive results of this study can contribute some sort of concrete evidence to the growing body of literature of the 'effectiveness of teacher written feedback' and 'second language writing', thus clarifying some mixed results of the previous research.

Chapter 2   Introduction

2.1   Feedback and its Significance on Second Language Writing

Feedback can be defined from different perspectives. According to Hattie & Timperley (2007), feedback can be defined as "Information provided by an agent with respect to one's performance or understanding" (p. 81). They further added that "A teacher or parent can provide corrective information, a peer can provide an alternative strategy, a book can provide information to clarify ideas, a parent can provide encouragement, and a learner can look up the answer to evaluate the correctness of a response. (ibid, p. 81)

From the teaching and learning perspective, Winne and Butler (1994) stated that "Feedback is information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies" (p. 5740). Some researchers postulate that feedback in learning and teaching is beneficial for learners (e.g. Bitchener, 2008; Evan, Hartshorn, & Strong-Krause, 2011; Leki, 1991).

From the psychological perspective, feedback is generally regarded as crucial for both encouraging and consolidating learning (Anderson, 1982; Brophy, 1981; Vigotsky, 1978), with its significance being "recognized by those working in the field of second language writing", "acknowledged in

process-based classrooms, where it forms a key element of the students' growing control over composing skills", and "acknowledged by genre-oriented teachers employing scaffolding techniques" (Hyland & Hyland, 2006 p. 1).

Situating feedback under the context of teaching second language writing, Hyland & Hyland (2006) argue that the provision of feedback to second language learners in a writing class, be it teacher written feedback, peer feedback, writing conference, or computer-mediated feedback is being regarded as one of the English teachers' most important tasks, for an effective feedback could help "create a supportive teaching environment for conveying and modeling ideas about good writing, for developing the ways students talk about writing, and for mediating the relationship between students' wider cultural and social worlds and their growing familiarity with new literacy practices" (p. xv).

Therefore, in the context of second language writing, provision of effective feedback is crucial to assist learners in facilitating their learning process, evaluating their performance, identifying their incorrectness, consolidating their understanding, sensitizing themselves to different genres as well as taking suggestions for improvements, over the process-based writing.

Apart from the support from the above empirical studies, the efficacy of teacher written feedback has been supported by some theories of second language acquisition. According to the 'Interaction Hypothesis' postulated by Long (1996), it is postulated that corrective feedback which creates opportunities for interaction and meaning-negotiation is hypothesized to support language

acquisition through which it provides learners the input needed for the second language acquisition and the input provided is made more comprehensible for the learners. In addition, Ene & Upton (2014) stated that "the value of feedback is connected to 'Schmidt's Noticing Hypothesis' (p.80), which holds that the "SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the importance or significance of noticed input to be" (Schmidt, 2001, p, 3-4). This means feedback is being valued as an input for SLA in that it can enable second language learners to focus on the linguistic input and notice the gap between the target language and their interlanguage. Lastly, the value of feedback is supported by Swain's 'Output Hypothesis' (1985) where it is postulated that pushing second language learners in the production of challenging output would raise their awareness of the gap between the target language and their interlanguage.

Despite the value of feedback being acknowledged from various perspectives, the 'actual' effectiveness of teacher written feedback has always been a subject of debates for decades. Such debate was further fueled up by Truscott (1996) who took an opposition position against the value of error feedback. The debate has continued till now, but the focus has been shifted to what types of corrective feedback work more effectively in remedying errors in the short term and long term (e.g. Truscott & Hsu, 2008; Bitchener & Knoch 2010b; Bitchener & Ferris, 2012). The controversy of teacher written feedback from the 'theoretical' perspective will be explored in Chapter 3.1.2 (A) 'The Role of Error Correction'. In addition, major 'practical' issues surrounding different strategies, foci and forms of teacher feedback practice and their implications on their effectiveness will be discussed in great details in Chapter 3.2 'Major Issues on Teacher Written Feedback'.

Before proceeding to the rest of the thesis, there are some terms which need to be clarified. There are feedback on an array of subject disciplines, and the feedback being reviewed and examined in this research study here is on second language learning, in particular on second language writing. Feedback on second language writing is usually referred in research studies as 'Teacher Feedback' (e.g. Enginarlar, 1993; Arndt, 1993), 'Teacher Written Feedback' (e.g. Hyland, 1998; Lee, 2008a & 2008b), 'Error Feedback' (e.g. Ferris, 2006; Ferris & Roberts, 2001; Lee, 2008a & 2008b; Rennie, 2000), or 'Corrective Feedback' (Bitchener & Knoch, 2009, 2010b; Ferris, 2010, 2012 & 2015; Bitchener & Ferris, 2012). Despite the differences in their names, basically they are all referring to the teacher written feedback on second language writing, and they seem to be interchangeable sometimes or being used together in the same studies, unless the terms are specified by a researcher like 'error feedback' in Lee's study (2008a & 2008b). This study mostly used the term 'Teacher Written Feedback' in a holistic sense which included the provision of grades, indication of errors and written commentary; and used the term 'Error Feedback' in a narrower sense when it came to the indication of errors, be it direct or indirect, coded or uncoded, on the areas of 'content', 'organization', 'language' and etc. The reason for employing these two terms was grounded on the fact that the methodologies adopted in this study were modeled on Ferris's study (2006) and Lee's studies (2008a & 2008b), who used the terms 'Teacher Written Feedback' and 'Error Feedback'. Recently, it has been noted that there is a trend to consistently use 'Corrective Feedback' (CF) as an overriding and holistic term when compared to others. CF refers to "an indication to the learners that his or her use of the target language is incorrect" (Lightbown and Spada , 1999, p. 172). Researchers have suggested that CF is associated with second language (L2) learning, because it leads learners to notice L2 forms (Bitchener & Knoch, 2010; Loewen & Erlam, 2006; Lyster & Mori, 2006; Varnosfadrani & Basturkmen, 2009). There

are different types of CF (or error feedback) which are 'direct feedback and 'indirect feedback'; and then the category of 'indirect feedback' is subdivided into 'indirect coded feedback' and 'indirect coded feedback'. Further explanation of these categories will be detailed in Chapter 4.3 'Clarification of Terms'.

2.2   *Background of Teacher Feedback on Writing*

The effectiveness of teacher feedback on writing has been always a subject of debate in second language writing for decades. The most prevalent debate in this area among ESL writing researchers and teachers was whether the teacher written feedback has any positive effects on student writing development (Ferris, 2006). The most notable argument against the value of giving teacher written feedback was represented by Truscott (1996) who argued that "practical problems such as teacher limitations and student inattention render even the hypothetical benefits of grammar correction unlikely (cited in Ferris, 2006, p. 82), concluding that in the end student writers cannot benefit at all. Ferris (1999, 2002, 2003, 2004 & 2006) argued that while the effectiveness of error feedback was not conclusive, more research should be done to explore if error feedback, as an integral part of teacher written feedback, should be embraced or abolished. Hyland and Hyland (2006) also concluded that little consensus about the kinds of teacher feedback in writing are effective, especially about the long-term effects of teacher written feedback on students' writing development. Thus, the question of what form, type and strategy of feedback is most beneficial to students deserves further scrutiny in this study.

A number of research studies (e.g. Willetts; 1992; Egbert, 2002; Milton, 2006; Lee, 2008a; Stevenson & Phakiti, 2013; Ene & Upton, 2014) suggested that integration of technology can improve academic performance, enhance motivation, and promote learning. Therefore, the use of technology might offer an alternative way of giving teacher written feedback in student writing more effectively and efficiently. To put this into perspective, the pedagogical use of computer-based educational technology for providing students with feedback on their writing is commonly known as 'Automated Writing Evaluation' ('AWE') (Stevenson & Phakiti, 2013). A few 'AWE' software programs have been developed, for example 'Criterion' by ETS, that can automatically detect and identify certain error types, mostly related to grammar, mechanics and organization errors, for student-users. And most of them can provide holistic scores for the writing texts (ibid). However, the critical review of thirty journal articles on evaluating ten 'AWE' software programs conducted by Stevenson & Phakiti (2013) suggested that the effectiveness of 'AWE' feedback on student writing is not conclusive due to 'paucity of research, heterogeneity of existing research, the mixed nature of research findings, and the methodological issues in some of the existing research' (p.62).

Against the above backdrop, this study focuses on examining on the effectiveness of the interactive-based electronic feedback system named 'Mark My Words' ('MMWs'), as a form of feedback treatment adopted in this experimental study, in improving students' error reduction during their writing revision process. This focus premises on the ground that the capability of a particular form of feedback treatment in improving the quality of students' revisions is central to claims made about:

(1) the effectiveness of teacher writing feedback in error reduction;

(2) the effectiveness of 'MMWs' feedback, as a form of feedback treatment, in error reduction.

It is noteworthy to point out that 'Mark My Words' ('MMWs') is not a fully automated writing evaluation tool. Unlike those 'AWE' software which can automatically detect and identify language errors based on techniques such as artificial intelligence, natural language processing and latent semantic analysis (Stevenson & Phakiti, 2013), the pedagogical functions of 'MMWs' is solely operated by teacher-users via an add-on tool bar specially designed for Microsoft Word. This add-on tool bar enables teacher-users to manually insert pre-set comments alongside boilerplate 'resource-rich' feedback into students' written assignments via the web-based platform (please refer to Chapter 3.4.2: Pedagogical Operation of 'Mark My Words'). Despite the difference between 'MMWs' and 'AWE' in terms of their technological level of advancement, the reason for the inclusion of 'AWE' into the literature review and discussion of this study are that, (1) if we take a broader scope and a looser definition in examining the use of technology in teacher writing feedback, both 'MMWs' and 'AWEs' can be viewed as under the same category of 'computer-generated feedback on writing', in which both of them involve the integration of technology in supporting teacher feedback on writing; and (2) the methodological designs of those 'AWE' research studies and their research outcomes provide this study much insights as to how to improve the quality of the electronic feedback under limited resources (unlike those 'AWE' programs which are mostly institutionally funded or commercially funded for the profit-making purpose), as well as to enhance the validity and reliability of the research design in this experimental study (which will be explained in greater detail in Chapter 3.5 The Use of Technology in Teacher Feedback on Writing and Chapter 5 Research Methodology).

2.3   Purpose of the Study

The title of my research is "The Study of Teacher Written Feedback: The Effectiveness of Electronic Feedback on Student Writing Revisions". This research is an experimental study of the effects of teacher written feedback on student writing revisions over 7 error categories with two different forms of feedback treatment (i.e. electronic feedback vs. traditional paper-based feedback), in improving the overall accuracy of student writing. This experimental study is a reflection of some theoretical considerations about Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' in the area of the second language writing. The 7 error categories covering 77 error types are appended in *Table 2.1* on the next page (or see Appendix A):

Table 2.1   Seven Error Categories Covering 77 Error Types

	Error Categories	Error Types
1.	"Awkwardness" Errors	Non-treatable clausal-level errors (e.g. Chinglish)
2.	"Clausal-level" Errors	Mixed construction; Imbalanced Structure; Sentence fragment; Run-on sentence; Wrong relative pronoun reference
3.	"Word-level" Errors	Redundant determiner; Missing determiner; Misuse determiner; Subject-verb agreement; Redundant word; Tense errors; Singular-Plural form; Wrong Preposition; Voice; Punctuation; Wrong word/expression; Wrong verb form; Adjective-adverb confusion; Noun-Adjective confusion; Verb-Noun confusion; Adjective-verb confusion; Missing auxiliary verb; Misuse auxiliary verb; Redundant preposition; Missing preposition; Missing noun/pronoun; Missing possessive; Finite vs. Non-finite verbs; Misuse of 'besides'; 'Comparing' vs. 'Compare'; Redundant conjunction; Wrong adverb/connective; Wrong adverb form; Intransitive verb; Missing adverb/connective; Misuse 'concern'; Missing verb; Wrong form; Missing word; Misuse of 'on the other hand'; Redundant adjective; 'Be'-'Do' confusion; 'Be'-'Have' confusion; Auxiliary or Non-finite; Capitalization; Wrong relative pronoun; Misuse of 'consist'; Subordinate

		clause/conjunction; Missing infinitive-'to'; 'Due to' vs. 'Because'; 'Word-order in question'; 'Negative form'; Misuse of 'on the contrary'; Superlative confusion; Spelling.
4.	"Collocation" Errors	Wrong Word Collocation (including wrong preposition).
5.	"Tone & Style" Errors	Inappropriate register; Level of courtesy/ confidence; Cliché.
6.	"Content" Errors	Too general / unclear; Logic flaw; Cause & Effect problem; Irrelevance; Inadequate support; Ambiguous.
7.	"Organization" Errors	Choppy ideas (cohesion); Choppy ideas across paragraphs (coherence); Missing / Inadequate topic sentence; New sentence; New paragraph; Inadequate / Missing a thesis statement; Unclear topic sentence; Missing / Inadequate scene setting; Supporting details do not follow the topic sentence; Reasoning is difficult to follow; Missing/Ineffective conclusion.

With the primary interest in improving the effectiveness of teacher feedback on writing while cultivating students' autonomy in their language development, the electronic feedback system called 'Mark My Words' ('MMWs') was designed, by the IT Team of the Center for Language Education at HKUST in which I was one of the members, in such a way to accommodate individual learners' interlanguage level and to be be more responsive to various error types.

'Mark My Words' ('MMWs') is an add-on tool bar for Microsoft Words, which enables teachers to insert preset comments into students' written assignments. Teachers can retrieve the assignments inside MSWord and use 'Mark My Words' to insert boilerplate 'resource-rich' comments as feedback, without having to rewrite the students' texts. These comments automatically direct students to detailed online explanations, lookup tools, etc. As such, 'Mark My Words', which is a computer-based educational technology, also regarded as a kind of Automated Writing Evaluation ('AWE') tool.

Given the primary interest above, the purposes of this study are (1) to investigate if and in what ways students' writing revisions were influenced by the electronic feedback and paper-based feedback; (2) to examine to what extent revision changes on each error category were made by students receiving the electronic feedback and the paper-based feedback respectively, (3) to compare and analyse students' perceptions of the effectiveness of teacher written feedback in improving the overall accuracy and appropriateness of their writing, and (4) to contribute to our growing understanding on how teacher written feedback can enhance the writing development of student writing through their responses to teacher written feedback.

In order to address the above, this project was undertaken by comparing and analyzing the ways students respond to and act on, respectively, the electronic feedback treatment and the traditional paper-based feedback treatment in a specific and identical context. It is hoped that this study can contribute to our growing understanding of the effects of teacher feedback practice on student writing revisions under two different feedback forms of treatment.

Chapter 3   Literature Review

In this chapter, the literature review plays a significant role in justifying the need for as well as identifying a direction and dimensions for the study. The literature review is divided into five parts. The first part identified a theoretical framework and language theories underpinning the rationales behind the electronic feedback practice ('MMWs'), not only in terms of its notion and values, but also its pedagogical operation in error correction (see Chapter 3.1). The second part summarized and discussed a range of prominent issues pertaining to teacher written feedback starting from 1980s, in an attempt to identify the knowledge gaps which were yet to be filled or narrowed down in this research study (see Chapter 3.2). The third part discussed the linguistic differences between English and Chinese (see Chapter 3.3). The fourth part discussed the pedagogical implications of these linguistic differences on the second language writing of Hong Kong ESL students (see Chapter 3.4). The fifth part examined the use of technology in teacher feedback on writing (see Chapter 3.5).

3.1   Theoretical Framework and Language Theories

The first part of the literature review aims at identifying a theoretical framework and language theories on which the rationales guiding my 'Three-Step Approach to Error Correction' is grounded. The theoretical framework scaffolding the pedagogical operation of 'Mark My Words' ('MMWs') is a model postulated by Nunan (1997) in promoting learner autonomy. Within this three-step pedagogical framework promoting learner autonomy in addressing language errors, the

19

'Role of Error Correction' and 'Krashen's Input Hypothesis', which are the second language acquisition theories, form the guiding principles in error correction, such as the level of explicitness needed for error identification and explanation.

3.1.1. Theoretical Framework

Nunan's (1997) model of framework, which postulates various levels for developing autonomy, is incorporated into my writing feedback practice through the pedagogical operation of 'Mark My Words' ('MMWs'). To understand how Nunan's model of framework is being represented in such pedagogical operation alongside its implications for developing learner autonomy in responding to writing feedback, we have to first understand the notion of autonomy and its role in language teaching and learning.

A.   Definition of Autonomy

The history of autonomy has begun with the publication of Holec's (1981) seminal report (Benson, 2006). In this report, Holec (1981) defines autonomy as 'the ability to take charge of one's own learning' (p. 3). On the other hand, autonomy is also described as 'the situation in which the learner is totally responsible for all of the decisions concerned with his learning and the implementation of those decisions' (Dickinson, 1987, p. 11). While Holec's definition (1981) treats autonomy as an attribute of a learner and Dickinson's definition treats autonomy as a learning situation, "Holec's definition of learner autonomy has proved remarkably robust and remains the most widely cited definition in the field" (Benson, 2006, p. 23). However, Holec's definition only

"explained what autonomous learners are able to do, it did not explain how they are able to do it" (ibid, p. 23).

B.   Role of Autonomy in Language Teaching and Learning

Allwright (1988) associates the idea of learner autonomy with 'a radical restructuring of language pedagogy' that involved 'the rejection of the traditional classroom and the introduction of wholly new ways of working' (p. 35). The proliferation of self-access centres in the 1990s as well as the recent developments related to computer-based modes of teaching and learning are the examples (Benson, 2006). Such 'deconstruction of conventional language learning classrooms' has blurred 'the distinction between classroom and out of classroom applications, leading to new and often complex understandings of the role of autonomy in language teaching and learning' (ibid, p.22).

However, the role of autonomy should not be conceptualized as an unconstrained freedom in student learning, as "most learners do not come in the learning situation with the knowledge and skills to determine the content and learning processes which will enable them to reach their objectives in learning another language" (Nunan, 1997, p.201).

According to Benson (2006), for learners to develop their autonomy many advocates of autonomy argue that some degree of freedom in learning is central to the development of their autonomy; However, resonating Nunan's view above, "freedom in learning is not the same thing as autonomy and this freedom will always be constrained" (ibid, p. 23). In other words, freedom in learning should be conceptualized as the provision of learning options rather than an unconstrained autonomy, where language teachers should shoulder the role of providing various learning options

catering to diverse learning needs and preferences. On the other hand, Nunan (1997) postulates three basic assumptions underlying a model of framework with respect to the notion of autonomy (p. 192). They are as follows:

(1) Developing some degree of autonomy is essential if learners are to become effective language users,

(2) The ability to direct one's own learning can be developed through pedagogical procedures of one sort of another.

(3) Autonomy is not an all-or-nothing concept, that there are degrees of autonomy, and that the extent to which autonomy can be developed will be constrained by the context in which the learning takes places.

C.   Nunan's Model of Framework for Developing Learner Autonomy

In the late 1990s, a number of studies attempted to define the notion of autonomy as a matter of degree (Benson, 2006). Nunan (1997) postulates a framework proposing five levels for encouraging learner autonomy in relationship to use of learning materials (p. 195). These five levels, which map out the sequence of language development, entail (1) awareness, (2) involvement, (3) intervention, (4) creation, and (5) transcendence under the dimensions of 'content' and 'process'.

At the 'awareness' level, learners are 'made aware of the pedagogical goals and content of materials' under the dimension of content, and 'identify strategy implications of pedagogical tasks'

and 'identify their own preferred learning styles / strategies' under the dimension of process (ibid, p.195)

At the 'involvement' level, learners are involved in 'selecting their own goals from a range of alternatives on offer' under the dimension of content. They then 'make choices among a range of options' under the dimension of process (ibid, p.195)

At the 'intervention' level, learners are involved in 'modifying and adapting the goals and content of the learning program' under the dimension of content. They could 'modify or adapt tasks' under the dimension of process (ibid, p.195).

At the 'creation' level, learners can 'create their own goals and objectives' under the dimension of content, and then 'create their own tasks' under the dimension of process (ibid, p.195). For example, learners are expected to generate written texts accompanied by a set of comprehension questions for others to answer (ibid, p.199).

At the 'transcendence' level, learners would 'go beyond the classroom and make links between the content of classroom learning and the world beyond' and 'become teachers and researchers' (ibid, p.199).

Nunan's (1997) model represented a continuum through which "learner autonomy can be worked towards, from relatively superficial awareness raising through to complete autonomy where learners transcend the classroom and take complete charge of their own learning" (ibid, p. 193).

Having said that, Nunan points out that "these levels involve considerable overlap, and that, in practice, learners will move back and forth between levels" (ibid, p. 195).

Nunan (1997) argues that his model illustrated how 'autonomy can be a normal, everyday addition to regular instruction' and 'how far one goes, or wants to go, in encouraging autonomy will be dictated by the contexts and environments in which the teaching and learning takes place' (p. 201). This argument lends weight to Farmer & Sweeney (1994) who argue that "autonomy is not an absolute but a relative term, and the degree of autonomy may vary from one context to another" (p. 139).

Since Nunan (1997) model of framework serves as an illustrative model for the use of learning materials in developing autonomy, adding to the above Farmer & Sweeney's (1994) argument which stated that the exercise of autonomy depends very much on particular contexts, the proposed model should not be regarded as an absolute model for any design of language teaching materials, let alone the practice of writing feedback which aims for promoting learner autonomy in this experimental study.

D. Implications for Developing Learner Autonomy in Writing Feedback: The Three-Step Approach to Error Correction within the Model of Framework for Developing Learner Autonomy

The framework scaffolding the pedagogical operation of 'Mark My Words' (MMWs) to a certain extent mirrors Nunan's (1997) model of framework for encouraging learner autonomy, in the

belief that 'developing some degree of autonomy is essential if learners are to become effective language users' (p.192). The only difference is that the model is incorporated into my writing feedback practice through the pedagogical operation of 'Mark My Words' ('MMWs'), rather than into teaching materials as the model primarily intended to do.

By taking learners through level one to level three of Nunan's model, 'Mark My Words' ('MMWs') partially incorporates this range of autonomy into its pedagogical operation in the following fashion:

The first level simply attempts to make students aware of the goals, types of feedback information and strategy implications underlying a range of pathways offered by the electronic feedback. It is a first step towards encouraging their autonomy. Before the electronic feedback is given, an instructor will demonstrate to learners the objectives of writing feedback, the range of pathways available, the functions of each pathway ranging from 'less explicit and less detailed' feedback to 'more explicit and more detailed' feedback alongside their corresponding pedagogical goals, and the strategies they can employ.

The second level allows learners to move from awareness to involvement in responding to the writing feedback by making choices from a range of feedback options which include: (1) Error identification and a brief explanation in the pop-up window, (2) Provision of a recommended lexico-grammatical form or an example sentence in another context through the web-based resources, and (3) An electronic English Grammar Guide which provides a more detailed

explanation of the identified error type, and then followed by some practice. Learners are encouraged to exercise their choice according to their learning needs and preferred learning styles.

The third level allows learners to intervene in their learning by acting on the writing feedback. That is to make corrections and if needed, to modify their learning goals by (1) referring to an error report summarized by error categories and frequency, and (2) visiting the corresponding online resources (e.g. the link to English Grammar Guide). Unlike the conventional paper-based marking which sees the provision of teacher written feedback as an end in itself along the continuum of assessment, the electronic feedback can be viewed as a cyclical or complementary practice in a teaching-learning-assessment cycle. To achieve this, not only does the electronic feedback guide students through error categorization, explanation and correction in level two, but it also generates an error report for a learner which is summarized by categories and by frequency, and it opens the door for an online tutorial where a learner who finds himself in need of further explanation or practice in a particular error can choose to go over the corresponding pages of the interactive English Grammar Guide by just simply a click at the bottom of the feedback window (i.e. "Click here for more advice and practice").

	The 3-step electronic feedback for each identified error entails the following features:	Details	Level of Explicitness
1.	Error identification and a brief explanation in the pop-up window	A pop-up window which identifies the lexcio-grammatical error (i.e. error type) and suggests ways for error correction or improvement is shown.	Low-Mid
2.	Provision of a recommended lexico-grammatical form or an example sentence in another context through the web-based resources	A link to 'Google Search' / 'Word Neighbor' is posted at the bottom of the pop-up window. Click on the 'Google Search' link (or 'Word Neighbors' link), then a recommended usage of the lexico-grammatical form or an example sentence will be presented in a different context.	Mid
3.	An electronic English Grammar Guide provides a more detailed explanation of the identified error type, followed by some practice	If students need a more detailed explanation and practice on a recommended lexico-grammatical form, simply just click on "Click here for more advice and practice" and they will be diverted to the electronic "English Grammar Guide" (EGG) where it will provide the full and relevant details along with practice.	Low-Mid
	Students can exercise their autonomy to follow or randomly select the step(s) according to their needs and interests.		

Figure 3.1 The Three-Step Feedback Approach

By responding to and acting on writing feedback through the above pedagogical operation, it is believed that 'Mark My Words' ('MMWs') can help engender the feedback situation, approximating to Dickinson's explanation of the situation required for developing autonomy (Dickonson,1987, p. 11), in which not only learners can identify, understand and correct their errors, but also they are involved in developing their own learning fashion and goals according to their language needs and learning preferences. As a result, learner autonomy can be encouraged

and promoted in language development. On the other hand, not only can learner autonomy help play down the teacher's overt control over students' language and writing process in future revisions by pointing out or even correcting their errors, but also it can help eliminate some possible teacher's 'blind spots' in ascertaining the students' level of interlanguage in the continuum in which teachers may encounter difficulties in deciding the appropriate level of explicitness to be given to individual students in their writing feedback. If students are to be given a choice as to the level of explicitness they want, it is believed that their subconscious awareness of their own interlanguage level would inform them of the best judgment. Further discussion on these practical issues alongside some suggestions will be explored in the second part of the literature review.

3.1.2. *Language Theories*

A. The Role of Error Correction

Views on the role of error correction are not conclusive but conflicting (Hedge, 2000). Krashen (1985) argues that adult second language acquisition is analogous to children's acquisition of their first language such that explicit grammatical feedback is not necessary, thus error correction has its questionable value. Krashen explains that children do not generally receive any explicit negative feedback on their accuracy when acquiring their mother tongue; similarly adult learners do not require explicit negative feedback which is discouraging and de-motivating (ibid).

However, criticism against Krashen's view about the role of error correction points out that "adult learners can be encouraged to process error correction in useful ways, and the role of the teacher is to provide feedback which learners can work on in order to refine their understanding and move to the next stage of interlanguage" (Hedge, 1997, p.15)

In fact, the role of error correction can be conceptualized as the continuum with one end on the left representing 'the most explicit' error correction under a broader category of 'Direct Feedback', the middle-left representing 'the more explicit' coded feedback and the middle-right representing 'the less explicit' uncoded feedback under a broader category of 'indirect feedback', and the other end on the right representing no feedback at all.

'Direct feedback' refers to teachers' provision of correct answers in response to student errors (Lee, 2008a). According to Ferris (2006), 'direct feedback' may take various forms including crossing out an unnecessary word, phrase, or morpheme; inserting a missing word or morpheme; or writing the correct word or form near the erroneous form (Ferris, 2006). It is used when teachers perceive the error in question is complex in a way that it is beyond students' ability for self-correction (Ferris, 1999; Frodesen, 1991)

'Indirect feedback' refers to teachers' indication of errors (by means of a circle, an underline, a code or a mark) with correction by students (Lee, 2008a). It can be further streamed into two categories which are coded feedback and uncoded feedback. Coded feedback indicates an error by pointing out the types of errors made such as 'preposition' or 'tense', whereas uncoded feedback indicates an error by simply underlining or circling it without indicating an error type (Lee, 2008a).

It is used when teachers want to engage students in problem-solving and develop their independent editing skills (Ferris, 2002; Ferris & Hedgocok, 2005; Lalande, 1982).

The inconclusiveness of the role of error correction, which is conceptualized as the various levels of explicitness in the continuum, reveals that there is no one best and most adequate way of giving writing feedback. This argument is supported in the literature where Ferris (1996 & 1999) argues that students' level of progress in error correction varied depending on error types. This implies that different feedback strategies should be prescribed for different error types. This issue will be revisited in the following sections.

While the 'explicit feedback' on error correction is usually represented by writing the correct word or form near the erroneous form (Ferris, 2006) and 'indirect coded feedback' is represented by means of a circle, an underline, a code or a mark with correction by students (Lee, 2008a), 'indirect uncoded feedback' only indicates an error by simply underlining or circling it without indicating an error type (Lee, 2008a). When putting Krashen's view on error correction with his 'Input Hypothesis' theory (1995) into perspective, 'indirect uncoded feedback' appears to approximate to Krashen's view on the role of error correction in the larger literature context of second language acquisition in which feedback with any explicit hints is discouraged.

Following Krashen's line of thought, the next question to arise is what kind of input should be given if adopting Krashen's view on error correction? If it is assumed that 'indirect uncoded feedback' approximates to Krashen's view on the role of error correction, then how should the input be represented in error feedback, other than merely underlining or circling errors without identifying their error types for students? Lastly at what point, pedagogically speaking, should

such input be given to students? The above questions, which involve a more practical and mechanical operation of Krashen's view on error correction, can not be answered without understanding Krashen's Input Hypothesis. Referring to the inconclusiveness of the role of error correction, it is worthwhile to reiterate that, in this study, Krashen's view on error correction is treated as one of the options in writing feedback in the belief that the responsiveness of specific errors towards error correction to a certain extent depends on particular feedback treatments (Ferris, 2006). Supporting Ferris's argument, the experimental results of this study also indicate that some error categorizes are relatively more sensitive to a particular feedback mode than the others.

B.   Krashen's Input Hypothesis

Following what we have discussed in the previous section, this section will explore what kind of input should be given if adopting Krashen's view on error correction as one of the options, how the input should be represented in error feedback, and at what point, pedagogically speaking, such input should be given to students. To answer these questions, we have to first clarify the notion of language input postulated by Krashen and understand the working principles behind one of his most prominent theories, 'input hypothesis'.

Input hypothesis postulated by Krashen (1985) is a significant idea which has emerged in second language acquisition. "The hypothesis makes a distinction between acquiring a language and learning it through conscious attention to language study" (Hedge, 2000, p. 10). According to Krashen (1985), there are two separate systems of second language performance, namely, 'the

31

acquired system' and 'the learned system'. The former refers to second language acquisition which postulates that language is picked up subconsciously and this is very similar to a child acquiring his or her first language. The latter refers to second language learning which postulates that language is consciously learned through formal instructions. Despite the distinction between acquisition and learning as they are defined by Krahsen, efforts on the part of the second language learner to understand the L2 through discovery rather than through their efforts to use it are mandatory. Hence, the nature of language input is very central to the difference between these two performances.

The 'Input Hypothesis' postulates that language is acquired in "only one way", that is, "by understanding messages, or by receiving 'comprehensible input'" (Krashen, 1985, p.2). The input hypothesis is a hypothesis in second language acquisition developed by Stephen Krashen, which states that a language learner gains the most benefit from receiving linguistic input that is just beyond his or her current interlanguage, or level of grammatical understanding. This type of input is known as comprehensible input or "i + 1," where "i" refers to the learner's interlanguage. Hedge (2000) further explains and clarifies the notion of comprehensible input as follows:

(1) The comprehensible input has to be meaningful such that it is 'relevant and topical to learners and their interests, or realistic in terms of simulating the authentic texts and speaking situations learners may eventually have to handle'.

(2) The value of providing input through out-of-class resources or encouraging students to make use of resources might be available to increase the output opportunities is emphasized.

(3) Teachers' own classroom language should be adjusted to be in line with students' proficiency

(p.12).

In summary, the notion of language input in second language acquisition comprises the following features:

(1) The input has to be 'comprehensible', 'meaningful', 'relevant' and 'authentic'.

(2) The level of input given has to be a bit above a student's level of language competence and understanding.

(3) Students have to be given opportunities to interact with this input through other available resources.

The 'Input Hypothesis' (Krashen, 1985) postulates that language acquisition takes place when a learner is exposed to a comprehensible input which is at the level 'i+1' where 'i' represents a learner's current level of language competence and understanding. When the learner exposes to this new input, s/he will formulate a hypothesis about a rule which governs the target language item. The hypothesis about the rule will then be tested out in the learner's future encounters, and then it will gradually be revised as a learner discovers contradictory evidence when receiving more exposure to other language input on his/her own, or receive feedbacks from others.

To put the working principles of input hypothesis into perspective, it is through the process of hypothesis making and testing such that a learner can make sense of these new but comprehensible input and then impose a rule governing its usage. These rules putting together form a developing system known as interlanguage, which "passes through a number of stages until it eventually

approximates to the rules of the target language or until it stabilizes, or fossilizes, in ways that deviate from these rules" (Hedge, 2000, p. 11). Though it is clear that some kind of language input is needed for language acquisition taking place, Hedge (2000) argues that "many questions remain about the kind of input which is most useful in facilitating the process" (ibid, p. 12). In the following sections, it will explain how 'Mark My Words ('MMWs') can provide a web-based platform such that it can help facilitate the process of input hypothesis in teacher written feedback.

C.   Implications for the Electronic Feedback ('Mark My Words')

The purpose of this paper is not to compare and evaluate the differences between second language learning and second language acquisition. With the primary interest in improving the effectiveness of teacher written feedback while cultivating students' autonomy in their language development, 'Mark My Words' ('MMWs') is designed in such a way to accommodate individual learners' language needs and learning preferences. This accommodation also extends to whether they are learning or acquiring a particular language item.

The following diagram (see *Figure 3.2*) illustrates how the three-step approach underpinned by the second language theories, namely the "Role of Error Correction" and "Krashen's (1985) Input Hypothesis", is pedagogically operated within Nunan's (1997) 'Model of Framework for Developing Learner Autonomy':

Pedagogical Operation of 'Mark My Words' ('MMWs')		
Nunan's Theoretical Framework for Developing Learner Autonomy		The 3-step Feedback Approach underpinned by the beliefs: (1) Role of Error Correction & (2) Krashen's Input Hypothesis
Level One:	Inform learners of the available feedback types	1. Error identification with a brief explanation of the error type shown in the pop-up window. 2. Provision of a recommended lexico-grammatical form or an example sentence in another context through the web-based resources (e.g. Google link). 3. An electronic English Grammar Guide provides a more detailed explanation of the identified error type, followed by some practice
Level Two:	Learners make their own choice of the feedback type	
Level Three:	Learners act on the writing feedback	
Student can exercise their autonomy to follow or randomly select the step(s) according to their needs and interests		

Figure 3.2 Pedagogical Operation of 'Mark My Words' ('MMWs')

Having said that, the following questions are yet to be answered:

(1) What kind of input should be given if adopting Krashen's view on error correction as one of

the options in error feedback?

(2) How should the input be represented in error feedback?

(3) At what point, pedagogically speaking, such input should be given to students?

These working principles of input hypothesis have been synchronized into the pedagogical

operation of 'Mark My Words' (MMWs). The following examples illustrating the pedagogical

operation of the electronic feedback provide some insights into the above questions.

When any error is identified, a recommended lexico-grammatical form is searched for using the

Google search engine embedded in the electronic feedback system. The example sentences

comprising the feature of recommended form generated are just comprehensible enough for the students to undergo the subtle process of hypothesis making and testing of the target form.

For instance, a student produces the following output "I concern my grammar"; and a teacher-user would like to correct the wrong usage/form of 'concern' in the sentence by using the electronic feedback system. Other than adopting 'indirect coded feedback' which clearly identifies its error type (i.e. adjective-verb confusion) alongside a detailed grammatical explanation, the teacher-user can also at the same time adopt 'indirect uncoded feedback' by activating the 'Google Link' function and inputting the phrase 'concerned about' (the structure the learner seems to be aiming for) in the search box of the Google search engine embedded in the electronic feedback system. When the student receives his or her electronic feedback (in the MS-word format), puts the curser over the highlighted error, and clicks on 'Google Link' icon on the pop-up window, it will show the correct and contextualized usage of 'concerned' (as an adjective) in the Google search results like the following:

	… and other policymakers who *are concerned about* the maltreatment of…
	I'*m very concerned about* the way …
	…feeling that the hospital staff members *were concerned about* the patient.
	etc.

As illustrated above, the Google search results generate a number of the recommended usage of 'concerned' in various contexts. As a departure point here, the student is expected to make a hypothesis about the recommended usage of 'concerned' in these new contexts without being

explicitly told that this belongs to 'adjective-verb' confusion, and 'concern' has to be changed into a past participle form preceded by a verb-to-be and followed by a preposition 'about'. In the end, it is anticipated that the student is able to act on such indirect uncoded feedback by imposing the correct form into his or her own context of writing and putting down "I am concerned about my grammar".

The merit of 'Mark My Words' ('MMWs') lies in its provision of an interactive-based electronic platform which can facilitate this subtle process of hypothesis making and testing such that a learner can make sense of the new input and then impose a rule governing its usage. The nature of language input postulated by Krashen (1985) coincides with 'indirect uncoded feedback' which is comprehensible, meaningful, relevant and authentic such that it can help students to identify the errors, understand the errors, form their hypothesis of the correct grammatical form or lexical choices while avoiding explicit error correction. The language input provided is comprehensible and meaningful as the recommended usage is embedded in new contexts; it is also relevant and authentic because the example sentences are derived from authentic texts in Google news or Google scholars.

Referring to the inconclusiveness of the role of error correction, it is worthwhile to reiterate that, in this study, Krashen's view is treated as one of the options in writing feedback in the belief that the responsiveness of specific errors to a certain extent depends on particular feedback treatments (Ferris, 2006). This argument is supported in the literature where Ferris (1996, 1999) argues that students' level of progress in error correction varied depending on error types. For example, students are more able to correct their errors through feedback on forms provided that the errors

are rule-governed and are less able to correct their errors if the errors are not rule-governed or the errors are more idiosyncratic such as 'word choice' or 'word order'. These differences imply that different feedback strategies should be prescribed for different error types. Hung (2000) seems to share Ferris's view by pointing out that when it comes to grammar teaching some language structures "are either hard to notice or hard to generalize" by students (p. 1). To this end, Ferris (1996, 1999) suggests a more directive feedback strategy such as reformulation or complete correction in these cases. However, the drawbacks of doing so would violate Krashen's propositions as well as defeating the principles underpinning learner autonomy as discussed earlier. To compromise such deviation, the provision of a recommended lexico-grammatical form or example sentences through the web-based resources in 'Mark My Words' ('MMWs') is being regarded as an alternative feedback strategy to help learners to 'acquire or pick up' the recommended language items or structures which are not rule-governed and are more idiosyncratic. This alternative strategy, to a great extent, has responded to Ferris's subsequent recommendation that a judicious combination of direct & indirect feedback, varying according to error types, may be most helpful to students (Ferris, 1999, 2006). The findings over the responsiveness of specific errors on different feedback different treatments will be covered in the 'Findings and Discussions: Error Correction' section in Chapter 6.

3.2   *Major Issues on Teacher Written Feedback*

The second part of the literature review is divided into nine areas within the research body of 'writing feedback', 'error correction', 'second language writing' and 'assessment'. These nine areas foreshadow and summarize major yet specific issues which are worth further exploration, discussion and clarification in this study. In addition, the findings from the previous studies have formed the basis for developing and experimenting with the electronic feedback (i.e. 'Mark My Words') which aims to address or even bridge the knowledge gaps by incorporating some of the untested or partially tested ideas proposed in this study and by other researchers in the previous research.

Error correction has been always a subject of debate in second language writing over decades. The most primitive debate on error correction among ESL writing researchers and teachers was whether teachers should provide any error feedback and whether the teacher written feedback has any positive effects on student writing development (Ferris, 2006). The most notable argument against the value of giving teacher written feedback was represented by Truscott (1996) who argued that "practical problems such as teacher limitations and student inattention render even the hypothetical benefits of grammar correction unlikely (cited in Ferris, 2006, p. 82), concluding that in the end student writers cannot benefit at all. In response to Truscott (1996) argument which denied the value of giving error feedback on student writing, Ferris (1999, 2002, 2003, 2004, 2006) argued that while the effectiveness of error feedback was not conclusive, more research should be done to explore if error feedback should be embraced or abolished.

Apart from arguing if error feedback should be given (Truscott, 1996), the more active and ongoing discussions on error correction concern when, what, and especially, how to correct student errors. These questions have led to a field of research relating to the most efficient way to correct student errors by means of different modes of teacher feedback on student writing (Paulus, 1999, p. 267). Among these various modes of teacher feedback under investigation, quite a number of them have attempted to explore if the level of explicitness of error feedback would have any positive or negative effects on writing revisions from students of various proficiency levels (Ferris, 1996, p.83). One important dichotomy is the distinction between direct and indirect feedback; and under direct feedback, it is further divided into coded feedback and uncoded feedback.

Unfortunately, the answers as to at what level of explicitness would make a difference in student writing revisions not conclusive (Lalande, 1982; Robb, Ross, & Shortreed, 1986, Leki., 1991; Arndt, 1993, Saito, 1994; Truscott, 1996; Ferris, 1995; 1997; 1999; 2003; 2006; Hyland, 1998; 2001; 2003; Radeck & Swales, 1998; Lee, 2008a; 2008b). The case in point is that while Robb, Ross, & Shortreed (1986) and Ferris and Robert, (2001) argue that coded and uncoded feedback make no significantly statistical difference to improving students' revisions, Lalande (1982) and Lee provide conflicting findings. Thus, the question of what type of error feedback is most beneficial to students deserves further scrutiny in this study.

3.2.1   Focus of Teacher Written Feedback

Some earlier studies indicate that teachers primarily attached importance to the 'form' of the language (grammatical accuracy) rather other its 'meanings' when giving written feedback (Cumming, 1985; Zamel, 1985). A number of studies also reported that a high proportion of teacher written feedback focused on error feedback rather than written comments (Hyland, 2003; Lee, 2008a) However, students, irrespective of proficiency level, wanted more written comments from teachers when comparing to the amount of errors they wished to be pointed out. This implies that students would like to receive more qualitative comments about what their teachers thought and felt about their writing holistically than seeing their texts awash in red ink of error identification and correction (Lee, 2008b).

However, the advent of process writing since 1990s has come with more studies reporting a shift in teacher written feedback focus from 'form' to other areas like 'content' and 'organization' (Hyland & Hyland, 2006a).

Regarding the above, to maximize the effectiveness of teacher written feedback, a more balanced feedback practice consisting of written commentary with error feedback which emphasizes equally on forms and meanings is advocated.

3.2.2   Teacher's centeredness of Teacher Written Feedback

A number of studies reported that teachers often take control over students' language and writing process in future revisions by explicitly pointing out and correcting their errors (Bronnon, 1982).

Explicit correction means a correct lexical form or a correct reformulation of language structure is explicitly imposed on student writing. As such, teacher written feedback cannot adequately empower students with the skills necessary for the control over future revisions (Cogie, 1999). Besides, doing so will discourage the development of learner autonomy as far as Nunan's model of framework for promoting autonomy in language learning is concerned.

Supporting Cogie's view (1999), Milton (2006) also argues that teacher writing feedback may be inadequate and not necessarily appropriate in helping students to be an autonomous and effective writer in the long run.

The study conducted by Lee (2008b) confirms similar findings. It is reported in her study that teachers' feedback which was mostly teacher-centered made students passive and dependent on teachers. This is evident in her observation that "students wanted teachers not only to indicate errors but also to provide corrections and indicate error types" (ibid, p.156). Students of both high and low proficiency showed a tendency to demand greater effort on the part of the teacher in giving more written comments and more explicit error feedback.

As a result, such explicit error correction renders a vicious cycle of escalating students' expectation for teachers identifying and correcting every single error, resulting to a passive and independent learning behavior in the learning-to-write process. This echoes Belcher & Liu (2004) who suggests that as students relinquish power to their teachers in the sense that they want to be told what to do rather than take initiatives to direct their own learning, which seems to describe this recurring pattern.

Nevertheless, there are concerns that students may run the risk of not being able to comprehend and of misinterpreting the error feedback when they are provided implicitly. Therefore, the crux of the issue lies on how to strike a balance between theses two extremes of the continuum; that is, the level of explicitness of the teacher feedback.

Students should be viewed as active and proactive agents in the feedback process. They are more likely to find teacher feedback useful and effective when feedback (1) can engage and interact with the student writers more cognitively and personally (e.g. 'reading aloud some good sentences in class' and 'asking the teacher for clarifications, explanations or help in class') and (2) when it is contextualized – that is, given in consideration of individual student needs (Hyland & Hyland 2006a & 2006b; Lee, 2008a, 2008b).

3.2.3   Effectiveness of Teacher Written Feedback

A.   Error Feedback & Misinterpretation of Findings in other Studies

Controversy continues as to whether feedback, or particularly which form of feedback, can help improve accuracy and overall quality of student writing (Robb et al., 1986; Truscott, 1996, Ferris, 2001, Lalande, 1982; Lee, 1997, Semke, 1984; Kepner, 1991; Fathman & Whalley, 1990).

In fact, a number of studies have been conducted to examine how students responded to and acted upon various types of teacher written feedback underpinned by different feedback strategies. For

example, direct feedback vs. indirect feedback, coded indirect feedback vs. uncoded indirect feedback, teacher feedback vs. peer feedback. However, no conclusive evidence could be drawn to support which form of feedback is relatively more effective than the others in improving accuracy in writing revisions. The controversy over the effectiveness and comprehensibility of error feedback is appended in *Figure 3.3* below; and the findings of some key studies regarding the effectiveness of error feedback with varying degree of explicitness (e.g. direct feedback vs. indirect feedback, or indirect coded feedback vs. indirect coded feedback) is appended in *Figure 3.4.*

Arguments against teacher feedback on writing		
Truscott (1996), Cohen & Robbin (1976), Zamel (1985)	Students ignore teacher feedback. Students cannot utilize teacher feedback effectively in their writing revision.	
Zamel (1985) & Lee (2008b)	Teacher feedback is not always understood due to its illegibility.	
Truscott (1996)	Teachers' Limitations	Students' Limitations
	Practical problems like incomplete, inconsistent and inaccurate teachers' error feedback have rendered writing feedback ineffective. Truscott (1996) also attributed the ineffectiveness of teacher feedback to the teachers' lack of skills to analyze and explain students' problems, as well as the students' lack of skills to understand and use the feedback.	Students' level of progress in error correction varies depending on error types and students' interlanguage levels, but unfortunately teachers often responded to errors of all categories in the same way without sufficient awareness that different types of linguistic form may take different order and sequence of L2 acquisition, while ignoring variations in students' interlanguage levels to comprehend, process and utilize teacher feedback successfully.
Ferris (1996 & 1999)	Ferris (1996 & 1999) argued that students' level of progress in error correction varied depending on error types (i.e. treatable error vs. untreatable error)	
Ferris (2003 & 2006)	Ferris (2003) suggested that a consistent system of coded feedback with systematic grammar instruction be in place. And Ferris (2006) also recommended a judicious combination of direct and indirect feedback, varying according to error types, may be most helpful to students.	
Hyland & Hyland (2006); Lee (2008a); Ferris (2014)	Hyland & Hyland (2006) stated that "The use of language corpora and concordancing offers one of the most exciting applications of new technologies to the writing class…(p.95)". They also concluded that "a final key for research must be the potential of automated essay evaluation and computer-mediated feedback for improving student writing and developing their independent writing skills. We have little information on students' views of these programs or the effects of computer-generated response, so researching examining students' perceptions, and use, of electronic feedback systems in naturalistic settings is needed. Since	

	there are likely to be many developments in such software in coming years, this will be a prime area of research related to feedback as will further research on areas of computer feedback such as online social interaction and revision practices" (p.96). Lee (2008a) also recommended that future research can explore the use of computer-based written feedback and how students respond to it in its specific context. Ferris (2014) recommended in her implications for future practice that "Teachers should explore how computer-based can benefit themselves and their students. Electronic feedback is superior to handwritten feedback in three different ways: it is legible, it is clearer and less cryptic; and it is permanent and can be saved for future reference or analysis" (p.21).
Lee (2008a & 2008b) & Hyland & Hyland (2006a &2006b)	Most teacher feedback is teacher-centered. Students will more likely find teacher feedback useful if teacher can engage and interact with student writers more cognitively and personally (e.g. asking teachers for clarification in class and reading aloud some good sentences) and when teacher feedback is contextualized in response to student needs.
Truscott & Hsu (2008)	47 EFL postgraduate students were divided into two groups in which the experimental group receiving indirect uncoded feedback and the control group receiving no feedback at all. The study which examined the total error reduction (but not by each error category) found that the indirect uncoded feedback (e.g. underline) was more effective in error correction in the revision of the first draft about the narrative description of a sequence of 8 pictures with prompts. However, the post-test results from another new narrative description of a sequence of 8 pictures with prompts showed that there was no significant difference between the two groups in terms of the total error accuracy. Truscott & Hsu (2008) concluded that successful error correction in revisions was not equivalent to the improved accuracy in a new writing text; and therefore the language gain from the revisions failed to transfer to language learning.
Bitchener & Knoch (2010b)	63 advanced ESL learners were divided into 4 groups examining the effectiveness of English 'article' correction with (1) direct focused feedback with written metalinguistic input; (2) direct focused feedback with written & oral metalinguistic input; (3) indirect circling feedback; and (4) no feedback. A pre-test which was about writing a description based on a picture was given and the texts were marked. It was found that both direct feedback with written and/or oral metalinguistic input performed better than indirect circling feedback, and no feedback being the worst in error correction, in a ten-week delayed post-test about writing a description based on another picture. Bitchener & Knoch (2010b) concluded that "provision of clear, simple metalinguistic explanation with examples" on a separate sheet of attached paper is the best type of written corrective feedback for long-term accuracy (p.216). Under Bitchener's new definition, his operation of 'direct focused feedback was equivalent to Ferris' 'indirect coded feedback'.

Figure 3.3 The Controversy over the Effectiveness & Comprehensibility of Error Feedback

Summary of the effectiveness of teacher feedback in varying degrees of explicitness		
	No difference in improving language accuracy in writing revision	Coded feedback makes a difference in improving language accuracy in writing revision
Coded (indirect) Feedback vs. Uncoded (indirect) Feedback (no non-feedback control group)	Robb et al. (1986) - compared one direct feedback treatment with three indirect feedback treatments in varying degrees of explicitness - all treatments can reduce errors over time (long run) - little longitudinal difference in long-term achievement in accuracy related to the level of explicitness of error feedback	Lalande (1982) - 60 intermediate-level German language learners at US university - compared indirect coded feedback with direct feedback - indirect coded feedback can reduce errors (over time), but direct feedback cannot. - indirect feedback triggers the "guided learning & problem solving" process (p.140)

	No difference in improving language accuracy in writing revision	Error Feedback makes a difference in improving language accuracy in writing revision
	Ferris & Roberts, B. (2001) -72 ESL university students -compared indirect coded feedback with indirect uncoded feedback -all treatments can reduce errors from draft 1 to draft 2 (short run) - no sig. diff. in error reduction when comparing draft 2 to draft 1 (short run) - # of errors vs. feedback treatment	Lee (1997) -indirect coded feedback was more effective in remedying errors
		Ferris (2006) - 92 ESL university students -compared direct feedback with indirect coded/uncoded feedback from draft 1 to draft 2 of essay 1,and between essay 1 and essay 4 - All treatments can reduce errors - no sig. diff. in error reduction between both treatment groups when comparing draft 2 to draft 1 (short run) – sig. diff. in error reduction when comparing essay 4 to essay 1 (long run) - specific errors vs. feedback treatment (verb & sentence)
	No difference in improving language accuracy in writing revision	Error Feedback makes a difference in improving language accuracy in writing revision
Content comments vs. error correction vs. combination of content comments & error correction vs. error identification without correction (with a non-feedback control group)	Semke (1984) -college-level German students	Fathman & Whalley (1990) -72 US ESL students -grammar feedback reduced error but content feedback not
	Kepner (1991) -62 Spanish ESL college students -form-focused feedback equally effective as meaning-focused feedback in error reduction (in the long run) - no sig. difference	
	Sheppard (1992) -26 upper-intermediate ESL learners -same results as Kepner's study (1991)	

Figure 3.4 Summary of the effectiveness of teacher feedback with varying degrees of explicitness

The study conducted by Lalande (1982) reported that students who received indirect coded feedback, which is by means of an underline, circle, code, or other mark but does not provide the correct form (Ferris, 2006, p. 83), had their errors reduced over time, whereas students who received direct feedback, which is by means of writing the correct word or form near the erroneous form (ibid), did not. Lalande (1982) attributed the effectiveness of indirect coded feedback to the fact that coded feedback in itself triggered the "guided learning & problem solving" process which helped students to acquire the target linguistic features over time.

Robb et al. (1986) studied their 134 Japanese college freshmen's ability in revising their drafts marked by the instructor with four modes of feedback. The first group was the correction group in which all errors were corrected by the instructor. The second group was the coded feedback group in which all errors were marked with a code in which the type of error was indicated in the student writing. The third group was the uncoded feedback group in which all errors were located but no codes. The fourth group was the marginal feedback group in which the total number of errors per line was written in the margins. Five compositions were analyzed and graded using 1 subjective and 18 objective measures of writing skills. The results of the study did not support the practice of direct correction of surface errors (e.g. explicit and direct feedback) and suggested that the less time-consuming methods may suffice since no groups outperformed the others. Hence, the result of their study (Robb et al., 1986) suggested that there was little longitudinal difference in student achievement in accuracy related to the level of explicitness of error feedback, and that "highly detailed feedback on sentence-level mechanics may not be worth the instructor's time and effort" (p. 91).

In another study, Lee (1997) examined how her 149 first-year Electrical Engineering students responded to different error correction practice. The error treatment task contained both surface and meaning errors, and the students were asked to make appropriate changes to errors under three conditions; namely, direct prompting, indirect prompting and no prompting. The results indicated that the students performed the best in error correction under the direct prompting condition where coded feedback, instead of overt correction, was given. Lee (1997) points out that abbreviated correction code is believed to be a very effective tool to help students correct their errors in Hong Kong schools (p. 467).

However, Ferris & Roberts (2001) conducted a research to compare the effectiveness of 'marked but uncoded feedback' versus 'coded feedback' (i.e. coded feedback vs. uncoded feedback) on student writing revisions. In her study, a group of 72 long-term immigrants at a US university composed their first draft. These immigrants then received either 'marked but uncoded feedback' or 'coded feedback' and revised their draft based on the feedback. The findings were that students receiving 'marked but uncoded feedback' had 75% of the errors corrected whereas students receiving 'coded feedback' had 77% of the errors corrected. Derived from the results, Ferris and Robert (2001) reported that though both forms of feedback treatment did help improve their grammar accuracy, the groups respectively receiving 'uncoded feedback' and 'coded feedback' made no statistical difference in error reduction. A noticeable issue in Ferris & Roberts's study (2001) which might have weakened the validity of their argument is that their generalization was premised on the responsiveness of students' error reduction to the two distinct forms of feedback treatment, without considering the responsiveness of different error types to these two distinct forms of feedback treatment (i.e. students' frequency of errors reduction by error categories). By

this, it means certain types of errors, for example, 'Word-level' errors, might be more insensitive to different feedback strategies representing different levels of explicitness. Due to the possibility that some 'Word-level' errors are just too obvious to be corrected, judging the effectiveness of feedback treatment from the responsiveness of students' error reduction alone might not provide an accurate representation of the extent to which it is the form of the feedback (i.e. the level of explicitness) which makes any statistical difference, or to what extent it is the error types (i.e. treatable or untreatable) which do. Ferris & Roberts (2001) investigated the importance of student attitudes and preferences about error feedback and their own assessment of their weaknesses in writing (Ferris, 2001, p. 166). The coded feedback received 48% popularity, the marked but not labeled feedback received 31% popularity, and the no feedback only received 19% popularity. In short, 'students clearly favoured the more explicit approach (i.e. 'coded feedback)' (Ferris, 2001, p. 178).

However, it remains a question as to whether it is valid to put on par the responsiveness of the 'Word-level' errors (i.e. 'verb tense/form' errors) and the sentence-level errors ('sentence structure' errors) towards the level of explicitness in error feedback; and then use the results to evaluate if the level of explicitness would have any effects on the students' improvement in accuracy. This is because 'Sentence structure' errors, representing at the later stage of L2 acquisition after 'verb tense/form', is supposed to be more difficult for students to acquire; whereas 'verb tense/form' errors tends to be easier to be identified and corrected. The distinction between the word-level linguistic forms and sentence-level linguistic forms in their sequence of L2 acquisition might have theoretically made 'verb tense/form' errors less responsiveness to the level of explicitness, while making 'sentence structure' errors theoretically more responsive to the level

49

of explicitness in the first place. Hence, this might have possibly weakened the validity of Ferris's case that argued for the superiority of indirect feedback over direct feedback for error correction.

In a subsequent study conducted by Ferris (2006) to compare students' ability of error correction in response to direct feedback and indirect coded / uncoded feedback, it is found that there was no significant 'short-term' difference between the responsiveness of certain error types (i.e. 'verb tense/form' and 'sentence structure') towards direct feedback and indirect feedback in improving students' accuracy in the next draft; however, there was a significant 'long-term' difference between responsiveness of the same error types towards direct feedback and indirect feedback in improving students' accuracy over time (i.e. the fourth draft). Ferris (2006) concluded that the findings in her study "make a strong case for the superiority of indirect feedback over direct feedback for facilitating student improvement over time, at least with this particular population of students" (p. 98). However, a question remains as to whether it is valid for Ferris (2006) to put on par the responsiveness of 'Word-level' errors (i.e. 'verb tense/form' errors) and the 'sentence-level' errors ('sentence structure' errors) towards the level of explicitness in error feedback; and then use the results to evaluate if the level of explicitness would have any effects on the students' improvement in accuracy. This issue of validity will be discussed more in depth in the 'Chapter 6: Findings and Discussions: Error Reduction'.

The findings reported by Ferris (2006) in her study lend support to Lalande's (1982) findings in that students who received indirect coded feedback had their errors reduced over time. Their findings also supported Lee's (1997) study where she pointed out that abbreviated correction code (i.e. coded feedback) was relatively more effective to help students correct their errors in Hong

50

Kong schools when comparing to the uncoded feedback (p. 467). By contrast, their findings (Lalande, 1982; Lee, 1997; Ferris, 2006) contradicted those reported by Robb et al. (1986) that there was little longitudinal difference in student achievement in accuracy related to the level of explicitness of error feedback.

The more recent development about teacher written feedback with the specific focus on examining the actual effectiveness of direct and indirect feedback on error correction can be represented by the studies conducted by Truscott & Hsu (2008) and Bitchener & Knooch (2010b) respectively as shown in Figure 1 on p.39. To further argue for his criticisms that teacher feedback was ineffective, unhelpful and time-wasting as it was held in his early work in 1996, Truscott with his partner conducted another study in 2008. In this study (Truscott & Hsu, 2008), forty-seven EFL postgraduate students were divided into two groups in which the experimental group receiving indirect uncoded feedback (i.e. underlining errors) and the control group receiving no feedback at all. The study found that the indirect uncoded feedback was more effective in error correction in the revision of the first draft about a narrative description of a sequence of eight pictures with prompts. However, the post-test results, which was about another new narrative description of 8 pictures with prompts, showed that there was no significant difference between the two groups receiving either indirect uncoded feedback or no feedback at all in terms of error accuracy by error totals but not by each error category. Truscott & Hsu (2008) concluded that successful error correction in revisions was not equivalent to the improved accuracy in a new writing text; and therefore the language gain from the revisions failed to transfer to language learning. However, their findings were based on the statistical difference of error rates between narrative 1 and narrative 2 (i.e. post test) by error totals but not by error categories as it was in this study. As such,

51

the responsiveness of each error category towards various forms of corrective feedback in the long run, evident by the frequency of error made by each category in the post-test, was unknown. Despite no statistically significant difference noted by error totals, in my opinion future research can look closely into to what extent different error categories can benefit from various feedback treatments in the long run, as in Ferris' study (2006) despite the methodological flaws identified (see Chapter 5.1.1). Having said that, Truscott & Hsu's findings might have possibly challenged the positive results claimed by Ferris in her study (2006) that indirect feedback was more effective than direct feedback in remedying errors in her longitude study where she argued "the direct intervention (on an untreatable 'sentence-level' error) did not appear to have any lasting effect over time" whereas "the indirect feedback that students received on verb errors (treatable error) may help them more over time because it consistently called this error to their attention…" (p.96). In two years' time after Truscott & Hsu's study (2008), the findings from Bitchener & Knoch (2010b) might have helped Ferris in restoring some of her credits from Truscott's criticisms. In their study examining specifically two functional uses of the English article system (i.e. "a" and "the"), 63 advanced ESL learners were divided into 4 groups examining the effectiveness of 'article' error correction with (1) direct focused feedback with written metalinguistic input; (2) direct focused feedback with written & oral metalinguistic input; (3) indirect circling feedback; and (4) no feedback. A pre-test which was about writing a description based on a picture was given and the texts were marked. It was found that both direct feedback with written and/or oral metalinguistic input performed better than indirect circling feedback, and no feedback being the worst in error correction, in a ten-week delayed post-test about writing a description based on another picture. Bitchener & Knoch (2010b) concluded that "provision of clear, simple metalinguistic explanation with examples" on a separate sheet of attached paper is the best type of

written corrective feedback for long-term accuracy (p.216). Under Bitchener's new definition, his operation of 'direct focused feedback' (2010b, p.212) was equivalent to Ferris' 'indirect coded feedback' (2006). Deviating from the conventional notion of indirect coded feedback characterized by coding erroneous forms, the major additions Bitchener & Knoch (2010b) made in her study was the integration of the written and/or oral metalinguistic input into direct feedback. Such additional input in the form of an attached paper for written explanation and/or of a 15-min full class discussion on errors (p.212) might have possibly enhanced the effectiveness of direct feedback in error correction, reversing the inferiority of teacher feedback in student writing revisions in Truscott's studies (1996, 2008). However, the question I raised here is the successful error correction in revision being seemingly transferred to the improved accuracy of a new text is only limited to English article system, but not the overall grammatical accuracy in their study. Another question is the practicality of the provision of grammatical explanation on a separate piece of paper if the types of error teachers attended add up.

B.   Teacher Written Commentary

Ferris (1997) conducted a research study to evaluate the impact of different types of teacher commentary on students' revised draft. Her commentary included 'questions asking for further information', 'requests', 'comments giving information to students', 'positive comments' and 'comments about grammar mechanics'. She examined over 1600 marginal and end comments written on 110 first drafts of papers by 47 advanced University ESL writers. She then looked into the revised drafts to check for the influence of different types of teacher commentary on students' revisions and assess whether the changes made in response to the teachers' feedback actually

improved their writing according to her subjective rating scale. The scale considered the degree to which the student utilized each of her commentary into the revisions by making 'no attempt', 'a minimal attempt', or 'a substantive attempt', and whether the repairs improved the revisions had mixed effects or had a negative effect on the revisions (Ferris, 1997, p.320). The results suggests that teacher commentary on 'grammar' and 'request for information' appeared to lead to the most substantive revisions (ibid, p.330).

Paulus (1999) examined the effects of peer feedback and the effects of teacher feedback on 11 ESL undergraduate students' essays. The results suggest that teacher written feedback exerted a greater impact and being prioritized by students (Paulus, 1999, p.283).

Hyland & Hyland (2001) examined students' responsiveness towards teacher commentary on praise, criticism and suggestions. The results suggest that feedback was important in providing helpful advice on student writing and negotiating an interpersonal relationship that helps facilitate student's development (p. 208).

C. Comprehensibility of Teacher Written Feedback

A number of studies reported that teacher feedback is not always understood due to its illegibility (Zamel, 1985; Lee, 2008b). Lee (2008b) reported from her study that not all the students were able to act on teacher feedback or found the feedback useful, especially those lower proficient students in the study. Truscott (cited in Ferris, 2006) argued that practical problems like incomplete,

inconsistent, and inaccurate teachers' error feedback have rendered the writing feedback ineffective.

Truscott (cited in Hyland, 2003, p.218) attributed the ineffectiveness of teacher feedback to the teachers' lack of skills to analyse and explain students' problems, as well as the students' lack of skills to understand and use the feedback.

Accounting for plausible factors other than lack of skills on the part of teachers in giving adequate error feedback, Lee (2008b) indicated that teachers who focused on comprehensive feedback were more likely to produce illegible feedback since comprehensive feedback placed a more demanding workload on teachers in terms of the amount of feedback given on each piece of writing.

In light of the issues discussed in this section and the previous section, the following suggestions were made by Ferris and Lee to enhance the effectiveness and comprehensibility of teacher written feedback. Ferris (2003) suggested that a consistent system of coded feedback which is supported by systematic grammar instructions should be put in place. Ferris (2006) also suggested that a judicious combination of direct & indirect feedback, varying according to error types, may be most helpful to students. Ferris's suggestion was based on her observation on teachers' intuitive selection of direct or indirect feedback by having regard to (1) 'treatable' or 'untreatable errors' types and (2) their students' abilities in self-correction. Having said that, Ferris (2006) and other previous researchers (Lalenda, 1982 & Lee, 1997) preferred indirect coded feedback to direct feedback when it came to translating short-term success in error correction to long-term gains (Ferris, 2006, p. 95-96). The findings in Ferris's (2006) indicated that students in the long run (i.e.

draft 1 to draft 4) made statistically significant reductions in an error type (i.e. 'verb tense / form' in her study) receiving indirect coded feedback; while the same group in the long run failed to make statistically significant reductions for another error type (i.e. 'sentence structure' in her study) receiving the direct feedback. However, a question remains as to whether it is valid to put on par the responsiveness of the 'Word-level' errors (i.e. 'verb tense/form' errors) and the 'sentence-level' errors ('sentence structure' errors) towards the level of explicitness in error feedback; and then use the results to evaluate if the level of explicitness would have any effects on the students' improvement in accuracy. This is because 'Sentence structure' errors, representing at the later stage of L2 acquisition after 'verb tense/form', is supposed to be more difficult for students to acquire; whereas 'verb tense/form' errors tend to be easier to be identified and corrected. The distinction between the word-level linguistic forms and sentence-level linguistic forms in their sequence of L2 acquisition might have theoretically made 'verb tense/form' errors less responsiveness to the level of explicitness, while making 'sentence structure' errors theoretically more responsive to the level of explicitness in the first place. Hence, this might have possibly weakened the validity of Ferris's case that argued for the superiority of indirect feedback over direct feedback for error correction.

Lee (2008b) recommended that future research can explore the use of computer-based written feedback and how students respond to it in its specific contexts (Lee, I., 2008b). With these suggestions in mind, an attempt has been made on my part to incorporate such consistent system of coded feedback covering content, organization and language into an electronic-based platform called 'Mark My Words' (MMWs).

D. Treatment of Different Error Types

Ferris (1996, 1999) argued that students' level of progress in error correction varied depending on error types. This implies that different feedback strategies should be prescribed for different error types. Her argument is based on the following premises:

1. Students were more able to correct their errors through feedback on form provided that the errors were rule-governed and that the rules governing concerned errors were taught such as 'subject-verb agreement' or 'article-problem'.
2. Students were less able to correct their errors through feedback on form if the errors were more idiosyncratic such as 'word choice' or 'word order'. In this case, a more directive feedback strategy such as reformulation or complete correction might be needed.

However, Truscott (1996) argued that teachers often responded to errors of all categories in the same way without the awareness that different types of linguistic forms (morphological, syntactic and lexical) represent distinct domains of linguistic knowledge and may take different order and sequence of acquisition. Having said that, Ferris (2006) pointed out that none of the studies reviewed focused specifically the responsiveness of specific errors (correction) to particular feedback treatments. In addition to the factor of error types which would affect students' level of progress in error correction, students' level of proficiency would also be another determining factor in comprehending teacher feedback. For example, Students at lower levels of L2 proficiency may not have sufficient linguistic knowledge to self-correct errors when pointed out implicitly (Ferris & Hedgcock, 2005)

As far as the above is concerned, it is important that teachers respond to errors according to student abilities (Straub, 2000). A flexible feedback policy which takes into account student abilities is more likely to help students develop interest, confidence and self-esteem in writing rather than a rigid policy which requires detailed and explicit error feedback across the board. To this end, future research could examine how teachers can vary their feedback according to students in order to maximize the benefits of feedback (Lee, 2008b).

E. Alternative Feedback Treatments

Critical examination of the teaching practice for EFL writing, such as controlled composition, the process approach and English for academic purpose, has cast doubt on the "adequacy of these strategies for the effective instruction or learning of writing" (Silva, 1990). In view of this, Milton (1997) argues that "the lack of demonstrable success by conventional instruction has been a justification for the exploration of independent learning strategies to improve writing skills among these learners" (p.238).

Many teachers find marking to be a 'tedious and unrewarding' task as they found that their feedback 'only rarely seems to bring about improvements in [students'] subsequent work' (Hyland, 1990). To enhance the effectiveness of teacher feedback in improving the revision process of student writing, some alternative feedback treatments were explored. Among those are 'Conferencing', 'Checklist', and 'Reformulation.

'Conferencing' takes place when teachers provide individual support during the progression of composition. It is a particularly useful technique at the early stage of the writing process when students are still thinking about the content and organization (Hedge, 2000, p. 313). Through careful questioning and elicitation like 'Who you are writing to?' or 'How have you organized your point?', the teacher can discuss any aspects of the writing with individual students ranging from organizing ideas to finding appropriate language (ibid). It is reported from a study conducted by Keh (1990) that students responded positively towards conferencing. It is noteworthy that the essence of 'Conferencing' lies with its real-time interaction between teachers and students through which an immediate response could be given. However, this feedback condition is subject to several constraints. Firstly, the real-time interaction can not be done without teacher's presence; secondly, individual consultation is highly likely hamstrung by a large class size; thirdly, it will possibly help students in the formation and organization of ideas but it may not render sufficient support in grammatical accuracy; and fourthly, it may work well during the planning stage of writing but may not work equally well at the later stage like editing and proofreading. Despite the study (Keh, 1990) reporting a positive student feedback on conferencing, these positive responses were affective in nature. That is, students would tend to appreciate the efforts made by teachers in giving individual consultation over future revisions. This positive affective response on individual consultation was reported in a study evaluating student reaction to teacher feedback (Lee, 2008b), and was also resonated in the findings of this study.

Another feedback treatment is using a 'Checklist' in which a list of guided questions covering the purpose of the writing, development and organization of the main ideas alongside their supporting details, and the language features of the particular genres are provided (Hedge, 2000, p. 313). The

'Checklist' enables students to self-evaluate their own writing against the expected standard without a teacher's presence. The 'Checklist' is popular device at a slightly later stage of the writing process (ibid). Though this feedback condition may help students scaffold the aspects of content and its logical flow expected in a particular discourse, the checklist which serves no more than a sheet of writing guidelines may not be able to cater for students' individual needs on one hand and may stifle students' creativity on the other.

The last treatment is 'Reformulation'. It is a useful procedure after students have produced their first drafts and are moving on to look at more local possibilities for improvement in the second or final drafts (Hedge, 2000, p.313). 'Reformulation' offers students with opportunities to notice and discuss any differences between the model writing and the students' own production, hoping they can acquire how ideas are paragraphed, how they are connected, and how the target language features are used in the model text and then incorporated them in their own writing (ibid). According to Allwright (1988), 'Reformulation' proceeds the following stages:

(1) All students carry out a guided writing task to ensure that the content and organization of their writing is similar overall.

(2) Each student writes a first draft and submits it to the teacher.

(3) The teacher marks the drafts by indicating the problems by means of underlining or highlighting.

(4) The teacher chooses one students' essay and reformulates it, following the ideas closely but improving the language use in terms of accuracy and appropriateness.

(5) The original essay and the reformulated essay are copied so that students can notice the

differences.

(6) Students work in pairs or groups, identifying the changes in the reformulation and discussing the reasons for the changes.

(7) The teacher discusses the changes with the class, providing a rationale and inviting comments and questions.

(8) Students then revisit their own first drafts and revise their writing based on the models suggested in the reformulated essay.

Despite its advantages mentioned, some possible drawbacks of this approach are identified. Firstly, it might stifle students' creativity in writing when students are given a guided writing task. Secondly, students should be encouraged to identify, comprehend and correct their own errors; however, encouraging them to imitate a reformulated lexical or clausal model situated in a sample essay, which might have detached from the contextual meaning in their original writing, might fail to address their individual writing needs and problems.

Regarding the above, effort was geared in this experimental study to explore an alternative feedback mode which on one hand can preserve some of the strengths and on the other compensate the shortcomings identified above.

F. Assessment for Learning or Assessment of Learning

With much of the discussion so far situated the study of teacher written feedback within the 'second language writing' and 'error correction' literature, this section attempts to situate teacher written

feedback within the 'assessment' literature, with an aim to find out the extent to which each of the electronic feedback and paper-based feedback serves as a means for assessment for learning (AfL) or assessment of learning (AoL). This is done by examining the nature and functions of each of the electronic feedback and paper-based feedback against certain criteria prescribed in the recent literature. The reason why this matters lies in the suggestion that assessment for learning (AfL) forms the basis for a good teacher feedback practice in the writing classroom (see Lee, 2007). To situate 'teacher written feedback' within the assessment literature, an attempt is made to clarify some key concepts of formative assessment vs. summative assessment, and assessment for learning (AfL) vs. assessment of learning (AoL).

"The AfL – AoL distinction is often seen as a parallel to the longstanding division between formative and summative assessment" (Lee, I., 2007, p.181). This means formative assessment shows a strong relevance to assessment for learning (AfL); whereas summative assessment shows a strong relevance to assessment of learning (AoL). To grasp their relationships is to understand the conceptual differences between formative and summative assessment.

The major distinction between summative and formative assessment lies in how the information gathered from the assessment is used but not the time of the assessment. That is either the information obtained is used for the improvement of teaching and learning or merely used for the measurement and judgment of student performance (See Genesee & Upshur, 1996; William, 2001; Yorke, 2003; Lee, 2007). In fact, formative assessment serves to "contribute to student learning through provision of information about performance" (Yorke, 2003, p.478), and it is usually conducted continuously (Black, Harrisom, Lee. Marshall, & William, 2003; Hout, 2002; William,

2001). For example, information obtained from the assessment is to improve teaching practice and inform students about their strengths and limitations (Lee, I., 2007). Thus, the assessment is exploited in a more prospective manner. On the other hand, summative assessment serves to "elicit evidence regarding the amount or level of knowledge, expertise or ability" (William, 2001, p. 169) for measurement of student performance and judgment of learning. The assessment is usually conducted at the end of the course (Black, Harrisom, Lee. Marshall, & William, 2003; Hout, 2002; William, 2001). For example, information obtained from the assessment is to measure what students have learnt against target objectives (Lee, 2007), assign grades or report level of performance to concerned parties (Genesee & Upshur, 1996). Thus, the assessment is used in a less prospective manner. According to Lee (2007), "Essentially what distinguishes formative from summative assessment is whether the assessment is used to work towards the improvement of teaching and learning" (p. 182).

The major distinction between assessment for learning (AfL) and assessment of learning (AoL) follows the identical traits of formative assessment and summative assessment. In assessment for learning, information gathered from the assessment is used for identifying the strengths and weaknesses of students' performance with an ultimate goal of enhancing their learning on the one hand, and for helping teachers reflect on their teaching objectives and strategies with an ultimate goal of improving their instructions on the other (Curriculum Development Institute [CDI], 2004). In assessment of learning, information gathered from the assessment is used for assessing or reporting student performance by measuring students' performance against a specified set of learning targets or objectives (Lee, 2007). Adding to this distinction, students under assessment

for learning play an active role in the assessment process rather than mere passive recipients under assessment of learning (Gardner, 2006).

Formative and summative assessments were treated as two sides of a coin in the past (see Scriven, 1967). In fact, recent studies suggest that they are not "two sides of the same thing" (Roos & Hamilton, 2005, p. 9) and they have their "overlapping functions" (Lee, 2007, p. 181). On a similar line, assessment for learning and assessment of learning are not "mutually exclusive" and an assessment can serve the dual purposes (Lee, 2007, p.182).

Implications for Teacher Written Feedback

Assessment for learning (AfL) is regarded as a good feedback practice in the writing classroom (see Lee, 2007). Applying William's (2001) definition of formative assessment which is to use feedback to promote assessment for learning, i.e., the following criteria has to be satisfied:

1. Students are told about their strengths and what needs to be done in their writing – e.g., areas for improvement, in terms of content, organization, language, etc.; the assessment is prospective;

2. Students act on the teacher feedback and are provided with opportunities to improve their learning based on the teacher feedback;

3. Information is communicated clearly and made intelligible to students in terms of what they have learnt, hence a close link between teaching, learning and assessment;

4. Students play an active role in managing their own learning.

Measuring against the above essential criteria, defined by William (2001), for the formative feedback being used as promoting assessment for learning, the electronic feedback practice is able to fulfill these requirements of a good feedback practice for the following reasons.

When comparing to the conventional paper-based marking which usually finishes off a composition by circling or underlining an error with a code or symbol, the electronic feedback offers more resources in helping students what needs to be done in their future revisions. These recourses includes categorizing error types, giving thorough explanation, suggesting ways for error correction implicitly (via Google Link or Word Neighbor), and directing web-based resources (e.g. English Grammar Guide) for further language support in the problematic area. All these have fulfilled criterion '1' and criterion '2'

Unlike the conventional paper-based marking which sees the provision of teacher written feedback as an end in itself along the continuum of assessment, the electronic feedback can be viewed as a recurring or complementary practice in a teaching-learning-assessment cycle. To achieve this, not only does the electronic feedback guide students through error categorization, explanation and correction, but also it opens the door for an online tutorial where a student who finds himself in need of further explanation or practice in a particular error can choose to go over the corresponding pages of the interactive English Grammar Guide by just simply a click at the bottom of the feedback window (i.e. "Click here for more advice and practice"). In addition, the electronic feedback can generate a statistical report diagnosing the category and frequency of errors. The statistical report about students' error pattern can be used as reference informing teachers the

effectiveness of the current teaching practice and illuminating them what further needs to be done in future teaching. Hence, this recurring and complementary feedback practice which is conducive to enhancing teaching and learning has fulfilled criterion '3'

Different from the conventional paper-based feedback which has already dictated the nature and amount of feedback to be received by students, the electronic feedback equipped with its interactive features allows students to play an active role in determining what information and how much information they would like to receive. This has fulfilled criterion '4'.

After all, the electronic feedback in this study distinguishes itself from the paper-based feedback by its enhanced capability in diagnosing students' errors as well as offering online pedagogical resources (e.g. Google Link, Word Neighbor, English Grammar Guide) for both short-term and long-term language development, whereby students are given the autonomy to manage their own language learning and act on the electronic feedback so as to improve their future revisions.

As discussed in the previous section, 'Mark My Words' ('MMWs') partially incorporates a range of autonomy postulated by Nunan in his model of framework (1997) into the pedagogical operation of the electronic feedback. It seems to suggest that Nunan's model (1997) has also fulfilled these three criteria prescribed for assessment for learning, whereby students are being encouraged to move progressively from raising their awareness of their errors alongside the options of feedback pathways available to making a decision on their favourite learning pathway and finally managing their future learning plans.

Though the conventional paper-based feedback in this study might have also provided students with comments or suggestions for future revisions, students are mere passive recipients of teacher feedback who have restricted autonomy in managing their own language learning on the one hand and the paper-based feedback is in itself hamstrung by the weak link between teaching, learning and assessment on the other. Hence, the above reasoning seems to suggest that the electronic feedback serves as a means for assessment for learning purposes, whereas the paper-based feedback serves as a means for assessment of learning purposes.

G.   Students' Perceptions of Teacher Written Feedback

Previous research studies on student views of feedback have consistently shown that students treasure teacher feedback and attach much greater importance to it than other forms of feedback, such as audio feedback, peer evaluation, and self-evaluation (Leki, 1991; Saito, 1994; Yang, Badger &Yu 2006; Zhang, 1995).

Most surveys of student preferences show that students are particularly positive about receiving feedback on language issues, although they also want teachers to comment on content and ideas of their writing (Hedgcock & Lefkowitz, 1994; Leki, 1991; Oladejo, 1993; Saito H., 1994).

As second language learners place a high premium on accuracy in writing, they are eager to have all their errors pointed out by the teacher (Komura, K., 1999; Lee, I., 2005; Leki, 1991; Rennie, 2000).

While studies by Radecki and Swales (1988) and Lee (2005) show that students wanted overt correction of errors (i.e., direct error feedback) from teachers, most of the other studies (e.g., Arndt, 1993; Hyland, 2001; Saito, 1994) suggested that students preferred indirect to direct error feedback, where they were given clues and also a more active role to play in the feedback process. This is resonated by Ferris, (2001) who investigated the importance of student attitudes and preferences about error feedback and their own assessment of their weaknesses in writing (Ferris, 2001, p. 166). The coded feedback received 48% popularity, the marked but not labeled feedback received 31% popularity, and the no feedback only received 19% popularity. In short, 'students clearly favoured the more explicit approach (i.e. coded feedback) (Ferris, 2001, p. 178).

Studies on students' perception toward teacher written feedback has begun to develop in the 1990s (e.g. Cohen, 1987; Eginarlar, 1993; Ferris, 1995; Hedgcock & Lefkowitz, 1994, 1996; Leki, 1990; Diab, 2005). However, most of them were investigated without a strong reference to the contextual factors in the sense that how these contextual factors might have influenced teacher feedback practice, which in turn affected the way students responded and acted on actual teacher feedback in a specific context. These contextual factors included instructional environments, students' level of proficiency, and etc, which will be explained in the following section.

These critiques were first evident in the study conducted by Hyland & Hyland (2006b) who reported that most of the previous studies on students' perception toward teacher written feedback were based on one-off questionnaire survey, and they rarely have their findings of students' reactions linked to actual teacher feedback in specific contexts. Similar critiques were also evident in the subsequent study conducted by Lee (2008b) who claimed that the existing research studies on student views of teacher feedback were limited by the weak link between student reactions and

actual teacher feedback situated in specific contexts. Lee, I. (2008b) pointed out that "These findings are presented in a decontextualized and broadbrush fashion. We know little about the nature of the teacher feedback these students have received and the context surrounding the feedback that has caused such student perceptions" (p.145). She further added that "When teacher feedback is investigated without reference to specific learner characteristics and classroom contexts, it is dangerous to generalize results from one group of learners to another group with markedly different characteristics, especially those operating in completely different contexts" (p.146).

Against the above background, there has been a call for examining how the context surrounding feedback practices may affect the way teachers respond to student writing and the way students respond to and act on actual teacher feedback in a particular context (Hyland & Hyland, 2006b, Lee, 2008a & 2008b). It is through analyzing students' perception toward teacher written feedback with reference to the contextual factors we can be more able to comprehend, for instance, if students' favouritism toward error feedback focusing on 'language error' in the findings should be merely attributed to the way the teacher writing feedback was given itself, or should be attributed to a more underlying reason that an exam-driven context has made the students place the premium on language accuracy only.

Avoiding a decontextualized and broad-brush generalization of students' attitudes on teacher feedback, this study has taken into consideration on how students respond to and act on teacher written feedback mediated by particular teacher feedback practice, which are possibly influenced

by teachers' beliefs as well as other contextual factors like instructional environment and students' stage of language development.

Students' perception toward teacher written feedback could have been influenced by students' stage of language development. For example, it is found from a study that students with a higher level of proficiency tended to demand more error feedback on language and would like all their errors corrected, whereas students with a lower level of proficiency would less prefer to have their errors corrected and were less demanding on error feedback on language due to the reason that weaker students might find it discouraging and de-motivating to 'receive papers awash in red ink' (Lee, 2008b).

Students' perception toward teacher written feedback could have been influenced in any one of the following instructional environments. For example, an exam-driven instructional context which place premium on accuracy would possibly drive the students to put more emphasis on receiving error feedback on language (Lee, 2008b). On the other hand, students under the instructional context of free-writing would benefit more on feedback on content in terms of improving fluency and developing confidence, but not enhancing accuracy (Lee, 1998).

Regarding the above, Lee (2008b) introduced a structured and systematic approach to evaluate the students' perception toward teacher written feedback with reference to particular contextual situations. Her design and format of the questionnaire has been adapted in this experimental study (In Chapter 7: 'Findings & Discussions: Students' Perceptions) with aim to investigate students'

perception of teacher written feedback in improving the overall accuracy and appropriateness of student writing in the following areas:

(1) Comprehensibility of feedback

(2) Student ability to correct errors

(3) Types of feedback (i.e. grades / error feedback / written comments)

(4) Focus of error feedback (i.e. content / organization / language)

(5) Amount of error feedback

(6) Focus of written comments (i.e. content / organization / language)

(7) Types of error feedback strategy (i.e. explicitness of error feedback: direct vs. indirect feedback)

(8) Students' responsiveness of error types (i.e. easier / more difficult to understand & correct)

(9) Types of post-writing activity

(10)   Usefulness of feedback

The presentation of the findings will be made reference to the contextual factors identified in the following section.

H. Contextualization of Feedback Studies

A number of studies have suggested that second language students believe that teacher writing feedback was useful and could help them improve their writing (e.g. Ferris, 1995; Hyland, 1998), and that they preferred having all their errors being highlighted (Komura, 1999; Lee, 2005; Leki,

1991; Rennie, 2000), and were particularly positive about receiving feedback on language issues. However, the study conducted by Lee (2008b) indicated that the students of lower proficiency appeared to show a less positive attitude toward error feedback than the students of higher proficiency.

In addition, findings over student preference between direct error feedback and indirect error feedback are not conclusive. The study conducted by Robb, Ross, and Shortreed (1986) reported that there was no statistically significant differences in long term gains of accuracy among four groups of students, in which one group received direct feedback and the others received indirect feedback in varying degrees of explicitness. Studies by Radecki and Swales (1998) and Lee (2005, 2008b) showed that students, irrespective of proficiency level, preferred direct error feedback. However, the study conducted by Lalande (1982) reported that students who received indirect feedback had their errors reduced over time, whereas students who received direct feedback did not. And there are other studies which suggest that students preferred indirect error feedback in which students were given a more autonomous role in self-editing errors with clues provided in the feedback process (Arndt, 1993; Hyland, 2001; Saito, 1994).

In fact, these findings are presented in a decontextualized and broad-brush fashion in the sense that the existing research studies on student views of teacher written feedback are limited by the weak link between student reactions and actual teacher feedback situated in specific contexts (Lee, 2008b). This means the link between student reactions and actual teacher feedback situated under specific contexts were explained in a loosely–tied manner. Lee (2008b) points out that "When teacher feedback is investigated without reference to specific learner characteristics and classroom

contexts, it is dangerous to generalize results from one group of learners to another group with markedly different characteristics, especially those operating in completely different contexts" (p.146).

In light of the above, there has been a call for examining how the context surrounding feedback practices may affect the ways teachers respond to student writing and the ways students react to and act on the actual teacher feedback in a particular context (Hyland & Hyland, 2006b, Lee, 2008a & 2008b).

The study conducted by Goldstein (2004 & 2005) and Hyland & Hyland (2006b) reveals that teacher feedback practices are influenced by a myriad of contextual factors:

1. Teachers' beliefs, values, understandings & knowledge
2. Cultural/Institutional context
   - philosophies about feedback
   - attitudes to exams
   - socio-political issues: power relations & teacher autonomy

Building on the understanding of the feedback context constructed by Goldstein (2004 & 2005) and Hyland & Hyland (2006b), Lee (2008a, 2008b) further developed and characterized the feedback context into the cultural, institutional and inter-personal dimensions, which were further divided into the following factors:

1.   Overall context of teachers' work (institutional & cultural dimensions)

2.   Accountability (institutional dimension)

3.   Examination culture (institutional & cultural dimensions)

4.   Teacher training (institutional context)

5.   Instructional context (institutional dimension)

6.   Teachers' beliefs and feedback practice (interpersonal dimension)

7.   Pedagogical activities (interpersonal dimension)

8.   Teachers' personalities (interpersonal dimension)

9.   Students' proficiency (interpersonal dimension)

10.   Students' motivation (interpersonal dimension)

Lee (2008a, 2008b) argued that studies of teacher written feedback were mostly decontextualized in the sense that they failed to look into the ways, on one hand, how the feedback context might have influenced teachers' feedback practices in the areas of the (1) types of feedback, (2) focus of error, (3) error feedback strategies, and (4) written commentary (Lee, I., 2008a) and how the feedback practices might have influenced student perceptions of the feedback practices, in the areas of the (1) types of feedback, (2) focus of errors, (3) error feedback strategies, (4) legibility of the feedback, and (5) ability for error corrections on the other (Lee, 2008b). Hence, any generalization of students' reactions to teacher written feedback without having regard to any plausible impact of such contextual factors in which any feedback practice is situated has to be taken with great caution because teachers' feedback practices as well as students' responses may vary depending on these cultural, institutional and inter-personal dimensions,

(i)   Instructional context

Instructional context can be exhibited by the writing approach adopted such as product writing or process writing, assessment criteria applied such as the percentage of weighting on written assignments in the overall grade and the percentage of weighting on the language part, marking policy imposed such as detailed marking or selective marking, form-focused marking or meaning-focused marking, and the pedagogy used in classroom setting.

Taking the writing approach as an example, a number of studies showed that, when comparing to single-draft writing, students tended to attend more details to teacher comments in multiple-draft writing in which they believed that teacher comments could help improve their writing (Diab, 2005; Ferris, 2003; Hedgcock & Lefkowitz, 1994). On the other hand, students in single-draft writing were likely to look up to teachers as the sole authority in feedback, making themselves become passive agents in the writing process (Lee, 2008b).

Taking marking policy as a further example, a flexible feedback policy which takes into account student abilities is more likely to help students develop interest, confidence and self-esteem in writing rather than a rigid policy which requires detailed and explicit error feedback across the board (Lee, 2008, p. 158)

(ii)   Teacher's Belief & Feedback Practice

Teacher written feedback underlies teachers' beliefs in feedback practice. The studies conducted by Lee (2008a) reported that most secondary teachers in Hong Kong attached paramount importance to written accuracy, placing priority on grammar and vocabulary. Most preferred responding to every single error and felt that students' written errors needed to be pointed out and corrected, otherwise students could not learn from and avoid making their mistakes. On the other hand, there are other studies recommending indirect error feedback in the belief that students would benefit if they were given a more autonomous role in self-editing errors with clues provided in the feedback process (Arndt, 1993; Hyland, 2001; Saito, 1994). This echoes the findings in the study conducted by Hyland & Hyland (2006b) reporting that students were more likely to find teacher feedback useful when it engaged student writers and when it was contextualized in consideration of individual student needs.

The more profound effect of teachers' feedback practices is that they could have a vicious bearing on student reactions and expectations. Lee (2008b) indicated from her study that the teacher-dominated feedback practice, where comprehensive and explicit feedback was given, has made the students become passive and dependent learners. "As teachers marked student writing in detail, responding to errors comprehensively and not providing opportunities for student-centered activities like peer/self-evaluation, students were rendered passive and became more and more reliant on the teacher. Therefore, students wanted teachers not only to indicate errors but also to provide corrections and indicate error types" (ibid, p.156). The explicit error correction renders a vicious cycle of escalating students' expectation for teachers identifying and correcting every

single error, resulting to a passive and independent learning behavior in the learning-to-write process. This echoes Belcher and Liu (2004) who suggested that as students relinquish power to their teachers, they want to be told what to do rather than take initiative to direct their own learning, which seems to describe this recurring pattern.

Addressing the problems of illegible feedback as a result of the overloaded marking on the part of the teachers, Lee (2008b) suggested (1) adopting selective marking which aimed at quality rather than quantity, and (2) exploring different modes of feedback such as the use of feedback forms with clearly pre-written criteria, audio feedback and computer-based feedback. Future research can explore these alternatives to teacher written feedback and how students respond to them in their specific contexts.

(iii)   Pedagogical activities used with error feedback

Error feedback is given through pedagogical activities. Error feedback alone should not be singled out as a sole factor accounting for students' responses and perceptions towards a particular feedback practice.

For instance, an error-focused pedagogy in classroom teaching, assessment and post-writing oral feedback, delivered in a mechanical, discouraging, and oppressive manner, might have caused frustration among students (Lee, 2008b), whereas students of a higher level of proficiency enjoyed reading aloud some of the beautiful sentences in class in the post-writing oral feedback (ibid).

(iv)   Teachers' personalities (e.g. appreciation vs. reprimand)

If the instructional approach delivered by the teacher was represented by negative and cynical comments, and criticisms, this could have further damped the students' motivation from learning about their written errors (Lee, 2008b). Such frustration and disappointment might possibly compelled the students to withdraw themselves from making improvement because they perceived that their teachers would have never expected them doing so. On the other hand, a number of studies have shown that students generally welcomed praise (Gee, 1972) and liked to receive both praises and constructive criticisms (Ferris, 1995; Hyland, 1998; Lee, 2008b)

(v)   Proficiency level of students

Learners' background, which is characterized by the learners' ages, motivations, and proficiency levels, is also a determining factor in the way they perceive and respond to teacher written feedback. Studies from Rizai (1997) and Leki (2006) reported that students at a more advanced level like the L2 graduate students particularly valued teacher feedback as a useful means to help them develop disciplinary literacy. Also, a study conducted by Lee (2008b) indicated that the students of lower proficiency were less interested in error feedback than those of higher proficiency, though both groups preferred more explicit error feedback from teachers.

(vi)   Motivation of students

Students' motivation is a factor determining students' responsiveness and perceptions towards teacher written feedback. This is supported that the research which found that motivation to be a main determinant in second language achievement (Dornyei, 1994). According to Guenette (2007), students with low motivation were less likely to take teacher feedback seriously and find it useful. Guenette (2007) argues that ''any type of feedback that does not take the crucial variable of motivation into consideration is perhaps doomed to fail'' (p. 52) .Or, this could work in the other around. Students of lower English proficiency would find it discouraging and de-motivating to 'receive papers awash in red ink', and thus would less prefer to have their errors pointed out (Lee, 2008b). Working from a social psychological perspective, MacDonald, R.B. (1991) suggested that students usually had one overall reaction rather than a separate set of reactions to teacher written feedback. When students, especially those weaker ones, received poor grades on their papers, their overall reaction was usually one of frustration and disappointment; and they were likely to reduce such tension by discrediting teacher written feedback.

To a certain extent, the above factors are mediated by the cultural, institutional and interpersonal dimensions in which feedback practice takes places, such as philosophies of feedback and attitudes towards examination, and the socio-political issues pertaining to power and teacher autonomy (Lee, 2008a, p.69). These factors are relating to on another intricately while working in a complementary fashion, which in turn contributes to the notions of teachers' practices as well as students' responses and perceptions towards the feedback practice, and so the overall effectiveness of written feedback. Hardly should a single factor be singled out for the main source of influence.

To remedy the critique that the existing research on teacher written feedback has been characterized by its decontextualized and broadbrush nature (Lee, I., 2008b), reference will be made to the contextual environment surrounding the study when it comes to analyzing the experimental study.

3.3   *Linguistic Differences between English and Chinese*

It appears that quite a number of grammatical errors made by Chinese ESL learners are possibly due to their first language Interference (L1 interference). L1-L2 Interference can be defined as "the native language effect" (Brown, 1994, p. 26). According to Brown (1994), ESL learners rely on their first language (L1) to facilitate the learning process of the second language (L2). It is at the beginning stages of L2 learning where the only reference learners have is L1, so they assume that L2 is similar to L1. Ellis (1997) argued that L1-L2 interference is a process under which ESL writers constructed their own L2 interim rules, more or less approximated to the rules of L2 structure, with the help of their L1 knowledge. Dulay et al (1982) defines L1 interference as the automatic transfer, due to habit, of the surface structure of the first language onto the surface of the target language. Brown (1994) also discusses 'interlanguage' when he refers to the developmental process learners undergo to become competent in L2. It is under this 'interlanguage' development process in which ESL learners' over reliance on the similarity between L1 and L2 as reference that they often made grammatical errors which are characterized by their L1 syntactic features (i.e. the Chinese syntactic features).

It is important to understand the linguistic differences between English and Chinese in order to understand the effects of L1-L2 interference on Chinese ESL learners. A number of studies have shown that the linguistic differences between English and Chinese have an effect on the development of Chinese ESL learners' interlanguage (Cai, 1998, 2002; Li, 2010). The effects of L1-L2 interference could possibly be explained by two language theories, namely, the 'Theory of Language Typology' and the 'Theory of Markedness'.

3.3.1 Theory of Language Typology

The 'Theory of Language Typology' suggests a dichotomy between the 'subject-prominent' structure of English language (SPL) and the 'topic-prominent' structure of Chinese language (TPL) (ibid.), which possibly gives rise to L1-L2 Interference. Based on the theory of Language typology, Li and Thompson's (1976) classification of the four types of world languages has shed lights on the nature of such dichotomy alongside its impacts on Chinese ESL learners in their writing. According to Li and Thompson (1976), English and Chinese represent the two major types of world languages among the four, with English being a 'subject-prominent' language (SPL) and Chinese being 'topic-prominent' language (TPL). According to Li (2010), "The key disparity lies in whether subject is the most fundamental syntactic element. Subject is obligatory in English with the exception of imperatives, and it plays syntactic and semantic roles, or just syntactic role as in the case of dummy subjects. In contrast, subject may be optional in Chinese, and null subjects are frequently used, particularly in colloquial Chinese" (p. 80-81). For example, the "subject-predicate" syntactic pattern might cause confusion to Chinese ESL learners as a 'subject', which plays a syntactical role and a semantic role in English, is not obligatory in the Chinese language.

In summary, the structure of the subject-prominent language (SPL) follows the subject-verb-object (i.e. S-V-O) pattern under which the 'subject' in SPL is both grammatically and semantically essential except in the case of imperatives and dummy subjects. The 'subject' itself is often being the focus/topic of the clause (Example 2), and there is only one main verb in an independent clause (Example 3). On the contrary, the structure of the topic-prominent language (TPL) does not necessarily follow the subject-verb-object pattern; that is, the absence of a 'subject' is very common for a sentence in TPL (Example 1), and it is acceptable in TPL to have more than one main verb in a clause (Example 3). If there is a subject in TPL, the subject may not necessarily to be the focus/topic of the content in the clause and/or the doer of the action (Example 4). Some of the distinctive features between the topic-prominent language (TPL) represented by the Chinese language and the subject-prominent language represented (SPL) by the English language can be exhibited in the examples below accompanied by explanations.

Example (1)	Dummy subject is the unique feature in SPL, but not in TPL.
Chinese	Is raining.
English	It's raining.

Explanation for Example (1): Zero subject is known that null subject is the typical characteristic of TPL, and it is acceptable to omit a subject in Chinese, but in highly SPL like English, a subject cannot be dropped. When Chinese-speaking learners translated Chinese sentences with verbal phrases occupying sentence-initial position, they may produce subject-less sentences in the target language (Li, 2010)

Example (2)	SPL follows the 'subject-verb-object' pattern (except imperatives) but it is not obligatory in TPL. A verbal phrase can function as a subject or topic in Chinese.
Chinese	Help the poor is meaningful.
English	Helping the poor is meaningful. / To help the poor is meaningful.

Explanation for Example 2: A verbal phrase can function as subject or topic in Chinese, but in English it has to be an infinitive or gerund if it functions as subject. This means that learners must reset the phrase by adding the infinitive particle to or the inflectional suffix -ing to indicate their syntactic role as a subject. Learners' errors may result from two situations. One is that null subject is the overt transfer of Chinese topic-prominent feature. The other is that Chinese is not an inflectional language as English, and this may lead to the lack of the gerund form of a verb that is derived by adding the inflectional suffix –ing.

Example (3)	SPL follows the 'subject-verb-object' pattern (except imperatives) but it is not obligatory in TPL. There is no distinction between the finite and non-finite verbs in Chinese.
Chinese	There were three persons went into the estate.
English	Three persons went into the estate.

Explanation for Example (3): In Chinese, there is no distinction between finite and non-finite verbs, with the former having different tenses and aspects by means of inflectional devices. In other words, it is common in Chinese to have more than one main verbs (i.e. there is only one finite verb) in a clause. Unlike Chinese, it is grammatically obligatory in English to change the form of the second verb in a clause into either a gerund or an infinitive (i.e. a non-finite verb). For example, "There were (finite verb / main verb) three persons going (non-finite verb) into the estate."

Example (4)	The main verb often refer to the subject (i.e. the doer) in SPL, but this may not always be the case in TPL.
Chinese	She die one horse then this much cry no stop.
English	She had a horse die on her, and she cried thus much without stopping.

Explanation for Example (4): Applying the S-V-O pattern of English in interpreting the Chinese sentence "She die one horse then this much cry no stop." would be problematic and confusing. If we try to say that "she" is the "subject" and "one horse" is the object, then it is natural to assume that "die" is a transitive verb, or at least has a transitive use. However, this would go against what Chinese speakers feel about this sentence. It is actually not that "she" caused the death of the horse, but that her horse died and this has affected her in some way.

3.3.2   Theory of Markedness

Apart from the 'Theory of Language Typology', the 'Theory of 'Markedness' can also be used to explain the impact of L1-L2 interference. According to Li (2010), "unmarked features are those that are universal or present in most languages, whereas marked ones are those that are specific to a particular language or found in only a few languages. For instance, dummy subjects are marked items in English (Example 1), e.g. 'it' is raining.) whereas nouns and pronouns are unmarked items (Example 2) when comparing sentential subject in Chinese and English" (p. 81).

Example (1)	Dummy subject is the marked feature in SPL (English), but not in TPL (Chinese).
Chinese	Raining. (A subject is not obligatory)
English	It's raining. (The subject 'it' has no semantic role but it just serves to satisfy the S-V-O pattern)

Example (2)	Nouns and pronouns are the common features in both SPL (English) and TPL (Chinese).
Chinese	He (Noun) go (no inflection for tenses in Chinese) back (no preposition in Chinese) his (Pronoun) room.
English	He (Noun) went back to his (Pronoun) room.

Another example of the marked feature for English is a clausal subject elicited by 'that' (Example) 3). In English, "'that' in a dependent clause plays a syntactical role but not a semantic role; and it is highly marked in English but Chinese clausal subject needs not such complementizer" (Li, 2010, p.84).

Example (3)	The clausal subjects 'that', 'who', 'which' as complementizers in a dependent clause is a 'highly' marked (unique) feature in SPL (English) but not in TLP (Chinese).
Chinese	The girl sit beside me is my sister.
English	The girl who/that is sitting beside me is my sister.

3.3.3 Implications of the 'Theory of Language Typology' and 'Theory of Markedness' on Common English Errors made by Hong Kong ESL learners

It is worth noting that the 'Theory of Language Typology' and 'Theory of Markedness' just serve to explain the some fundamental differences between English and Chinese, and how the so called 'L1-L2' interference gives rise to some common English grammatical errors made by Hong Kong ESL learners. Despite so, this is not to say that all differences found between these two language types could be attributed to or explained by these theories, for example 'English collocations'.

The 'Theory of Language Typology' postulates the dichotomy between the 'subject-prominent' structure of English language (SPL) and the 'topic-prominent' structure of Chinese language (TPL). Such dichotomy, which is mostly at the clausal level characterized by the 'S-V-O' pattern which is not obligatory in Chinese, is highly associated with the English grammatical errors such as 'sentence fragment', 'misuse of a verbal phrase as a subject', and 'misuse of finite and non-finite verbs'.

The 'Theory of Markedness' argues that some language features (i.e. marked features like 'complementizer', 'tenses', 'articles', 'singular/plural form in English) are specific and unique to a particular language and some (i.e. unmarked features like noun and pronoun) are common to more than one language. Some English errors often made by Hong Kong ESL students associated with the marked features, ranging from word-level to clausal-level, which include 'tense problems', 'wrong/missing articles', 'wrong/missing preposition', 'missing a suffix 's' for a plural form', 'missing a clausal subject as a complementizer like 'that', 'which' and 'who''. According

to Li (2010), 'clausal subjects as a complementizer' is a highly marked feature of English and they might cause confusion to Chinese ESL learner because there is no distinction between an independent clause and dependent clause in the Chinese language.

3.4 Pedagogical Implications: The Role of L1 in ESL Writing for Hong Kong ESL nnn Students

The question followed after understanding the linguistic differences between English and Chinese is to discuss to what extent as well as in what ways such L1-L2 Interference has an impact on Chinese ESL learners in their writing. A number of studies have been conducted to explore this area.

Ying (1987) examined 120 essays from Taiwan Chinese ESL college students and categorized errors on the basis of three criteria: overgeneralization, simplification, and language transfer. The results of this study indicated that, among 1,250 errors identified in these essays, as much as 78.9% of the errors were a result of language transfer, 13.6% were overgeneralization of the target language, and 7.5% were forms of simplification.

Cai (1998) and Cai (2002) investigated the influence of topic-prominence on Chinese ESL learners' writing by virtue of the theory of L1 transfer. In Cai's study (1998) where he investigated Chinese ESL learners' acquisition of subject-predicate structures, it is found that at least 23% of the errors resulting from the 'topic-prominent' features were attributed to L1 transfer and there

was no significant difference between the influences of topic-prominent features on beginning and intermediate learners, but there was between the two groups and advanced learners.

Chan (2004) examined English writing samples from 710 Hong Kong ESL college students with the focus on five error types: (1) lack of control of the capula, (2) incorrect placement of adverbs, (3) inability to use the 'there be' structure for expressing the existential or presentative function, (4) failure to use the relative clause, and (5) confusion in verb transitivity. Her findings revealed that many Chinese ESL learners tended to rely on their L1 linguistic repertoire as reference when producing the English equivalent in such a way that "the surface structure of many of the interlanguage strings produced by the participants were identical or very similar to the usual or normative sentence structures of the learners' first language (L1), Cantonese" (Chan, 2004, p.56), where the extent of L1 syntactic and vocabulary transfer was getting more serious, especially among learners of a lower proficiency level, when the participants were writing complex L2 syntactic structure comprising more complicated meanings. Chan's findings (2004) resonates the study conducted by Liu, Sung, and Chien (1998) that the less English proficiency learners possess, the more L1 interference was found in their English writings.

A study conducted by Chen (1998) reported that most Taiwan Chinese students experienced difficulties in using correct tenses in English due to the reason that Chinese is not an inflected language (i.e. a marked feature for English language). Another study conducted by Chen (2000) indicated that articles (i.e. a marked feature in English language) in English could be one of the most difficult grammatical items for Taiwan Chinese ESL students as Chinese has no equivalent

syntactical device to the English article system. Both tenses and articles are marked features for English language but unmarked features for Chinese language.

Apart from investigating the impact of L1-L2 interference on word-level grammatical errors like 'tenses' and 'articles', Hsin (2003) examines the run-on sentences identified in the writing samples of Taiwan Chinese ESL students. It was observed that the dichotomy between the 'subject-prominent' structure of English language (SPL) and the 'topic-prominent' structure of Chinese language (TPL) was the main cause for this error type. Hsin (2003) concludes that such a linguistic difference between English and Chinese has resulted in the recurring run-on sentence errors for Chinese ESL students.

The above studies have contributed to our understanding on how L1-L2 interference gives rise to negative transfer in ESL writing for Chinese students. Although the findings of these previous studies suggested that L1-L2 Interference might have posed difficulty for Chinese ESL learners in writing 'standard' English, the results of this study seem to suggest that L1-L2 interference (or negative transfer) might possibly be overcome at the 'revising' stage as long as an effective feedback strategy was provided.

The issues identified alongside the proposed pedagogical approach 'Mark My Words' ('MMWs') within the context of this study will be presented in Chapter 4 'Research Questions and Background of the Study'. A thorough discussion in this aspect will be presented in Chapter 6 'Findings and Discussions: Error Reduction'.

3.5   *The Use of Technology in Teacher Feedback on Writing*

The use of technology in teaching and learning has been getting common, which is well-received by both teachers and students (Hyland, 2010). However, the application of technology in provision of feedback on writing is still not as common, and the studies of which are quite limited in terms of quantity and scope. The use of technology in this particular study generally refers to 'Computer-assisted language learning' (CALL). Under the banner of CALL with respect to teacher writing feedback, Ene & Upton (2014) suggested a dichotomy between 'computer-facilitated feedback' and 'computer-generated feedback'. 'Computer-facilitated feedback' refers to "electronic feedback produced by either the teacher or student peers with the help of a computer and delivered electronically to the student" (p. 82), and this usually takes the form of "comments, track changes, or feedback provided in synchronous or asynchronous computer mediated communication (CMC)" (ibid, p. 82). On the other hand, 'Computer-generated feedback' provides "automated algorithm-derived feedback drawn from an existing database of comments or corrections" (ibid, p. 82).

The software program that can deliver 'Computer-facilitated feedback' is the Microsoft Word program in which one of its functions is called 'review' that allows users (e.g. teachers or markers in this study) to annotate a word document (Ene & Upton, 2014). This form of feedback is generally known as 'electronic feedback'.

The software programs that can deliver 'Computer-generated feedback' is generally known as 'Automated Writing Evaluation' ('AWE') (Ene & Upton, 2014). A few 'AWE' programs have been developed, for example, 'Criterion', 'My Access', 'Summary Street, and etc., which can automatically detect and identify certain errors or specific range of errors mostly related to content, grammar, mechanics or organization (Stevenson & Phakiti, 2013), based on the existing database of comments or corrections (Ene & Upton, 2014).

As such, the major dichotomy between 'Computer-facilitated feedback' (or generally named as 'electronic feedback') and 'Computer-generated feedback' (or generally named as 'Automated Writing Evaluation' ('AWE')) lies with the former being a kind of human-operated feedback, whereas the latter being a kind of fully automated feedback, capitalizing on the technology of natural language processing, latent semantic analysis and artificial intelligence (Stevenson & Phakiti, 2013). The electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis', which are the two second language acquisition (SLA) theories, being adopted and tested out in this study is called 'Mark My Words' ('MMWs'). 'MMWs' is being placed under the category of electronic feedback for it is not automatic but solely operated by teacher-users via an add-on tool bar on Microsoft Word that enables teacher-users to manually insert pre-set comments alongside boilerplate 'resource-rich' feedback into students' assignments via the web-based platform. (see Chapter 4.4.2. 'Pedagogical Operation of 'Mark My Words' ('MMWs')).

Despite the differences between 'Computer-facilitated feedback' and 'Computer-generated feedback' in terms of their technological advancement, both are regarded as the identical form of feedback treatment in supporting teacher feedback on writing, and they were being placed under a critical review in this study.

The reason for the inclusion of both 'computer-facilitated feedback' and 'computer-generated feedback' into the literature review and discussion of this study are, firstly, if we take a broader scope and a looser definition in examining the use of technology in teacher writing feedback, both 'computer-generated feedback' and 'computer-facilitated feedback' can be viewed under the identical form of feedback, with contrast to paper-and-pen feedback in particular, where both of them involve the integration of technology in supporting teacher feedback on writing. Secondly, the methodological designs of both 'computer-facilitated feedback' and 'computer-generated feedback' research studies alongside their research outcomes provide this study much insights as to how (i) to improve the quality of 'computer-facilitated' feedback under limited resources (unlike those 'computer-generated feedback' programs which are mostly institutionally funded or commercially funded for the profit-making purpose), as well as (ii) to enhance the validity and reliability of the research design in this experimental study (which will be explained in great detail in Chapter 5 'Research Methodology'). And lastly, it is hoped that a critical review of both strands can help contribute some sort of evidence to the growing body of literature of the effectiveness of the pedagogical use of computer-mediated feedback (which covers both computer-facilitated feedback and computer-generated feedback, in a broader sense), informing us the knowledge gaps and clarifying some mixed results of the previous research.

As far as the above is concerned, the focus of this critical review of computer-mediated feedback in the following sub-sections does not aim to make any comparisons or draw any conclusions about the relative effects of electronic feedback and various 'AWE' programs, nor any of those with 'Mark My Words' ('MMWs'), the electronic feedback program being adopted in this study.

3.5.1 Computer-facilitated Feedback (or Electronic Feedback)

According to Ene & Upton (2014), "The current body of research on electronic feedback from teachers is limited" (p. 80). There are few studies which examine the electronic feedback in second language learning, and these studies were mostly situated in the context of evaluating corrective feedback in online discussions (e.g. Sauro, 2009; Samburskiy & Quah, 2014; Martin-Beltran & Chen, 2013), rather than evaluating corrective feedback in multiple-draft writings (i.e. Ene & Upton, 2014), which is quite similar to this research study.

A study conducted by Sauro (2009) investigated the impact of two types of computer-mediated corrective feedback on the development of 23 Swedish adult learners on the use of zero articles with abstract nouns (i.e. target form knowledge). The investigation took place through a real-time (synchronous) online text chat. 9 Swedish participants at a time were scheduled to chat with their respective American chats partners in individual chat rooms using the Virtual Classroom chat tool of Blackboard. Participants were provided with writing prompts and certain words and abstract nouns which they had to use during the 20 minutes online text-chat. The American chat partners, which were divided into two condition groups and one control group, provided corrective feedback

whenever they noticed any errors related to the use of the target form. One condition group provided 'recast', the other provided 'metalinguistic prompt' and the last one provided only 'topic relevant response that does not contain the target form in the same context. At the end, the computer-delivered acceptability judgement tests containing a mixed of repeated abstract nouns and new abstract nouns in a new context were administered to the participants as a post-test to check which type of corrective feedback (i.e. recast or provision of metalinguistic information) was more effective for immediate gains and gains over time in remedying the errors on the use of zero article with abstract nouns. The study reported that both 'recast' and 'metalinguistic information' groups supported gains in the target form knowledge but neither type was significantly more effective than the other in either the immediate term or over time.

Similar to Sauro's study (2009), a study conducted by Samburskiy & Quah (2014) examined the types of corrective feedback (i.e. 'recast', 'metalinguistic feedback', 'textual enhancement', 'elicitation', and 'explicit correction') and their frequency of use in marking students' errors by 6 novice language teachers. These teachers were paired with groups of 6-12 Belorusian students. The online correspondence between teachers and learners lasted for two weeks during which students were required to produce a minimum of five posts per week. The topics included 'Introducing Myself', 'Culture and Language' and 'Online Learning'. The experiment took place through the Moodle course management platform, where the written data were collected in the form of transcripts of asynchronous interaction between novice online language instructors and learners. The discussions were extracted from the transcripts so that the sequence and original attributes of the posts were captured which allowed the researchers to trace the types of corrective feedback to learners' posts. The study was not intended to evaluate student uptake of corrective

feedback provided by their teachers against their posts online. It is because the research design did not require the learners to reproduce the post in the correct form in response to the feedback. The study concluded that these novice language teachers used 'recast' and 'textual enhancement' most frequently in their online corrective feedback.

Using a similar tellecollaborative setting with English language learners from Taiwan and student teachers from the U.S., Martin-Beltran and Chen (2013) conducted a case study of one tutor's feedback to two graduate ESL students in an online discussion forum. Their analysis of 47 comments revealed that the tutor used a variety of speech acts, leading to the researchers to conclude that asynchronous comments can be as interactive as synchronous ones; 92% of the comments resulted in uptake. The feedback formulated as interrogatives and hedges led to most revisions, and the learners reported increased awareness of language and revisions.

Despite the fact that the above studies concerns examining the nature of electronic feedback in second language learning and its impact on students' responses, as well as evaluating its effectiveness in remedying certain grammatical errors and encouraging interaction between students and the text, it is found that the nature of these studies does not appear to serve as good reference to the purpose of my research study. There are three reasons accounting for such irrelevance. First, we were employing very different feedback types, and none of their studies were focusing on testing the explicitness of corrective feedback on error correction. Second, our feedback environments were very different. Their studies were targeting at online conversation exchange whereas this study were targeting process writing which took the form of multiple drafts.

Third, our primary objectives were very different. Their studies primarily aimed at identifying relationship between feedback types and their frequency of use, or identifying the relationship between feedback types and their corresponding elicited output (which was not necessarily about accuracy). This study primarily aimed at evaluating student uptake (which was all about improved accuracy) of corrective feedback by different feedback types. For the sake of relevance, it is necessary to locate some studies of electronic feedback which examines the relationship between student uptake of teacher written feedback on different error types mediated by varying degree of explicitness. Such degree of explicitness is moderated by direct or indirect feedback (Lalande, 1982; Ferris, 1999, 2002, 2006; Ferris & Hedgocok, 2005; Lee, 2008a).

However, the research on 'Computer-facilitated feedback' (or electronic feedback) in second language writing (SLW) is very scarce. According to Ene & Upton (2014), "In particular, to our knowledge, there are no studies that describe comprehensively the use and impact on student writing of e-feedback (i.e. electronic feedback) by skilled teachers in non-experimental SLW classrooms' (p. 82). Having said that, there are some research studies on examining electronic feedback strategies used by tutors in written interactions with language learners in discussion forums (e.g. Zourou, 2011; Martin-Beltran and Chen, 2013; Samburskiy & Quah, 2014), but they are not included in the review here in that their feedback was targeted for the online discussion genre which is web-based and interactive-based, but not on essay writing with multiple drafts.

As such, the study conducted by Ene & Upton (2014) was the first study which examined the effectiveness of teacher electronic feedback in essay writing. The purpose of their study was to

examine what kinds of electronic feedback student received, and how they acted on teacher electronic feedback in two writing courses over two semesters where two essays of different genres (with the former being expository while the latter one being argumentative) were to submit and be evaluated in an electronic environment, for each of the two courses respectively. The composition took a process writing approach in which each essay (i.e. W1 & W2) was to be submitted twice (i.e. draft 1 & draft 2). Students made revisions on draft 2 of each essay based on the teacher electronic feedback indicated in draft 1. The same applied to two essays which resulted in two drafts for each essay. The participants were 12 non-native English speakers who had matriculated into undergraduate engineering programs in a Midwestern U.S. university. The teacher-markers were the three teachers who taught the participating students with an average of 10 years of experience teaching ESL composition.

For each paper submitted by a student, a teacher-marker would mark the paper electronically by means of 'directives', 'requests to change', 'provision of rules and explanation', 'provision of examples', 'indication of errors', 'provision of correction', or 'praise'. Without any rules imposed on the focus of error feedback nor their degree of explicitness, these various feedback comments were usually communicated electronically through inserting a 'bubble comment' on the margin of a paper. The number of feedback comments, by total and by category, was counted in draft 1 and draft 2, for each essay. Then student uptake, by total and by category, was tallied in draft 2. This was done by counting the number of errors, by total and by category, reduced from draft 1. The feedback categories were divided into 'content', 'organization', vocabulary', 'grammar', 'mechanics' and 'process'. 'Uptake' is conceptualized as learner revisions made in response to teacher feedback and is considered as a measure of feedback effectiveness (Lyster & Ranta, 1997;

Ellis et al., 2008; Storch & Wigglesworth, 2010; Ene & Upton, 2014). Thus, the number of student uptake was taken as the measure of the effectiveness of teacher e-feedback between draft and draft 2 of an essay respectively. The means of the student uptake from two groups (draft 2 of essay 1 in the first semester and draft 2 of essay 2 in the second semester) were calculated by category, and then paired t-tests were used to determine if any difference is statistically significant.

The results of Ene & Upton's study revealed that electronic feedback shared quite a number of feedback characteristics with written feedback. Most written feedback given by teachers was directive and consisted of marginal comments (Ferris et al., 1997), in the same way as most electronic feedback analyzed in Ene & Upton's study (2014) was directive (i.e. 27% in Essay 1 and 16.8% in Essay 2) and was placed in marginal comment bubbles. Ferris's studies (1997, 2011) show that written feedback primarily focuses on content and meaning-oriented. This outcome is the same as the electronic feedback analyzed in Ene & Upton's study (2014). Ene & Upton's study (2014) claims that direct explicit electronic feedback adopted in their study and the indirect coded feedback (or direct focused feedback) adopted in paper-and-pen feedback are both preferred by students and are both effective in resulting in more student uptake than indirect implicit feedback. They also claim that students' success with uptake matches the predictions of second language acquisition (SLA) theories, where teacher electronic feedback was noticed (Schmidt, 1990) and the presence of push output was evident (Swain, 1985, 1995).

However, such generalization of similarity between teacher written feedback and teacher electronic feedback might be susceptical to (a) logical flaws, (b) methodological flaws, and (c) lack of insightfulness:

(a) The findings about the characteristics of teacher written feedback (i.e. 'content-focused' and 'meaning-oriented') depicted by Ene & Upton (2014) are over-generalized. Even those feedback characteristics exhibited in the studies cited by Ene & Upton existed, it is not adequate to generalize to these results from these few studies to other studies, leaving an impression that the identical patterns exhibited in the majority of studies in teacher written feedback. In fact, quite a number of studies found that teacher feedback tends to be 'form-oriented' or 'grammar-focused' (Cumming, 1985; Zamel, 1985; Hyland, 2003; Lee, 2008a; Lee, 2008b), including this study.

(b) According to Ene & Upton's study (2014), it appears that neither the researchers imposed or defined a specific way that the teacher electronic feedback should be given nor the electronic feedback adopted in their study offered any distinctive marking features other than those of the traditional paper-and-pen marking. Teacher-markers in their study were seemingly providing their feedback 'electronically' via the 'review' function of Microsoft Word by typing up their feedback into the comment bubbles. This, by principle, remains no difference as if they were providing their feedback on paper. Despite the fact that the electronic feedback adopted in Ene & Upton's study was found to be effective with as much as 70% and 50% of successful student uptake in essay 1 and essay 2 respectively, any effects produced from electronic feedback should have been attributed to the teacher-markers' individual practice per se, but with little relevance to the nature of electronic feedback itself (i.e. the change in feedback strategies because of the change of medium from the paper mode to electronic mode which allows more flexible, autonomous and interactive feedback practice). In addition, no attention was paid to how the contextualization of feedback studies (see Chapter 3.23(H)) would have an impact on individual teacher feedback

practice, which might have in turn influence the way the electronic feedback was given in the study, namely teacher's belief, institutional dimensions, cultural dimensions, instructional dimensions, pedagogical dimensions and etc. To attribute any effects to the adoption of electronic feedback, researchers are advised to maintain the extraneous variables constant just as what they were done in this study. In addition, Ene & Upton attempted to compare the student successful uptake rate between essay 1 (an expository essay) and essay 2 (an argumentative essay) is ill-conceived in that essay 1 and essay 2 are of different nature and represent different levels of complexity, which in turn demand different levels of cognition and writing skills. As such, putting on par the student uptake of essay 1 with the student uptake of essay 2 for comparison is inappropriate. To remedy this methodological flaw, the researchers should have selected two writing tasks of the same type and the same level. For example, the first and second writings on a narrative description of 8 pictures with prompts in Truscott & Hsu's study (2008), and the first and second writing on composing a description based on a picture in Bitchener & Knoch's study (2010b).

(c) In Ene & Upton's (2014) study, there seems no new application of other SLA theories, apart from reiterating the three SLA theories (i.e. Long's Interaction Hypothesis, 1996; Schmidt's Noticing Hypothesis, 2001; Swain's Output Hypothesis, 1985) to teacher written feedback. To put this into perspective, there has been no discovery of other SLA theories or other learning theories that can form some sort of a new theoretical basis exclusive to electronic feedback. As such, it appears as if the only apparent difference between the electronic feedback being examined in Ene & Upton's study and the conventional paper-and-pen feedback is the feedback environment mediated by the mode of feedback itself. There was no mention if such adoption of electronic

feedback would change or facilitate the way feedback was conventionally given. For example, does the proposed electronic feedback offer a new feedback practice which is more effective in remedying errors or certain errors? Is the electronic feedback more conducive to second language writing? What are the students' perception towards electronic feedback?

In summary, Ene & Upton's (2014) study only confirms the value of electronic feedback just as paper-and-pen feedback, but fails to provide evidence if electronic feedback would have a potential to serve as a better alternative to the traditional paper-and-pen feedback. For example:

1. Does electronic feedback offer any distinctive feedback features which are different from or unrivaled by paper-and-pen feedback?

2. To what extent does electronic feedback change, reshape or enhance the way error feedback is given?

3. To what extent does the student uptake of teacher feedback on each error category respond differently to electronic feedback, when compared to paper-and-pen feedback?

4. What the students' perceptions of electronic feedback in identifying, understanding and correcting errors?

As far as the above is concerned, the potential of teacher electronic feedback is yet to be greatly explored, and the above questions are among those being raised in the research questions and being answered in this research study.

3.5.2   *Computer-generated Feedback*

As previously mentioned, despite the differences between 'Computer-facilitated feedback' and 'Computer-generated feedback' in terms of their technological advancement, both are regarded as the identical form of feedback treatment in supporting teacher feedback on writing, and they were being placed under a critical review in this study.

The focus of the critical review on the development of computer-generated feedback in this sub-section does not aim to make comparisons or draw conclusions about the relative effects of various AWE programs, nor any AWE program with 'Mark My Words' ('MMWs') as a kind of electronic feedback. Instead, this review aimed to identify issues, in particular to the methodological flaws of some AWE studies, which might have rendered the effects of computer-generated feedback invalid or inconclusive in their research studies. Hence, the identification of these issues has enabled this experimental study to (1) identify the research gaps and (2) avoid these flaws in this experimental design, thus enhancing its reliability and validity. These flaws are to be discussed in this sub-section.

A number of research studies (e.g. Milton, 2006; Lee, 2008a) suggested that integration of technology can improve academic performance, enhance motivation, and promote learning. Therefore, the use of technology might offer an alternative way of giving teacher written feedback in student writing more effectively and efficiently. To put this into perspective, the pedagogical

use of computer-based educational technology for providing students with feedback on their writing is commonly known as 'Automated Writing Evaluation' ('AWE') (Stevenson & Phakiti, 2013). A few 'AWE' software programs (see *Figure 3.5*) have been developed, for example 'Criterion' by ETS, that can automatically detect and identify certain error types, mostly related to grammar, mechanics and organization errors. And most of them can provide holistic scores for the writing texts of certain designated topics (ibid).

		Examined by # of research studies			Examined by # of research studies
1.	Criterion	11	6.	LSA semantic space	1
2.	My Access	5	7.	Summary Street	3
3.	Writing Roadmap	1	8.	ECS	1
4.	ETIPS	2	9.	SAIF	1
5.	IEA	1	10.	QBL	4

Figure 3.5   Types of Automated Writing Evaluation Tool

However, a review of research articles on examining the effectiveness of 'AWE' software programs (e.g. Chen,1997; Warden, 2000; Hyland & Hyland, 2006; Warschauer, 2006; Cheng & Cheng, 2008; Lee et al., 2009 & 2013; Schroeder et al., 2008; Grimes, 2008; Chodorow et al, 2010; Wang & Wang, 2012; Stevenson & Phakiti, 2013) suggested that the effectiveness of 'AWE' feedback on student writing is not conclusive due to paucity of research, heterogeneity of existing research, the mixed nature of research findings, and the methodological issues in some of the existing research.

3.5.2.1 Paucity of research

Stevenson & Phakiti's study (2013) on the development of computer-generated feedback has revealed that only a relatively small number of research, around thirty studies on ten 'AWE' programs, were conducted to investigate the effects of computer-generated feedback on student writing revisions. As stated by Warschauer (2006), such small number has been possibly due to the fact that 'AWE' is a relatively newcomer to the field of second language writing. Within this body of research studies, 'scores' (24 studies) and 'error frequency' (5 studies) have been the most common measurements to the effectiveness of 'AWE' feedback, whereas the remaining one is about 'citations' (Stevenson & Phakiti, 2013). It is because the capability to improve the quality of students' texts, which is more easily evident by 'scores' and 'error count', is central to claims made about the effectiveness of 'AWE' feedback. As such, it is hoped that this experimental study can contribute to our growing understanding about the effects of computer-generated feedback on student writing revisions. 'Error count' which is one of the most common measurements, was adopted to evaluate the effectiveness of teacher feedback in the experimental and control groups.

On the other hand, the majority of 'AWE' studies were done in the US college setting. In Hong Kong, there seems to have only two 'AWE' studies carried out so far, in which the feedback focus was primarily on 'content' and 'organization' (Lee et al., 2009, 2013). In light of this, more investigations into the effects of computer-generated feedback on error reduction for Hong Kong students can fill in the niche.

3.5.2.2 Heterogeneity of existing research

As shown in *Figure 3.4*, a number of studies have been conducted to investigate the effects of 'AWE' programs on student writing revisions. Indeed, among those 'AWE' programs, they offer various kinds of feedback with different focus to users. As stated by Stevenson & Phakiti (2013), 'Criterion' and 'My Access' provide feedback on language use and global writing skills, 'QBL' provides feedback on language use, and 'Summary Street', 'IEA', 'LSA' and 'ECS' provide feedback primarily on content knowledge. *Figure 3.6* below summarizes the various focus of 'AWE' programs:

1.	Criterion	They provides scores and feedback on both content and language.
2.	My Access	
3.	Writing Roadmap	
4.	Summary Street	They are based on a technique known as latent sematic analysis that purports to focus primarily on content feedback
5.	IEA	
6.	LSA	
7.	ECS	
8.	ETIPS	It provides feedback for pre-service teachers on tasks carried out in an online case-based learning environment.
9.	SAIF	It provides feedback on the citations in a text.
10.	QBL	It provides comments on language errors only.

Figure 3.6   Heterogeneity of AWE's Feedback Focus pointed out by Stevenson Phakiti (2013)

In addition, most 'AWE' research investigated students from diverse backgrounds in diverse teaching contexts. Due to such heterogeneity of feedback focus and contexts, it would appear difficult and broad-brush to draw a general conclusion on the effectiveness of 'AWE' on student writing revisions as (1) each of those programs might entail different feedback features and functions for the purpose of different outcomes, (2) the responsiveness of different students towards different 'AWE' feedback might be different under different circumstances, and (3) the student uptake of different error types (e.g. 'content', 'organization' or 'language') on different AWE programs with different focus might vary. Rather, it would seem more meaningful to investigate in what ways and to what extent a particular feedback mode, with the help of computer-generated technology, would be more effective in helping particular kind of students, under a specific instructional context, to identify, comprehend, correct and avoid language errors of various types. This was where the primary interest of this experimental study lies with, and this was the way the research design of this study was being set up.

3.5.2.3 The mixed nature of research findings

Partly due to the heterogeneous issues discussed above, variation in the kinds of feedback alongside their relative effects on student writing revisions has not only rendered the effectiveness of 'AWE' programs inconclusive, but also has given rise to controversy over the value of 'AWE' feedback. On the one hand, some researchers see 'AWE' as releasing language instructors from the workload of marking so that they could devote more time to writing instructions (Hyland & Hyland, 2006; Philips, 2007). Dikli (2006) acknowledged the positive effects of 'AWE' on student writing revisions due to the immediate online feedback. Cheng & Cheng (2008) also stated that

'AWE' has positive effects on student autonomy. On the other hand, the effectiveness of integrating computer technology in providing effective writing feedback has aroused considerable suspicion (Stevenson & Phakiti, 2013). Anson (2006) stated that computer lacks human inferencing skills and background knowledge to score human texts. Hyland and Hyland (2006) argued that the comments provided by 'AWE' have placed too much emphasis on surface features of writing such as grammaticality. While it may sound ethereal to expect 'AWE' to be omnipresent or completely replace human marking, this experimental study attempted, in a pragmatic manner, to identify if the electronic feedback would complement the short of the papered-based feedback practice.

3.5.2.4 Methodological Issues & Misinterpretation of Findings in other Studies

A number of methodological issues of 'AWE' research studies were identified (Chen, 1997; Grimes, 2008; Schroeder et al., 2008; Chodorow et al., 2010; Stevenson & Phakiti, 2013). These issues might have possibly undermined the validity and reliability of their findings.

Effects of 'AWE' feedback are not always based on empirical evidence but at times appear to reflect author's own 'techno-positivistic' or 'technophobic' stance towards technology in writing classrooms (Stevenson & Phakiti, 2013). As such, an empirical approach was undertaken in this experimental study. It was reported by Stevenson & Phakiti (2013) that some studies have largely ignored the negative evidence drawn from the findings, and hence have drawn conclusions about the positive effects of computer-based feedback that are more optimistic than they should have

appeared. For example, in a study conducted by Schroeder et al (2008), it was found that one of the three experimental groups using 'AWE' feedback (i.e. 'Criterion' program) didn't receive significantly higher final grades than the control. However, the reasons accounting for such insignificant outcomes were not provided, and the researcher rather arrived at a firm conclusion that "results from this study overwhelmingly point toward the value of technology when teaching writing skills" (p.444). Another example showing the ignorance of researchers, be it deliberate or not, is the study conducted by Chodorow et al. (2010), in which it was concluded that 'AWE' feedback (i.e. 'Criterion' program) was significantly more effective in reducing 'article' errors only for non-native speakers' group but not for the native speakers' group. However, according to Stevenson & Phakiti (2013), the study did not provide the 'article' error rate respectively for the native-speakers' group and non native-speakers' group, and did not raise the point that 'AWE' feedback might seem less effective for native speakers for that native speakers tended to make much fewer 'article' errors. In this case, "the lack of a significant effect for native speakers should not be taken at face value as negative evidence for the effectiveness of 'AWE' (ibid, p.60).

On the one hand, *Figure 3.7* below (adapted from Stevenson & Phakiti, 2013) illustrates the form of feedback treatment received by control groups across those thirty 'AWE' studies. It can be seen that the most common form of feedback treatment received by these control groups in the studies is 'no feedback' at all; and eight of these studies among these thirty had no control group. Among these thirty 'AWE' studies, only three studies were explicitly compared 'AWE' feedback with teacher feedback. With as many as 26 studies comparing the experimental groups using 'AWE' feedback with the control groups either receiving no feedback or even in the absence of any control

group, any positive or negative claims drawn from these research studies might appear vulnerable to challenges of their validity and reliability. It sounds normal or natural for any groups receiving feedback performed better in their quality of writing revisions than those receiving no feedback; and also the kind of comparison made in these studies might have offered little value and weight to any research outcomes because improvement made by 'AWE' users to successive drafts could have been attributed to other extraneous factors other than the 'AWE' feedback itself.

Experimental Group (form of feedback treatment received)	Control Group (form of feedback treatment received)	# of studies
'AWE feedback'	No feedback	18
	Teacher feedback	3
	Different 'AWE' conditions	1
	No control	8
Total # of studies		30

Figure 3.7   Form of feedback treatment received by experiment and control groups

On the other hand, even when I reviewed those studies with control groups, it was found that the comparability of students in experimental groups and control groups was in doubt. For example, in Grimes' (2008) studies, the results indicated that students who used 'My Access' (a 'AWE' program) had higher external test scores than students from a previous year who did not receive 'My Access' feedback. Despite the positive effects, Grimes acknowledged that it was difficult to attribute the positive effects to the 'AWE' feedback as Grimes also could not exclude the possibility that such language gain might have been attributed to improvements of the writing instructions provided by the teachers in the experimental group during the intervention period.

Another study conducted by Chen (1997) was identified with a similar problem. In Chen's study, it was found that both experimental group receiving 'AWE' feedback and the control group receiving no feedback were equally able to reduce errors in the second draft. However, such positive effects could not be attributed to the merits of the 'AWE' feedback as both groups, in the same class, had the teachers spending time to review the most common error types identified by the computer in the presence of all students including those in the control group. Hence, the improvement in error reduction might not be possibly attributed to the use of 'AWE' feedback itself but the instructions in class.

Another methodological flaw I identified is about the inadequate comparison of feedback treatments with different feedback focus. The 'AWE' study conducted by Warden (2000) is an example above, whose flaws were quite similar to the paper-and-pen feedback study conducted by Ferris (2006). Warden's experimental group receiving the 'AWE' feedback was provided with specific error feedback, whereas its control group receiving the teacher feedback was only provided with general comments on 'content', 'language' and 'grammar'. This means the nature or kind of feedback received by both groups was not comparable. As such, any language gain obtained from a particular group should not be attributed to the form of feedback treatment (i.e. 'AWE' feedback or teacher feedback), but more likely to the feedback practice (i.e. feedback focus and specific/general feedback) themselves.

The last methodological flaw identified is that a number of the studies provided only vague and little descriptions of their participations such as language background and literacy; for example, the study conducted by Elliot & Mikulas (2004). In one study (Wang & Wang, 2012), only 1 participant in each condition was involved; hence the sample size is in doubt. In another 'AWE' study conducted by Lai (2010), no information about the proficiency level of the 21 Taiwanese college participants was provided except their gender and age.

3.5.2.5 Implications on the Methodological Design of this Study

As far as the above is concerned, it appears most of the computer-mediated feedback studies (including both computer-facilitated feedback in Chapter 3.5.1 and computer-generated feedback in Chapter 3.5.2) were not rigorously designed. The methodological issues discussed have rendered the researchers difficulties in drawing a firm conclusion as to the effectiveness of the computer-generated feedback on written production.

As such, efforts were made to avoid the impact of these extraneous factors on the validity and reliability of the research outcomes (see Chapter 5.4.1 'Measures for Minimizing Methodological Flaws'). It was in this way any positive or negative claims drawn from the findings would become more meaningful in the sense that any language gains evident in revisions could be attributable to the use of the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krahsen's (1985) 'Input Hypothesis'.

Chapter 4   Rese arch Questions and Background of the Study

4.1   Research Problems

Research on teachers' beliefs has indicated that beliefs have an important impact on teacher practice. However, little is known in teacher feedback research about teachers' beliefs and the extent of which they influence practice (Lee, I. 2008a). "Teachers should be encouraged to undertake action or classroom-based research and to share good feedback practices. Entrenched feedback practices are difficult to change, but if teachers' classroom research shows that alternative feedback practices can lead to better student motivation, more effective learning, and even improvement in student writing, teachers gradually deconstruct their preferred ways of feedback. In Hong Kong, and perhaps in many similar contexts, however, professional development work on feedback is in its infancy. This is a long way to go (ibid)".

My continuous self-reflection of my own teaching has prompted me to investigate the effectiveness of teacher written feedback with an ultimate goal to empower students with the necessary skills to become more independent, autonomous and effective writers.

To put this into perspective, the following questions have been raised. What are some common written feedback strategies adopted by language teachers? To what extent these written feedback strategies adopted by language teachers have effectively helped students identify, comprehend and correct errors in process writing, as well as helped them avoid the identical errors in their future

writing. If they are not effective, what are the limitations or constraints making the feedback practice ineffective; and in what ways the applications of those language theories and learning theories can help reduce these limitations and constraints in error reduction? Can teacher written feedback be more student-centered such that students can take a more active role in the feedback process under which individual learners' language needs and learning preferences are accommodated?

4.2   Research Questions

With the above questions in mind, the study undertaken endeavoured to address the following research questions, in which the first and second research questions remain the primary focus of the study, whereas the third one is the secondary:

(1) Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a difference on student writing revisions, when comparing to the paper-based feedback?

(2) Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a difference on student writing revisions for each error category, when comparing to the paper-based feedback?

(3) What were the students' perceptions of the effectiveness of teacher written feedback?

4.3   Clarification of Terms

The two common types of error correction used by second language teachers are 'direct feedback' and 'indirect feedback', in which the former refers to teachers' provision of correct answers in response to student errors whereas the latter refers to teachers' indication of errors (by means of a circle, an underline, a code or a mark) with correction by students (Lee, 2008a).

According to Ferris (2006), "direct feedback may take various forms including crossing out an unnecessary word, phrase, or morpheme; inserting a missing word or morpheme; or writing the correct word or form near the erroneous form. Indirect feedback occurs when the teacher indicates in some way that an error has been made by means of an underline, circle, code, or other mark but does not provide the correct form, leaving the student to solve the problem that has been called to his or her attention" (p. 83).

Direct feedback is used when teachers perceive the error in question is complex in a way that it is beyond students' ability for self-correction (Ferris, 1999; Frodesen, 1991), whereas indirect feedback is used when teachers want to engage students in problem-solving and develop their independent editing skills (Ferris, 2002; Ferris & Hedgocok, 2005; Lalande, 1982).

There are two types of indirect feedback which are coded feedback and uncoded feedback. Coded feedback indicates an error by pointing out the types of errors made such as 'preposition' or 'tense', whereas uncoded feedback indicates an error by simply underlining or circling the error without indicating the error type (I, Lee, 2008a).

In addition, for easy reference and clear differentiation in the study, the following terms alongside their meanings were used throughout the thesis:

Table 4.1 Clarification of Terms

1.	Types of Feedback	Grades or Error Feedback or Written Comments
2.	Focus of Error Feedback	Content or Organization or Language
3.	Form of Feedback	Paper-based Feedback or Electronic Feedback or others
4.	Feedback Strategy	Underline Errors and/or Circle Errors and/or Categorize Errors and/or Provide Examples and/or Provide Correction
5.	Feedback Practice	A general term refers to a particular feedback strategy adopted by a teacher-marker.

4.4   Background of the Study

Seven years ago, the Director of the Language Center of the Hong Kong University of Science and Technology commissioned the IT team, in which I was one of the members, to develop various online writing pathways which could help students to improve their language proficiency while fostering learner's autonomy and independence in language learning.

As part of the efforts in designing the writing pathways, I was tasked with refining the pedagogical aspect of (instead of the technological aspect) and promoting an electronic marking tool called 'Mark My Words' ('MMWs') developed by the IT team.

I was tasked with piloting this electronic marking tool ('MMWs') in an English language course for the second-year engineering students (course code: LANG 206) with the following objectives:

1.   To determine the effectiveness of the electronic marking tool ('MMWs')

2.   To compare the effectiveness between the traditional paper-based teacher written feedback and the electronic-based teacher written feedback ('MMWs)

In coincidence with my research interest in the field of 'Writing Feedback', investigation on the effectiveness of this electronic marking tool ('MMWs') become the topic of my Ed.D. study.

4.4.1   Teacher Beliefs behind 'Mark My Words' ('MMWs')

The purpose of developing 'Mark My Words' ('MMWs') in LANG 206, along similar lines to my teaching belief in a second language learning classroom, is to empower students with the self-learning skills to become independent, autonomous and effective writers in the long run.

'Mark My Words' ('MMWs') is an MS Word add-on, which goes beyond a mere electronic marking tool providing electronic coded feedback in the form of a pop-up window. Its electronic coded feedback also makes web-based functions and resources available to students during the writing process and as part of the feedback on their written assignment. These integrated web-based functions and resources include the 'lexis-and-grammar checking' functions as well as the 'English Grammar Guide' (EGG) whereby students can check for accuracy while composing or editing their texts throughout the process writing.

This electronic marking tool ('MMWs') was developed and integrated, throughout my course, to support students during their writing process under the 'Discovery-based' learning principle, hoping that students could become more independent, autonomous and effective language writers in the long run.

4.4.2   The Pedagogical Operation of 'Mark My Words'

Since my thesis topic is about the use of an electronic marking tool called 'Mark My Words' ('MMWs') and this electronic marking tool is the 'feedback environment' where this experimental study and all other related activities took place, it is necessary, before presenting the research design, to explain how 'MMWs' functions and operates in a technical sense on one hand, and how the design of 'MMWs' reflects the theoretical propositions about error treatment and good feedback practices on the other. This section includes the instructions on the use of 'MMWs', featuring the demonstration of the so-called 'electronic feedback'.

After the 'MMWs' software is downloaded (with authorization) into a computer, 'MMWs' will become an add-on toolbar on the Microsoft (MS) Word (see *Figure 4.1*).

Figure 4.1

For instance, a teacher-user would like to correct the wrong usage/form of 'concern' in the sentence below (see *Figure 4.2*).
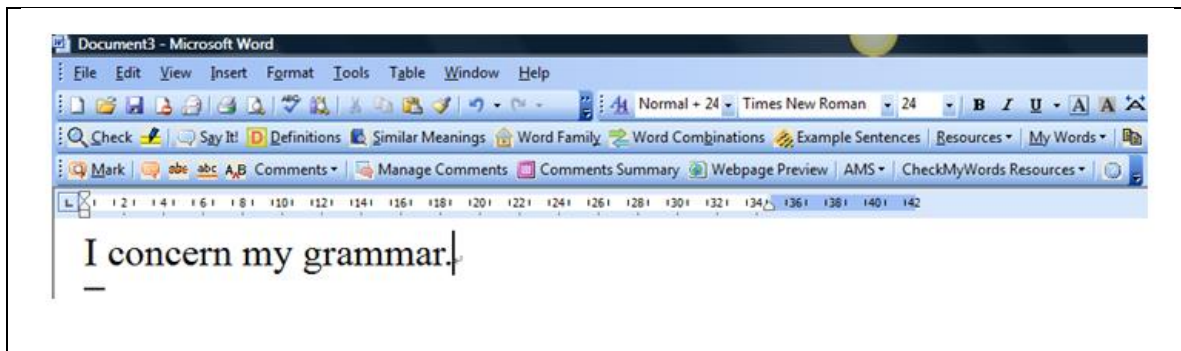
Figure 4.2

The teacher-user has to highlight the error 'concern', go to the 'MMWs' toolbar and click on 'Comments'. Then s/he has to select the corresponding lexico-grammatical item by browsing and choosing the relevant cell, 'Adjective-Verb Confusion' (see *Figure 4.3*).
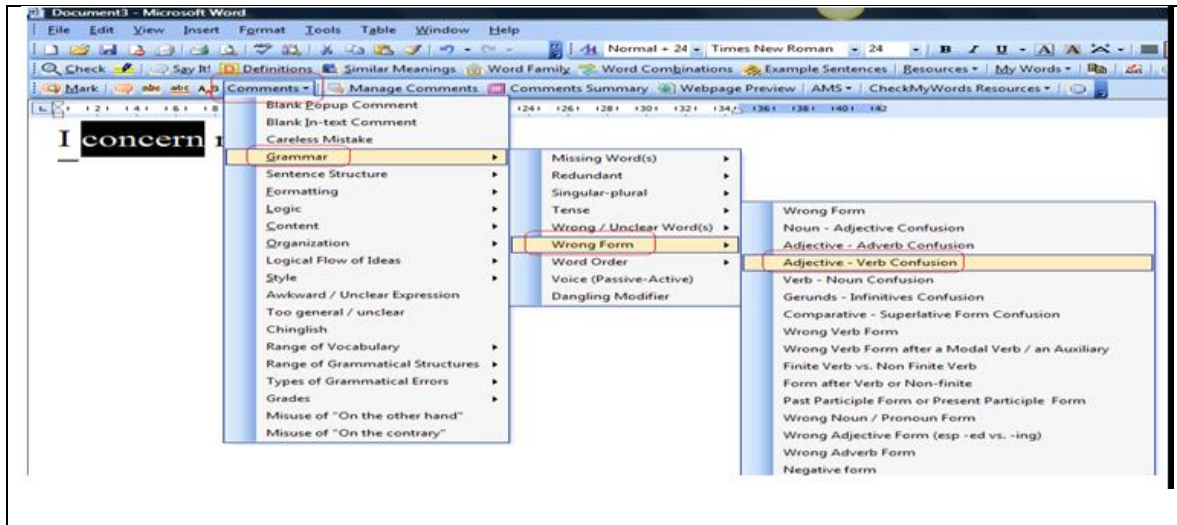
Figure 4.3

After selecting the 'Adjective-Verb Confusion' cell, a comment window will pop up and show the

pre-written explanation of the highlighted error (see *Figure 4.4*).
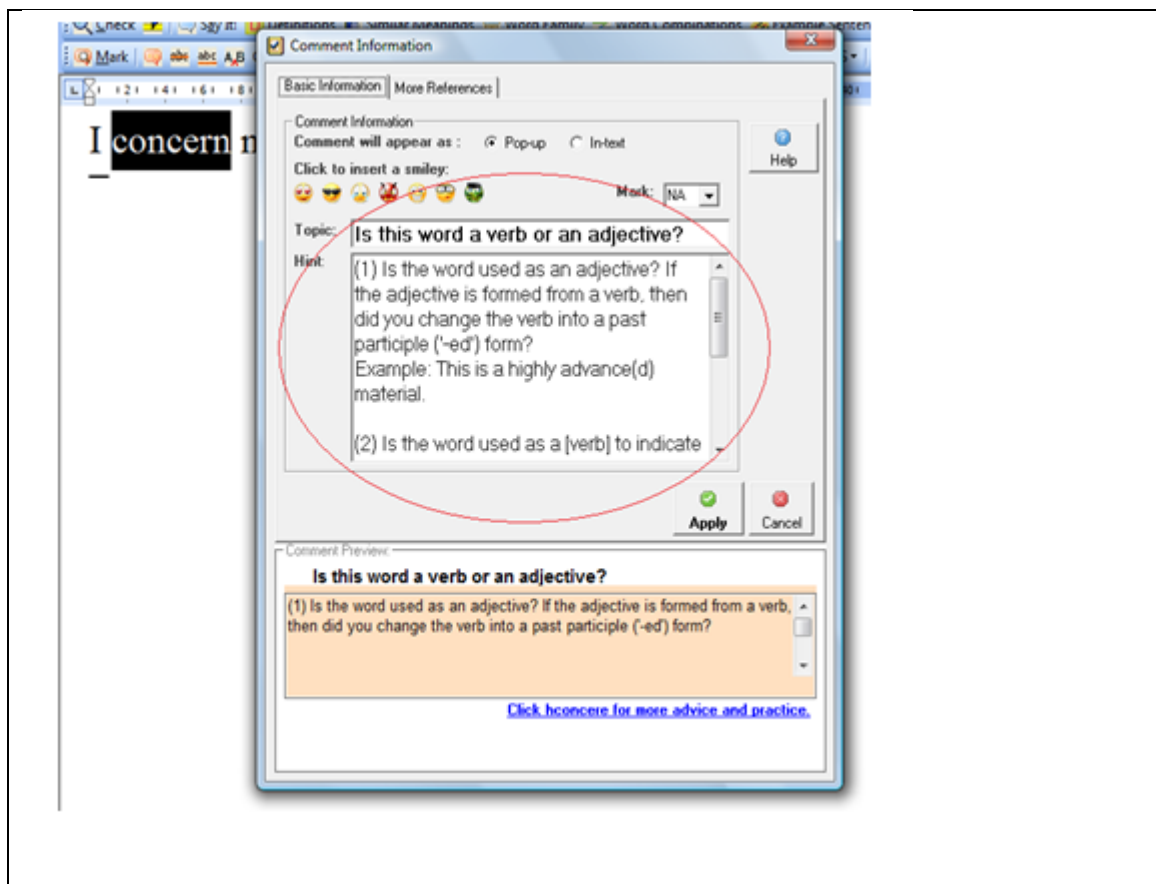
Figure 4.4

Click on the 'More References' tag, and then the relevant page of the 'English Grammar Guide' ('EGG') and the 'Google Link' which show the correct and contextualized usage of 'concerned' as an adjective form will be embedded in the electronic coded feedback (see *Figure 4.5*). All information and references are pre-written in a way that teacher-user can easily and conveniently adopt them completely or partially for their own use when giving feedback. If the information or reference is determined by the language teacher as adequate or applicable, the teacher-user can finally click on the 'Apply' button.
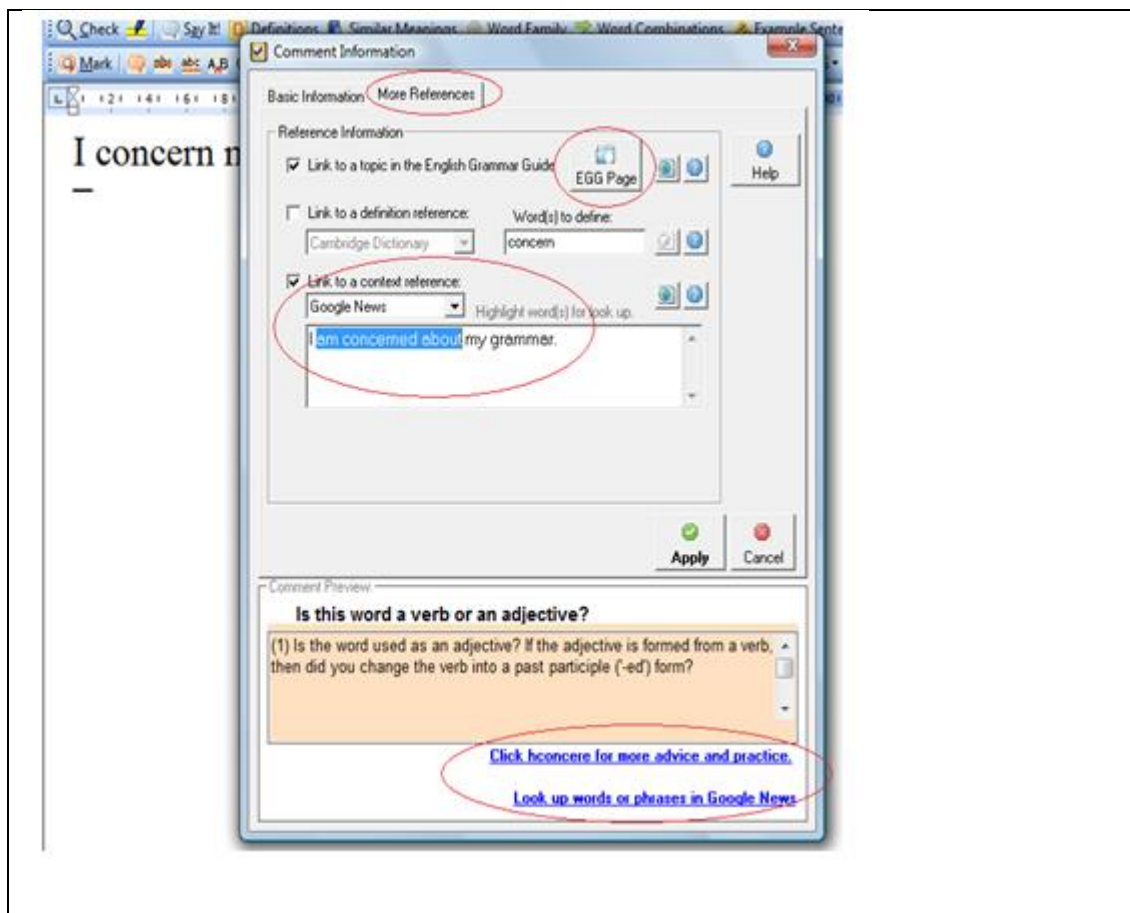
Figure 4.5

When receiving the electronic coded feedback, students will find that each error is identified with a code that comes with an error identification, explanation, and web-based resources (i.e. [u1] in this example). When moving the cursor to the code, a comment window which identifies and explains the highlighted error will pop up. In addition, a student can click on the 'Google Link' (to see the adequate use of the highlighted language item 'concerned' in an authentic source from Google) or click on the 'more advice and practice' to check out the 'English Grammar Guide' for detailed explanation and practice (see *Figure 4.6*).

121
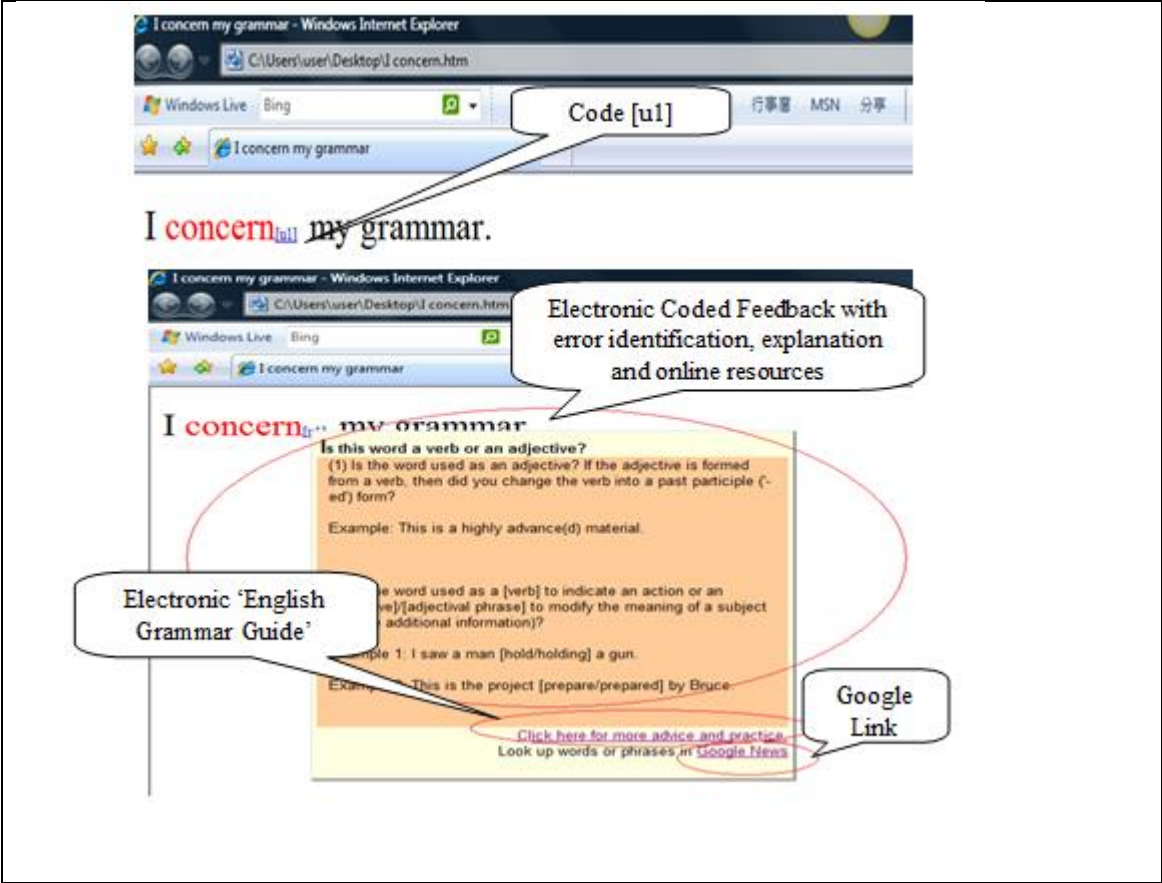
Figure 4.6

Chapter 5   Research Methodology

As suggested by Antwi & Hamza (2015), "All research is based on some underlying philosophical assumptions about what constitute a valid research and which research method(s) is/are appropriate for the development of knowledge in a given study. The selection of research methodology depends on the paradigm that guides the research venture" (p.217-218).

Section 5.1 will discuss the philosophical assumptions of two generic or conventional research paradigms within the educational research context, and conclude with identifying a particular research paradigm which is suitable to the research topic, purpose and objectives of this research study in Section 5.2. The reason for exploring the philosophical assumptions of research paradigms is that a research paradigm forms the theoretical basis for guiding a research process: research design (that guides how the research is set up), research methods (that determines the way(s) the data is collected), and data analysis (that provides the techniques for analyzing data). By the end of this chapter, it should be able to state, describe and justify the research paradigm alongside its research design, research methods and data analysis underpinning my thesis.

5.1   Research Paradigms: Ontology, Epistemology & Methodology

According to TerreBlanche and Durrheim (1999), the research process has three major dimensions: ontology, epistemology and methodology, which form a 'research paradigm' that defines the nature of enquiry and "guides the entire research process including strategies, methods and analysis" (Antwi & Hamza, 2015, p.218). 'Paradigm' in the context of research methodology refers to a set of philosophical assumptions about the phenomena to be studied, about how they can be understood, and even about the proper purpose and product of research (Hammersley, 2012). Hitchcock and Hughes (1995) describe the logical relationship about the philosophical assumptions of these three dimensions in this way: ontological assumptions, which are the assumptions about the nature of reality and nature of things, give rise to epistemological assumptions, which are ways of researching and enquiring into the nature of reality and the nature of things; these two assumptions in turn give rise to issues of instrumentation and data collection (p.21).

When considering which research methodology to be used for conducting a research study, it is of importance to understand and give some thought to research paradigms, and then adopt a paradigm that suits your understanding of the phenomena of the research context being studied and the nature of enquiry (e.g. does the intervention make an improvement in scores vs how did the intervention make such improvement). Research paradigms, namely positivism, post-post-positivism, interpretivism, critical theory, etc., are ways of explaining the basic set of beliefs (i.e. philosophical assumptions) that a researcher has and how these influence the way a researcher conducts research.

The selection of a research paradigm for a research study depends very much on how a researcher perceives the construction of social reality. According to Cohen et al. (2011), the nature of social reality can be represented by two generic conceptions, namely 'positivism' (or called 'normative paradigm' as one of the categories subsumed under 'positivism' (p.17), and 'anti-positivism' (or called 'interpretive paradigm' which will be generally used here) as one of the categories subsumed under 'anti-positivism' (p.17).

Positivists take the view that social reality is external to individuals which is 'out there' in the world, and that knowledge is hard, objective and tangible which will demand of researchers an observer role. For its implications on the research methodology, positivists tend to analyze the relationships between selected factors (i.e. variables) and its research design tends to be quantitative and is concerned with identifying and defining causation relationship between variables with the goal of searching for universal laws that can explain and govern which is being observed. Interpretivist researchers take the view that social reality is the product of individual consciousness and cognition which is created or interpreted by one's own mind, and that knowledge is personal, subjective and unique. For its implications on the research methodology, Kirk & Miller (1986) point out that anti-positivism places emphasis on the explanation and understanding of individuals and relativistic social reality rather than collectivity and absolute social reality, and its research design relies on exploring a subjective relationship between the researcher and subjects such as interviewing or participant observation (Antwi & Hamza, 2013).

To examine the philosophical differences between the positivist paradigm and interpretive paradigm in detail, they will be analyzed from the ontological, epistemological and methodological perspectives in the following subsections.

5.1.1   Ontological Dimension of the Research Paradigm

Ontology refers to "a branch of philosophy concerned with articulating the nature and structure of the world" (Wand and Weber, 1993, p.220). Antwi & Hamza (2015) added that "it specifies the form and nature of reality and what can be known about it" (p.218). To put it simply, it concerns with the question of what is real, or a researcher's worldview.

The positivist paradigm of exploring social reality is based on the philosophical ideas of the French Philosopher, August Comte, who argues that observation and reason are the best ways to understand human behavior and true knowledge, which can be obtained by observation and experiment (Antwi & Hamza, 2013, p.218). From the ontological perspective, "positivists assume that reality is objectively given and is measurable using properties which are independent of the researcher and instruments; in other words, knowledge is objective and quantifiable" (ibid, p.218). In other words, the ontological assumption is single and objective (Creswell, 1998). Similarly, Henning, Van Rensburg & Smit (2004) argues that positivism deals with discovering the truth and presenting it empirically. This is conducted by measurement oriented methodologies (as opposed to meaning-oriented methodologies) such as empirical examination, hypothesis testing or questionnaires, which identifies statistical relationships among variables (Antwi & Hamza, 2013).

On the other hand, interpretive paradigm takes the view that the understanding of social reality is sought by observation and interpretation, under which observation refers to the collection of information about events, while interpretation refers to making meaning of that information by drawing inferences or by judging the match between the information and some abstract pattern (Antwi & Hamza, 2013). Deetz (1996) argues that interpretive paradigm attempts to understand phenomena through having participants to assign meanings to them. In other words, the ontological assumption is multiple and subjective (Creswell, 2005). This is conducted by using meaning-oriented methodologies such as interviewing or participant observation, that rely on a subjective relationship between the researcher and subjects (Antwi & Hamza, 2013).

In summary, the positivist paradigm explores the understanding of social reality through devising a universal rule which governs human behaviors being examined, and this is conducted by empirical experiments. The interpretive paradigm is concerned with understanding social reality as it is from subjective experiences of individuals, which are collected by interviews or observations.

5.1.2   Epistemological Dimension of the Research Paradigm

Epistemology refers to the nature of the relationship between the researcher (the knower) and the known (Antwi & Hamza, 2015). It poses the questions about what is the relationship between the knower and what is known, how do we know what can be known, and what counts as knowledge.

Broadly speaking, there are two epistemological positions which are 'positivism' and 'interpretivism' (ibid).

Positivists view the study of reality as an organized method that combines deductive approach with empirical observations of collective behavior in order to derive a rule or a set of rules that can govern and predict the general patterns of human activity (Neuman, 2003). The study of reality is based on empirical facts rather than personal and subjective experience, and the phenomena is generally governed by cause and effect (ibid). As such, the person contact and interaction between the researchers and the subjects being studied is alienated. The basic assumption of this paradigm is that the goal of science is to use the most objective methods to approximate reality (Ulin, Robinson & Tolley, 2004). According to Antwi & Hamza (2015), researchers adopting positivist paradigm "explains in quantitative terms how variables interact, shape events, and cause outcomes" (p.219).

Contrary to its positivist counterpart, interpretivist view the study of reality as being constructed, interpreted and experienced by individuals through their interaction with others under a relatively less controlled setting (Maxwell, 2006; Guba & Lincoln, 1985). There tends to be a closer contact and interaction between the researchers and the subjects being studied. The basic assumption of this paradigm is that the goal of research is to explore and understand a phenomenon by gaining deeper insights as to how individuals interact with the reality being studied. The depth of such insights is the key but not the prediction of any general patterns (Antwi & Hamza,2015; Ulin, Robinson & Tolley, 2004; Neuman, 2003). According to Merriam (1998), researchers adopting

the interpretivist paradigm study the reality inductively through firsthand experience, truthful reporting and quotations of actual conversation from insiders' perspectives than identifying and testing if there is a rule governing general patterns of human activity (Bryman, 2001). As such, "interview, focus group discussion and naturalistic observation are the most widely used data gathering methods for researchers using qualitative research methodology" (Antwi & Hamza, 2015, p.219)

5.1.3 *Methodological Dimension of the Research Paradigm*

Methodology refers to the practical ways for a researcher to find out what he believes can be known (Antwi & Hamza, 2015). It is a research strategy that translates ontological and epistemological principles into practically how a research study is conducted (Sarantakos, 2005).

The positivist paradigm "requires a research methodology that is objective where emphasis is on measuring variables and testing hypothesis that are linked to general causal explanations" (as cited in Antwi & Hamza, 2015). As such, positivist research adopts a quantitative research approach by using "experimental designs to measure effects" and "the data collection techniques focus on gathering hard data in the form of numbers to enable evidence to be presented in quantitative form" (as cited in Antwi & Hamza, 2015, p.220). The verification of the collected data is conducted by statistical analysis (Bryman, 1998).

Contrary to its positivist counterpart, the interpretivist paradigm adopts a qualitative research approach which assumes that reality is socially constructed and that the meaning is created by individual participants as a result of their experience and interaction with the reality through the mediation of a researcher (Merriman, 1998). The role of a researcher under the interpretivist paradigm has a closer contact with the participants being studied in the sense that not only the researcher is taking an observant role but the researcher often participates in activities, interviews the subjects, constructs case studies and analyze existing documents (Antwi & Hamza, 2015). Qualitative researchers' goal is to explore and understand the views of insiders with a group under study.

5.2   Selected Research Paradigm: Justifications for the Mixed Methods Paradigm

After evaluating the ontological, epistemological, methodological assumptions of the positivist paradigm and interpretivist paradigm against the suitability to answering the research questions, it is found that the positivist paradigm is more suitable to addressing research questions (1) and (2), while the mixed-methods paradigm is applicable to research question (3). These three research questions, where (1) and (2) serve as the primary objective whereas (3) serves as the secondary (or supplementary) objective of this research study, are restated as follows:

(1) Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a

difference on student writing revisions, when comparing to the paper-based feedback?

(2) Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a difference on student writing revisions for each error category, when comparing to the paper-based feedback?

(3) What were the students' perceptions of the effectiveness of teacher written feedback?

The primary objective of this study was to test out a hypothetical assumption; that is, if and to what extent the electronic feedback modeled on Nunan's (1997) and Krashen's (1995) theories would make statistically significant improvement in error reduction, by total and by each error category, over student writing revisions (i.e. Research Questions (1) and (2)). Through testing this hypothesis on a total of 62 students including both experimental and control groups in this study, it was hypothesized that the results of which can validate the effectiveness of electronic feedback which is built on these theories on error correction, and can generalize the results to a larger population. With this primary interest setting on this, this lends weight to what Cohen et al. (2011) describe the suitability of adopting the positivist paradigm in which "the researcher's ultimate aim is to establish a comprehensive 'rational edifice', a universal theory, to account for human and social behavior" (p.18). In other words, this study attempted to establish the cause-and-effect rule that governs a statistical relationship between 'using electronic feedback' as a cause for the' improved error reduction' as an effect. While the testing out of these theories on my participants was intended to provide a general and broad picture if the electronic feedback statistically works better for recipients who participated in the study, it is believed that the some sort of the recipients'

comments (as the supplementary data) should also be collected to explain the general picture of the results for further understanding of the phenomena, for example, the reasons accounting for merits or demerits of electronic feedback. As such, the research methodology adopted here also extended to the interpretivist paradigm in which subjective realities are also valued. After an exploration of the two generic and conventional research paradigms in the previous sub-sections, the mixed methods paradigm was selected for this research to best focus on the above research questions on the premise that this mixed paradigm would build on the strengths of both conventional paradigms. The philosophical assumptions of the mixed method paradigm will be discussed below.

5.2.1 Mixed Methods Paradigm: Purpose, Research Design, Methods & Analysis

Researchers in favour of mixed methods design argue that using both confirmatory method and exploratory method are of significance in one's research (Johnson & Onwuegbuzie, 2004). Such mixture implies philosophical assumptions of both generic paradigms were called into play in this study. The following explains the implications of the positivist paradigm and interpretative paradigm on the corresponding research design (i.e. experimental design: quantitative supplemented by qualitative approach), data collection instruments (i.e. research methods), data analysis (i.e. mainly t-tests supplemented by coding) and research purpose (i.e. mainly causal explanation). The positivist research adopts the quantitative research design which primarily follows the confirmatory scientific method on hypothesis testing and theory testing. As pointed out by Antwi & Hamza (2015), "Quantitative researchers consider it to be of primary importance

132

to state one's hypotheses and then test those hypotheses with empirical data to see if they are supported" (p. 220). As such, 'performance measures' (in the form of error count) and 'attitudinal measures' (in the form of administering questionnaires) were used as the major instruments (refer to Chapter 5.2.2) in the research design under which the frequency of errors before and after an intervention (i.e. error count modelled on Ferris's (2006) study) were collected and subsequently categorical questions (modelled on Lee's (2008) study) were asked, and finally statistical tests (i.e. t-tests) were run to analyze if there was a significant relationship between variables (i.e. feedback practice vs. error reduction). This part is about the quantitative data analysis, and the research purpose is to develop a causal explanation by hypothesis testing. The justifications for the selected data collection instruments will be discussed in Chapter 5.2.2, and the process of data collection and analysis will be provided in Chapter 5.3 ('Explanatory Mixed Methods Research Design: Setup & Operation'), Chapter 5.4 ('Validity & Reliability'), Chapter 5.5 ('Data Collection'), Chapter 5.6 ('Error Reduction'), Chapter 5.7 ('Students' Perceptions on Teacher Written Feedback') and Chapter 5.8 ('Data Analysis & Triangulation').

The interpretivist research adopts the qualitative research design where the researcher "relies on the views of participants, asks broad, general questions, collects data consisting largely words or text from participants, describes and analyzes these words for themes, and conducts the inquiry in a subjective, biased manner" (Creswell, 2005, p.39). As such, for each question in my questionnaire, a comment box (i.e. empty box) was given where the participants could choose to freely express their opinions after they had selected the categorical item in the questionnaire. It is believed that this arrangement could help identify some patterns, themes or features in their written comments that can explain or enrich the meanings of the statistical results. This part is about the

qualitative data analysis, and the research purpose is to provide the insiders' viewpoints. The justifications for the selected data collection instrument will be discussed in Chapter 5.2.2, and the process of data collection and analysis will be provided in Chapter 5.3 ('Explanatory Mixed Methods Research Design: Setup & Operation'), Chapter 5.4. ('Validity & Reliability'), Chapter 5.5 ('Data Collection'), Chapter 5.6 ('Error Reduction'), Chapter 5.7 ('Students' Perceptions on Teacher Written Feedback') and Chapter 5.8 ('Data Analysis & Triangulation').

Having said so, the quantitative component remains the backbone and anchor of this research study as the primary interest is the evaluation of actual effectiveness of the electronic feedback on error reduction as compared to that of the paper-based feedback.

5.2.2   Mixed Methods Design: Selection of & Justifications for Data Collection Instruments

According to Creswell (2005), three general types of research designs are classified with correspondence to the degree of intervention of a researcher, namely 'descriptive research' which provides a report of what is actually happening in the classroom for a specific purpose, 'interventionist research' in which some kind of change is intentionally introduced in one aspect of the teaching and learning process and the focus is set on monitoring the effects of this change, and lastly 'experimental research' which involves a formal control of the variables and is usually conducted with the control group of participants without participating in the treatment such that the results of which can be compared and contrasted with those of the experimental group. With the primary objective setting on evaluating the actual effectiveness of teacher written feedback,

the research design of this study falls into a broader category of 'experimental research' under which the explanatory mixed methods research design is one of its branches. The term 'explanatory' will be explained in Chapter 5.3. In the experimental research (or a research design which places a premium on the quantitative component), 'performance measures', 'attitudinal measures', 'behavioral observations' and 'factual information' are the common sources of data collection instruments (Creswell, 2005, p.159). The description of each is provided as follows:

1. Performance Measures

'Performance Measures' assess an individual's ability to perform on an achievement test, intelligence test, aptitude test, interest inventory, or personality assessment inventory. Through past research, researchers might have already developed norms or standardized benchmark for the tests. But one drawback of these measures is that they do not measure individual perceptions or attitudes.

2. Attitudinal Measures

'Attitudinal Measures' measure attitudes of individuals, which is a popular form of quantitative data for surveys, correlational studies and experiments. According to Creswell (2005), "researchers use attitudinal measures when they measure feelings toward educational topics. To develop attitudinal measures, researchers often write their own questions or they find an instrument to use that measures the attitudes" (p.156).

3.   Behavioral Observations

'Behavioral Observations' are used to collect data on specific behaviors through which an instrument is selected to observe individuals for that behavior, and to check points on a scale that reflect the behavior (i.e. behavior checklist). According to Creswell (2005), "the advantage of this form of data is that you can identify an individual's actual behavior, rather than simply record their views or perceptions" (p.156). However, the challenge for this form of data collection lies with its difficulty in giving scores to behaviors and the process of data collection is time consuming (ibid).

4.   Factual Information

'Factual Information' refers to "numeric, individual data available in public records" (Creswell, 2005, p.158). This source of data also refers as secondary data.

As mentioned, the primary objective of this study was to evaluate the actual effectiveness of the electronic feedback on error reduction against that of the paper-and-pen feedback. Having regard to the choice of instruments above, it appears that 'Performance Measures' served the primary objective of the study as such instrument allowed me to assess an individual's ability to make error correction in the second draft based on feedback given on their first draft. Still, through reviewing the past research, an 'error count' approach adopted by Ferris (2006) in her study serves the primary objective of this study very well. In addition, Ferris's 'error count' approach provides a

systemic and objective way of conducting the performance measurement above. Further explanation for the statistical analysis of and justifications of the 'error count' approach will be provided in Chapter 5.3 and Chapter 5.4.2.

To fulfill the secondary objective of this study which was to evaluate students' perception of teacher written feedback as well as to remedy the drawback of 'Performance Measures', it appears that 'Attitudinal Measures' could serve this supplementary objective as such instrument could measure the attitudes of the participants with respect to the teacher written feedback. As such, a questionnaire was a suitable instrument to understand students' perceptions for it could examine the attitudes, opinions, beliefs and characteristics of the participants (Brown, 1998). Having reviewed the literature measuring students' perceptions towards teacher feedback, it was found that the questionnaire designed and adopted by Lee in her study (2008b) was a good fit to this study as the supplementary objective of which was almost identical to that of Lee's (2008b) study. The questions fielded were comprehensive which were completely what this study would like to find out from the participants. Further justifications of this questionnaire are provided in Chapter 5.4.2.

Different types of measurement scales were used in the questionnaire of this study. A measurement scale is a set of rules for quantifying or assigning numerical scores to a particular variable, and there are four types of scale (Bulter, 1985; Creswell, 2005):

1. Nominal Scale (or Categorical Scale) – basically it does not measure but rather names. Observations are simply classified into categories with no necessary relation or order existing between the categories. It also provides response where participants can check one or more categories that "describe their traits, attributes, or characteristics" (Creswell, 2005, p.167). For example, Question 14 of the questionnaire (Appendix D) adopted the nominal scale as below:

	14. Which of the following **types of error** indicated in the feedback is/are **more difficult to understand**? You can tick a **maximum of 2** boxes.	
	Types of Error	**Preference**
a.	Wrong Word / Wrong Word Choice	
b.	Missing / Redundant Word	
c.	Word-level errors (e.g. Tense; Preposition; Singular-Plural; Wrong Verb Form)	
d.	Clausal-level errors (e.g. Wrong Sentence Pattern; Run-on Sentence; Sentence Fragment)	
e.	Awkward Expression (e.g. 'Chinglish'; Unclear / Confusing Meaning due to the Wrong Grammatical Structure used)	
f.	Tone & Style errors (e.g. informal word; informal phrase)	
g.	Organization-related errors (e.g. Paragraphing; Lack of Transitions; Formatting)	

Comments: _____

_____

2. Ordinal Scale – It is a rank ordering of things, with a categorization in terms of more than or less than.

3. Interval Scale – Not only it represents the order of things but also defines the interval or distance between judgements. For example in Question 1 of the questionnaire (Appendix D):

1. Was your language instructor's feedback **legible (understandable)**? (Please circle only **ONE** answer) ('3' = Partially)

5	4	3	2	1
Totally < - -------------------------------------- Partially -------------------------------> Not Legible at all				

Comments: _____

_____

In this study, both nominal scale and interval scale were adopted in the formation of closed-ended questions (i.e. structured response) in the questionnaire (Appendix D), where the nominal scale was used from Question 13 – Question 18, whereas the interval scale was adopted in Question 1, Question 2 and Question 19.

Question 1 and Question 2 asked if the participants understood the teacher feedback and were able to correct the errors identified in the feedback. Questions 3 – 6 asked about the preference of the feedback types (i.e. Grades / Error Feedback / Written Comments / Combination), Question 7 and Question 8 asked about the preference on the focus of error feedback (i.e. Content / Organization / Language), Question 9 asked about the amount of errors to be indicated, Question 10 and Question 11 asked about the preference on the focus of written comments (i.e. Content / Organization / Language), Question 12 examined the participants' preference on error feedback

strategies (Underlining Errors / Circling Errors / Provision of Examples / Provision of Corrections / Combination), Question 13 attempted to understand the degree of explicitness of error feedback preferred by participants, Questions 14 – 17 examined which types of error were perceived by the participants to be more difficult and easier to understand and correct. Question 18 solicited which kinds of post-writing activities were preferred by the participants, and lastly Question 19 asked the participants to what extent they regarded teacher written feedback as a whole was conducive to their long-term writing.

Other than fielding the closed-ended questions, an open-ended question in the form of a 'comment box' was given at the end of each closed-ended question, which allowed the participants to give their opinions in whatever form. The open-ended questions are the only qualitative part of the study. While maintaining the quantitative components ('Performance Measures' and 'Attitudinal Measures') as the backbone for the primary objectives they served, the collection of qualitative data was used to help probe further into some key quantitative results or patterns identified by coding key recurrent information into themes. Since a premium was placed on the quantitative data and analysis, the mixed methods research design employed here is called 'explanatory mixed methods design' which will be defined in Chapter 5.3. It is also for this reason other qualitative instruments such as interviews, observations and the like were excluded from this study.

The next section will define what the 'explanatory mixed methods design' is, and provide a detailed account of how the quantitative data and qualitative data were collected, analyzed and synthesized in a complementary and triangulated manner.

5.3　Explanatory Mixed Methods Research Design: Setup & Operation

The purpose of this research is to investigate the actual effects of teacher written feedback on the writing revision process with two different forms of feedback treatment, namely, the electronic feedback (i.e. electronic feedback) and the traditional paper-and-pen feedback (i.e. paper-based feedback), in improving the overall accuracy of student writing. An explanatory design of the mixed methods research was adopted in this experimental study.

A mixed methods research design is "a procedure for collecting, analyzing and "mixing" both quantitative and qualitative data in a single study to understand a research problem" (Creswell, 2005, p.510). The reasons for adopting the mixed methods approach was in the belief that it would provide a better understanding of my research problems than either type by itself, and that this would build on the strengths of both quantitative and qualitative data in the providing us a fuller and more meaningful picture of the research findings.

In this study, a quantitative component (i.e. Error Count) was followed by another mix of quantitative and qualitative component (i.e. Questionnaire). This research design is called the "explanatory mixed methods design". According to Creswell (2005):

An explanatory mixed methods design consists of first collecting quantitative data and then collecting qualitative data to help explain or elaborate on the quantitative results. The rationale for

this approach is that the quantitative data and results provide a general picture of the research problem; more analysis, specifically through qualitative data collection, is needed to refine, extend, or explain the general picture…the mixed methods researcher places a priority on quantitative data collection and analysis. This is done by introducing its first in the study and having it represent a major aspect of data collection. A small qualitative component follows in the second phase of the research. (p.515).

The quantitative component contributed to the major part of this study as it can "collect data from a large number of people" (Creswell, 2005, p.510). As shown in the table on the page after next, the 'error count' method was a primary and major quantitative component modeled on Ferris's (2006) study. The reason for adopting the 'error count' method was that it served the primary objective of examining the actual effects of a particular form of feedback treatment in error reduction, by error category, in student writing revisions of a group of students, addressing research questions (1) and (2). On the other hand, the 'Questionnaire' method which comprised both the quantitative and qualitative data (see Appendix D) was a secondary and supplementary component modeled on Lee's (2008b) study. The reason for adopting the 'questionnaire' method was that it served the secondary objective of evaluating students' perceptions of teacher writing feedback by means of fielding both close-ended nominal and interval questions and open-ended 'comment box', addressing research question (3). The advantage of these close-ended questions was to collect quantitative data to support theories (e.g. learner autonomy and Krashen's Input Hypothesis) and hypothetical assumptions (e.g. consistency with the 'Error Count' results in phase one) in this study, whereas the advantage of the open-ended responses allowed me to explore reasons for the close-ended responses and identify comments people might have which were out

of the scope of the responses to close-ended questions. As illustrated in Creswell's quotation above, the questionnaire results representing the second phase of the research was used to help elaborate and explain the quantitative results obtained in the first phase. And it was done by the means of triangulation (cross-comparison of data from different sources) for cross-checking of consistency, and identification plus quantification of overlapping qualitative themes (i.e. overlapping themes / key words in qualitative comments) for understanding the perspectives and viewpoints of the respondents.

The quantitative component, comprising mainly the 'Error Count' method (modeled on Ferris, 2006) supplemented by 'close-ended responses of the questionnaire' (modeled on Lee, 2008b) can provide a greater confidence in the generalizability of the results which is the primary objective of this study, especially this was the first time evaluating the actual effectiveness of 'Mark My Words' ('MMWs') to a certain population of students. The qualitative component, comprising the open-ended questions in the form of a 'comment box' in the questionnaire can help clarify quantitative statistical relationships and numeric findings, and obtain certain overlapping perspectives and viewpoints of the respondents. To put this into perspective, the explanatory mixed methods design on the one hand allows quantitative and qualitative data, alongside the results obtained from some previous studies with similar research objectives, being triangulated in a single study. Such triangulation allows me to compare data from various sources, and make interpretation if these results support or contradict each other (Creswell, 2005). For example, it was found from the 'error count' that the experimental group receiving electronic feedback made statistically significant reduction across all error categories when compared to the control group receiving the paper-based feedback (see Chapter 6.2). Simultaneously, it was also reported from the quantitative

results of the questionnaire that a higher percentage of students from the experimental group than

the control group indicated that they were able to comprehend and correct errors (see Chapter 7.1

& 7.2). As such, both quantitative results from two sources confirmed each other about the

effectiveness of the electronic feedback over the paper-based feedback. On the other hand, this

explanatory mixed methods design allows the qualitative data being used to "refine the quantitative

data such that this refinement can lead to exploring a few typical cases, probing a key result in

more detail, or following up with outliers or extreme cases" (Creswell, 2005, p.515). As such,

certain overlapping qualitative themes or key words were identified in the 'comment' boxes of the

questionnaire for trend, and they were quantified by frequency of occurrence for magnitude. These

high-frequency qualitative themes or key words were a reflection of certain common perspectives

and viewpoints held by the respondents with respect to certain questions being examined. And this

information can provide us some explanation of the quantitative results, so does a fuller picture of

the situation. For example, when scrutinizing the qualitative results of the questionnaires with

respect to the comprehensibility of errors and ability of error correction (see Q1 of Chapter 7.1 &

Q2 of 7.2), a prominent qualitative theme or key words called 'Google link / search / example'

was identified most frequently and overwhelmingly in these qualitative comments when the

respondents from the experimental group receiving electronic feedback attributed to their

satisfaction over such comprehensibility and ability. Deriving from the above, it was logical and

reasonable to arrive at some observation that their statistically significant error reduction was

associated with the distinguished feature of 'Google Link' (i.e. a recommended lexcio-

grammatical form in authentic texts) provided by 'Mark My Words' ('MMWs'), which was not

offered to the control group.

Figure 5.1 below illustrates the methodologies adopted in the explanatory mixed methods design which was divided into two phrases where priority was placed on the 'Error Count' method.

Phase	Explanatory Mixed Methods Design	Objectives	Instruments	Results
1.	Error Reduction [modeled on Ferris's study (2006)]	[Primary Approach] Examining the actual effects of the electronic feedback and paper-based feedback in reducing the frequency of errors in student writing revisions	Error Count (Quantitative – primary instrument)	Chapter 6: Findings & Discussion: Error Reduction
2.	Students' Perceptions [modeled on Lee's study (2008b)]	[Supplementary Approach] Evaluating the students' perceptions of the electronic feedback and paper-based feedback	Questionnaire (Quantitative by close-ended questions & Qualitative by open-ended questions – supplementary instrument)	Chapter 7: Findings & Discussion: Students' Perceptions

Figure 5.1

The first phase is the primary and major method which involved the 'Error Count' approach modeled on Ferris's (2006) study. In this study, the frequency of identified errors by each error category was counted before and after a particular form of feedback treatment respectively for the experimental group receiving the electronic feedback, and for the control group receiving the paper-based feedback. The mean of error frequency was compared within and across the treatment

groups before and after the intervention (i.e. either the electronic feedback or paper-based feedback). A number of t-tests were run to determine if there was a statistically significant difference for each of the error categories (7 error categories in total) between the experimental group and control group, respectively in the 'between-groups' comparison and the 'within-group' comparisons. As such, the actual effects of the electronic feedback and paper-based feedback in reducing the frequency of errors in student writing revisions were statistically determined, addressing research questions (1) and (2). The details of the statistical analysis process will be provided in the following sub-section (Chapter 5.3.1 "Statistical Analysis: Two Sample T-Tests").

To ensure a higher level of confidence in and provide more detailed explanation for the statistical results obtained in 'Error Count', the 'Questionnaire' method representing the second phase of the explanatory mixed methods research, was used to help cross check, elaborate and explain the quantitative results obtained in the first phase. It was done by the means of triangulation (i.e. cross-comparison of the quantitative data respectively collected from 'Error Count' and 'Questionnaire') for cross-checking of consistency, and by identification plus quantification of overlapping qualitative themes (i.e. high-frequency themes / key words in qualitative comments) for understanding some common perspectives and viewpoints of the respondents. The questionnaire was administered to 32 students in the experimental group add 30 students in the control group with the objective of collecting their views after they had received the feedback treatment. This second method served the secondary objective of evaluating the students' perceptions of the electronic feedback and paper-based feedback by fielding 12 close-ended questions followed by an open-ended question after each (see Appendix D). The results and explanation of which are

presented in Chapter 7 'Findings & Discussions: Students' Perceptions on Teacher Written Feedback'.

5.3.1 *Statistical Analysis: Two Sample t-tests for Error Count*

To operationalize the data analysis for the 'error count' approach, the most direct and explicit way of such evaluation was to count the number of errors by each error category in the first draft and the second draft respectively for the experiment group ("E") and the control group ("C") while remaining other variables constant. The data obtained from error count from all students for the experimental group and control group were compiled into the Microsoft excel file appended below (see *Figure 5.2*) where:

(1) Student Names & IDs were entered into the left hand columns labelled with S_Name and S_ID.

(2) These students were divided into the experimental group (32 students) and the control group (30 students) which were represented by a letter "E" and "C" respectively.

(3) The pre-test and post-test total scores, the total number of errors, the number of errors for each of the 7 error categories were represented by the following labels in the excel spreadsheets: PRETEST (= pretest score), POSTTEST (= post test score), tot (= total errors), AWK (awkwardness errors), cl_err (= clausal errors), colloc (= collocation errors), sp (= word errors), tone (= tone & style errors), cont (= content errors), org. (= organization errors).

(4) One excel spreadsheet was catered to each error category as it can be seen in the figure below the excel file was subdivided into various index tags at the bottom.

(5) Each error category was also divided into draft 1and draft 2. Taking 'Awkwardness' errors as an example, there were AWK_D1 which refers to the number of 'Awkwardness' errors identified in draft 1, and AWK_D2 which refers to the number of 'Awkwardness' errors identified in draft 2. If student "A" from the experiment group made 1 count of awkwardness errors in draft 1 but none in draft 2, then in the corresponding row under the student name, a numeric value of "1" was put in AWK_D1 column and a numeric value of "0" (or blank space) was put in AWK_D2 column. The same was done for all students across all error categories, including the pre-test results, post-test results and the total errors.

Figure 5.2

After inputting all the data in the corresponding spreadsheets of the excel file, the next step was to calculate the mean of the pre-test total scores & post-test total scores respectively for the experimental group ("E") and the control group ("C"), the mean of the total number of errors in draft 1 and draft 2 for "E" and "C" , and the mean of the number of errors for each of the 7 error categories in draft 1 and draft 2 for "E" and "C". They are all the dependent variables of this study, whereas the independent variables are the two forms of feedback treatment – electronic feedback vs. paper-and-pen feedback. Then, I needed to find out (see *Figure 5.3*):

149

Between-groups comparison	(1) if there was a statistically significant difference between the means of the number of errors respectively made by the electronic feedback group ("E") and the paper-and-pen feedback group ("C") in <u>draft 1</u>, for each of the dependent variables (e.g. awkwardness errors)
	(2) if there was a statistically significant difference between the means of the number of errors respectively made by the electronic feedback group ("E") and the paper-and-pen feedback group ("C") in <u>draft 2</u>, for each of the dependent variables (e.g. awkwardness errors)
Within-groups comparison	(3) if there was a statistically significant difference between the means of the number of errors respectively in draft 1 and draft 2 (e.g. awkwardness errors) <u>made by the experimental group</u> ("E")
	(4) if there was a statistically significant difference between the means of the number of errors respectively in draft 1 and draft 2 (e.g. awkwardness errors) <u>made by the control group</u> ("C")

Figure 5.3

The quantitative data was analyzed using R program. The original excel file (see *Figure 5.2*) was converted into csv. file with the student names deleted, and then loaded into R for statistical computation.

The two-sample t-test was adopted for the statistical analysis as it is one of the most commonly used hypothesis tests (Cohen et al., 2011; Creswell, 2005). It is applied to compare whether the mean difference between two treatment groups is really significant or if it is due instead to random chance. It helps to answer questions like whether the reduction in average number of errors (i.e. mean of the number of errors) in a certain error category is statistically significant after implementing a new form of feedback treatment.

Specifically, the four steps involved to conduct any hypothesis test are:

1. Specify the null alternative hypotheses, for example:

> (1) Null hypothesis (H0): electronic feedback will not improve students' language accuracy (*overall or for certain error category*) in their writing revisions.
>
> (2) Alternative hypothesis (H1): electronic feedback will improve students' language accuracy (*overall or for certain error category*) in their writing revisions.

Using the sample data and assuming the null hypothesis is true, calculate the value of the test statistic. Again, to conduct the hypothesis test for the population mean μ, the t-statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

was used which followed a t-distribution with $n - 1$ degrees of freedom. The statistical computation was to be processed by R program.

2. Using the known distribution of the test statistic, calculate the P-value: "If the null hypothesis is true, what is the probability that we'd observe a more extreme test statistic in the direction of the alternative hypothesis than we did?" (Note how this question is equivalent to the question answered in criminal trials: "If the defendant is innocent, what is the chance that we'd observe such extreme criminal evidence?"). The statistical computation (e.g. calculations of t-statistics, p-values, means) involving Step 2 and Step 3 was to be processed by R program, by inputting certain commands which could be read and executed by the program. For each dependent variable (except pre-test and post-test total scores), a two sample t-test was to be run fourth times in the following manners:

Example: Two Sample t-tests for 'Awkwardness' Errors	
Between-groups comparison (e.g. Awkwardness errors in draft 1 between "E" & "C")	For example, a two sample t-test was conducted to examine if there was a statistically significant difference between the means of the number of errors respectively made by the electronic feedback group ("E") and the paper-and-pen feedback group ("C") in draft 1, for the awkwardness error category. The following commands needed to be entered into R: t.test (bma$AWK_D1[bma$Group == "E"] and bma$AWK_D1[bma$Group == "C"]
Between-groups comparison (e.g. Awkwardness errors in draft 2 between "E" & 'C")	For example, a two sample t-test was conducted to examine if there was a statistically significant difference between the means of the number of errors respectively made by the electronic feedback group ("E") and the paper-and-pen feedback group ("C") in draft 2, for the awkwardness error category. The following commands needed to be entered into R: t.test (bma$AWK_D2[bma$Group == "E"] and bma$AWK_D2[bma$Group == "C"]
Within-groups comparison (e.g. Awkwardness errors between draft 1 & draft 2 in "C"	For example, a two sample t-test was conducted to examine if there was a statistically significant difference between the means of the number of errors respectively in draft 1 and draft 2 (e.g. awkwardness errors) made by the control group ("C") t.test (bma$AWK_D1[bma$Group == "C"] and bma$AWK_D2[bma$Group == "C"]
Within-groups comparison (e.g. Awkwardness errors between draft 1 & draft 2 in "E"	For example, a two sample t-test was conducted to examine if there was a statistically significant difference between the means of the number of errors respectively in draft 1 and draft 2 (e.g. awkwardness errors) made by the experimental group ("E") t.test (bma$AWK_D1[bma$Group == "E"] and bma$AWK_D2[bma$Group == "E"]
Remark: 'bma' is the file name of the data spreadsheets to be located and executed by R program.	

Despite the fact that conducting the between-groups comparison for draft 2 (in the absence of draft 1) would have already given me some confidence if a particular form of feedback

treatment performed better in remedying certain errors, it was felt that the same should also be done for draft 1 in the between-groups comparison to examine if there was no statistically significance in the two groups' abilities for avoiding errors before either form of feedback treatment was conducted. This was to ensure that the language ability of two groups was on par with each other before the participants received the feedback treatment.

In addition, despite the possibility that some statistical results might show that a particular form of feedback treatment was statistically more significant than the other in remedying certain error categories in the final draft (i.e. draft 2), this should not exclude any possibility that the less-performed feedback treatment still could make statistically significant reduction in the same error categories within its treatment group (in the within-group comparison), but it just performed relatively less effectively than its counterpart (in the between-groups comparisons). In order not to exclude such possibility, the within-groups comparison was conducted between draft 1 and draft 2, for the total error category and each of the seven error categories.

3. Set the significance level, α, the probability of making a Type I error to be small — 0.01, 0.05, or 0.10. Compare the *P*-value to α. If the *P*-value is less than (or equal to) α, reject the null hypothesis in favour of the alternative hypothesis. If the *P*-value is greater than α, do not reject the null hypothesis. The significance level was decided by taking p-values into the consideration where $p > 0.05$ meant there was not a meaningful difference, and $p < 0.05$ meant there was a statistically significant difference. According to Creswell (2005), "A p-value is the probability (p) that a result could have been produced by chance if the null hypothesis is true" (p.188). "A significance level (or alpha level) is a probability level that reflects the maximum

risk you are willing to take that any observed differences are due to chance" (ibid). If the significance level is set as 0.05, this means an extremely low probability value will actually be observed if the null hypothesis is true (i.e. 5 out of 100 times it will be due to random chance).

As such, the interpretation of the p-value against the significance level (which was set at 0.05 level), for the total errors and each of the 7 error categories, was made as follows:

Example: Two Sample t-tests for Certain Error Category	
Null hypothesis (H0)	Electronic feedback will not improve students' language accuracy (*for certain error category*) in their writing revisions.
Alternative hypothesis (H1)	Electronic feedback will improve students' language accuracy (*for certain error category*) in their writing revisions.

As it has been noted, a p-value is a probability. This means that it is a real number from 0 and 1. While a test statistic is one way to measure how extreme a statistic is for a particular sample, p-values are another way of measuring this. P-value if the value determining if a result is statistically different or not. It is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

When a p-value and other statistical data (e.g. t-statistic, means, etc.) were generated by R program, the question is, "Is the sample obtained which shows a difference the way it is by chance alone with a true null hypothesis, or is the null hypothesis false?" There are two possible outcomes for using p-values in such hypothesis testing. The interpretation of such is demonstrated by using an example showing the "between-group comparison of a certain error category in draft 2", which is as follows:

Outcome	Between-groups comparison of a certain error category in draft 2
P-value is less than 0.05	It means if the H0 holds (there is no difference between the means of the number of errors for a certain error category respectively made by the electronic feedback and paper-and-pen feedback group in draft 2), the chance of getting a sample that shows a difference is less than 5%. The chance of getting such a sample is pretty small like winning a lottery; so getting such a sample which shows a difference in reality is not likely by random chance, and this would be enough evidence to reject H0 and accept H1.
P-value is more than 0.05	It means if the H0 holds (there is no difference between the means of the number of errors for a certain error category respectively made by the electronic feedback and paper-and-pen feedback group in draft 2), the chance of getting a sample that shows a difference, by random chance, is more than 5%. The chance of getting such a sample is not small; thus getting such a sample which shows a difference in reality is likely by random chance. This means it is not likely to find a difference when a sample is taken, and this would not give me enough evidence to reject H0.

In general, the smaller the p-value, the more evidence that we have against the null hypothesis. The above interpretation of the p-value against the 0.05 significance level was applied to the subsequent statistical analysis for the total pre-test scores, the total post-test scores, the total errors and each of the 7 error categories. The results and explanation of which are presented in Chapter 6 'Findings & Discussions: Error Reduction'.

Despite a statistical result showing that there is statistically significant between the variables, it does not tell about the strength of such relationship, be it strong or weak. As such, effect size was also calculated. According to Creswell (2005), "Effect size is a means for identifying the practical strengths of the conclusions about group differences or about the relationship among

155

variables in a quantitative study" (p. 186). There are different types of effect size (Huberty, 2002). In this study, the standardized difference between the means of two groups was used for determining the effect size results. According to Baguley (2009), the standardized mean difference includes two scales: Cohen's d and Hedge's g. In this study, Cohen's d was used because it is commonly used in social sciences. The effect size in standardized mean difference indicates the magnitude of treatment effect as follows (Cohen, 1988, p. 25):

- Effect size >= 0.2   Small effect
- Effect size >= 0.5   Medium effect
- Effect size >= 0.8   Large effect

The above mechanism was applied to the interpretation of p-values concerning the questions stated in *Figure 5.4* and *Figure 5.5*. The results and explanation of which are detailed in Chapter 6 'Findings & Discussions: Error Reduction'. This is a way of quantifying the size of the difference between two groups. It is valuable for quantifying the effectiveness of a particular form of feedback treatment (i.e. electronic feedback for the experimental group), relative to some comparison (i.e. paper-based feedback for the control group), for the effect size suggests "how well does electronic feedback work" rather than merely "does electronic feedback work or not" in hypothesis testing (i.e. p-value).

Figure 5.4: Between-Groups Comparison	
*Whether there is a difference between the Electronic feedback and Paper-based feedback in their effectiveness in reducing errors in draft 1 (D1).	
**Whether there is a difference between the Electronic feedback and Paper-based feedback in their effectiveness in reducing errors in draft 2 (D2).	
Between Groups	**Electronic (D1) vs. Paper (D1)** **Electronic (D2) vs. Paper (D2)**
Pre-test vs .Post-test **(Ch. 6.1)**	Is there a statistical difference between total scores respectively made by the electronic feedback group and the paper-based group, respectively in pre-test and post-test?
Total Errors **(Ch. 6.2)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?
'Awkwardness' Errors **(Ch. 6.3)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?
'Clausal-level' Errors **(Ch. 6.4)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?
'Word-level' Errors **(Ch. 6.5)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?
'Collocation' Errors **(Ch. 6.6)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?
'Tone & Style' Errors **(Ch. 6.7)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?
'Content' Errors **(Ch. 6.8)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?
'Organization' Errors **(Ch. 6.9)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?

Figure 5.5: Within-Groups Comparison		
#Whether the Paper-based feedback makes a significant difference in reducing errors in draft 2 (D2). ##Whether the Electronic feedback makes a significant difference in reducing errors in draft 2 (D2).		
Within Groups	**Paper (D1) vs. Paper (D2)** **Electronic (D1) vs. Electronic (D2)**	
Pre-test vs. Post-test **(Ch. 6.1)**	Is there a statistical difference between the mean scores respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in pre-test and post-test?	
Total Errors **(Ch. 6.2)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	
'Awkwardness' Errors **(Ch. 6.3)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	
'Clausal-level' Errors **(Ch. 6.4)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	
'Word-level' Errors **(Ch. 6.5)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	
'Collocation' Errors **(Ch. 6.6)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	
'Tone & Style' Errors **(Ch. 6.7)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	
'Content' Errors **(Ch. 6.8)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	
'Organization' Errors **(Ch.6.9)**	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	

5.4 Validity & Reliability of the Research Methodology

5.4.1 Measures for Minimizing Methodological Flaws

A number of methodological flaws were identified in my critical review of the 'teacher written feedback' in Chapter 3.2, which is mostly represented by Ferris's works in 2006, and 'computer-mediated feedback' studies in Chapter 3.5. These methodological flaws might have possibly given rise to some extraneous variables which were not in the interest of this study and would certainly undermine the reliability and validity of this study without addressing them properly in the research design.

The two major issues identified in Ferris's study (2006) are, firstly, Ferris's (2006) inadequate comparison of different error types against different forms of feedback treatment in the absence of a controlled condition in which teachers were allowed in her study to arbitrarily use direct or indirect feedback to respond to all error types based on their intuitive choices. It was found that 'untreatable errors' (i.e. errors that cannot be identified with and explained by grammatical rules) was overwhelmingly marked by direct feedback while 'treatable errors' (i.e. errors that can be identified with and explained by grammatical rules) tended to be marked by indirect feedback. Such unconstrained condition might have possibly rendered the evaluation on the effectiveness of a particular form of teacher feedback (i.e. indirect feedback) under examined inaccurate and impartial because teachers in Ferris's study (2006) inclined to adopt a quick-fix approach to remedy 'untreatable errors', and this practice has made no difference from the provision of an

overt correction while indirect feedback was being used on 'treatable errors' which is relatively easier for students to understand and correct. To remedy the above methodological flaw, a compulsory marking principles were imposed in this study (see Chapter 5.5.2 (f)).

Secondly, another flaw is Ferris's study (2006) is her inadequate comparison of different error types against different forms of feedback treatment without consideration for the sequence of second language acquisition in her research design. This implies that students making different error types respond in varying degrees to the same form of feedback treatment, thus by putting the responsiveness of 'treatable errors' (i.e. 'verb' error in her study) on par with the responsiveness of 'untreatable errors' (i.e. 'sentence-level' errors in her study) against either the direct/indirect feedback under the free choice of the teachers and then using the results of error reduction to evaluate if a particular form of feedback would have any positive effects on students' writing revisions in the short and long term might seem ill-conceived and impartial as well. To remedy two methodological flaws, unlike Ferris's study (2006) which compared the responsiveness of 'different' error types against two different forms of feedback without maintaining the error type constant each time, this study only compared the responsiveness of the 'same' specific error type respectively against the electronic feedback and paper based feedback each time. This can ensure that any significant reduction in each of the error types identified with either the experimental group receiving the electronic feedback or the control group receiving the paper-based feedback could be attributed to the adoption of a particular form of feedback treatment, rather than other extraneous factor exhibited in Ferris's study (2006).

On the other hand, as already discussed in Chapter 3.5.2 regarding the critical review of 'computer-generated feedback' studies, it appears most of the studies (e.g. Chen, 1997; Warden, 2000; Hyland & Hyland, 2006; Ware & Warschauer, 2006; Chen & Cheng, 2008; Lee et al., 2009; Schroeder et al., 2008; Grimes, 2008; Chodorow et al, 2010; Wang & Wang, 2012; Stevenson & Phakiti, 2013) were not rigorously designed such that the methodological issues arisen might have possibly rendered the effectiveness of the computer-generated feedback on student writing revisions inconclusive. To summarize, ten main methodological issues identified in the previous discussion of 'computer-generated feedback' studies which might also have possibly undermined their validity and reliability are (1) non empirical-based research, (2) inadequate comparison between groups receiving feedback with groups receiving no feedback, (3) absence of control groups, (4) inadequate comparison between students of different instructional settings on the effectiveness of a particular feedback treatment, (5) inadequate comparison between students of different proficiency levels on the effectiveness of a particular feedback treatment, (6) inadequate comparison of feedback treatments with different focus (e.g. content feedback vs. grammar feedback), (7) sketchy description of participants and their institutional or instructional contexts alongside their possible implications on the research outcomes, (8) avoidance of extraneous factors which might have an impact on the validity and reliability of the studies, and (9) ignorance of negative evidence, (10) inadequate sample size, and (11) inter-rater reliability.

With the above issues in mind, a number of measures were also implemented to minimize the methodological flaws and extraneous factors such that any statistical difference identified in students' writing performance could be more likely to attribute to the different forms of feedback adopted:

161

(1) An empirical-based research was conducted in this study. The mixed methods approach was adopted, namely the 'error count' method and 'questionnaire' were administered to both experimental and control groups. The 'error count' method was modeled on Ferris's study (2006) which served to examine the actual effects of the electronic feedback and paper-based feedback in reducing error frequency in student writing revisions. According to Stevenson & Phakiti (2014), the capability of a particular form of feedback treatment in improving the quality of students' writing revisions is central to claims made about the effectiveness of that feedback treatment, and 'error count' is one of the most common measurements for such evaluation in Automated Writing Evaluation ('AWE') studies. After the intervention, a questionnaire was administered to both groups. The questionnaire, which was modeled on Lee's study (2008b), served to evaluate students' perceptions of the electronic feedback and paper-based feedback. It was anticipated that the qualitative comments could help make some insightful meanings to the quantitative results. Through conducting the error count which was the primary focus of the methodology and then supplemented by the explanation obtained from the qualitative comments, it was anticipated that the results obtained could answer the research questions (Please refer to Chapter 4.1 for Research Problems and Chapter 4.2 for Research Questions).

(2) Both experimental group and control group received teacher feedback despite in different forms.

(3) The control group receiving the conventional paper-and-pen feedback was in place for comparison with the target groups receiving the electronic feedback.

(4) Students from both experimental group and control group were exposed to an identical instructional setting as all students were all enrolled in the identical second-year language

course for engineering students, going through the same syllabus, teaching and learning materials, and assessment requirements.

(5) Both participants in the experimental group and control group were local students of comparable English proficiency level which was evident by their streaming results based on their grades received in the first-year English language course and a pre-test before the experimental intervention.

(6) This study examined the effectiveness of teacher feedback in both treatment groups on 7 error categories.

(7) A clear description of participants and the institutional and instructional contexts was provided (see Chapter 4.4 and Chapter 5.5)

(8) Efforts were evident in minimizing and avoiding extraneous variables in this study. Two teacher markers, who were experienced markers and teachers of the language course where the experimental setting took place, were selected based on their reliability and consistency of the marking standard in the language course over the last five years, which was evident by their course means and standard deviations. Certain marking principles were imposed to ensure that the focus, comprehensiveness (i.e. detailed marking but not selective marking), and level of explicitness (i.e. indirect coded/uncoded feedback only) were applied to both electronic feedback for the experimental group and the paper-based feedback for the control group. Failure to adhere to the above principles would make the subsequent data analysis difficult.

(9) All positive and negative findings were presented, and efforts were made to explore their causes, interpreted the results and draw the implications (see Chapters 6-8).

(10)　To make efforts for this study to fulfill an adequate sample size, 30 students were selected for the experimental group and 32 were selected for the control group. The total sample size is

hence 62, which is considered to be within an acceptable range in 'AWE' research studies (Please refer to *Figure 5.6* below).

(11)   To ensure the inter-rater reliability in this study, an additional procedure was implemented as in Ferris's study (2006) such that both teacher-markers (i.e. Teacher A and Teacher B) of this study marked independently those five scripts of paper (the same scripts used in the course standardization meeting) comprehensively (which was not mandatory by other teacher-markers who did not participate in this study). Under the moderation of the course coordinator who was as well the researcher, they all came together in another separate occasion before the course standardization meeting to discuss the writing issues of those scripts, achieving an agreement on error totals for each error type. This additional procedure further enhances the validity and reliability of the study.

Sample Size	# of 'AWE' studies
Less than 10 participants	1
11 – 50 participants	4
51 – 100 participants	9
101 – 200 participants	5
More than 200 participants	10
Unspecified	1

Figure 5.6 Number of 'AWE' studies and Sample Size (adapted from Stevenson & Phakiti, 2014)

As seen above, efforts were made to avoid the impact of extraneous factors on the validity and reliability of the research outcomes. It was in this way any positive or negative claims drawn from the findings would become more meaningful in the sense that any language gains evident in revisions could be attributable to the use of computer-based feedback.

5.4.2   Justifications for Adoption of Ferris's (2006) & Lee's (2008b) Methodologies

The justifications for adopting both Ferris's (2006) methodology for evaluating the 'Error Reduction' as a primary approach, and Lee's (2008b) methodology for measuring 'Students' Perceptions as a supplementary approach are based on the following:

First, despite the fact that the focus of Ferris's study (2006), Lee' studies (2008b) and this study might be slightly different, they share a lot of similarities, or perhaps they are even in common, with respect to their directions alongside their objectives of the research studies. That is we all investigated the impact of teacher written feedback on student writing revisions. Ferris's study (2006) examined the impact of teacher written feedback on student writing revisions of 16 error types (only covers 'language') with various feedback strategies (e.g. indirect feedback vs. direct feedback, indirect coded feedback vs. indirect non-coded feedback). Lee's studies (2008a & 2008b) examined the impact of teacher written feedback on student revising process with respect to the various contextual factors like the banding of schools, teacher's belief, school requirements and etc. This study examined the impact of teacher written feedback on student writing revisions of 7 error categories (covering 77 error types) with two forms of feedback treatment (i.e. electronic feedback vs. paper-and-pen feedback).

Second, Ferris's (2006) model (i.e. 'Error Count') adopted in her study was a good fit to the primary objective of this study in which the objective of Ferris's (2006) study and the primary objective of this study were both identical. That is they both examined the actual effects of a

particular form of feedback treatment in reducing the frequency of errors on student writing revisions. Moreover, Ferris's (2006) model provides a systematic and objective way of conducting the error count, which can easily be applied in other research environments with similar research purposes. This study is the case in point. In addition, Dana, R., Ferris is one of the most prominent scholars and experts in the area of 'Writing Feedback' under the academic discipline of 'Second Language Writing'. Her research has focused extensively on responses to student writing and on written corrective feedback in second language writing. The details of the Ferris's model and the way it was implemented in this study will be explained in Chapter 5.6 'Error Reduction'.

Third, Lee's (2008b) model adopted in her study was a good fit to the secondary objective of this study in which the objective of Lee's (2008b) study and the secondary objective of this study were both identical. This is to evaluate the students' perceptions of particular teacher written feedback. Moreover, Lee's (2008b) model provides a systematic way of evaluating students' perceptions, which can easily be applied in other research environments with similar research purposes. This study is the case in point. Lee's (2008b) questionnaire might possibly be one of the most comprehensive questionnaires in the measurement of students' perceptions towards teacher written feedback. It is believed that scope and relevance of questions (See Appendix D) being asked in the questionnaire can effectively probe into the students' attitudes towards particular feedback treatments in various dimensions. Given its comprehensiveness, the results of which were able to provide some good explanation to the quantitative data obtained in the 'Error Count'' results. In addition, Lee, Icy is probably the most prominent Hong Kong scholar and expert in the area of 'Writing Feedback' under the academic discipline of 'Second Language Writing'. She has an extensive research track-record in the area of second language writing and teacher written feedback

for Hong Kong ESL students. As such, the results and the methodologies of her studies would probably lend considerable weight to this study on the premise that our studies shared quite similar subjects which are all Hong Kong ESL students despite different institutions and proficiency levels of students, as well as quite identical local educational context. Lee's (2008b) model which was adopted in this study will be explained in Chapter 5.4 'Students' Perceptions on Teacher Written Feedback''.

Fourth, other than addressing the methodological flaws identified in computer-generated feedback studies which were presented in Chapter 5.4.1 'Measures for Minimizing Methodological Flaws', reasonable efforts were also made to minimize extraneous factors which might have undermined the results identified in Ferris's (2006) study and Lee's (2008a, 2008b) studies. This will be explained in Chapter 5.5 'Data Collection' where it can be seen that pre-emptive measures were done on the selection of participants and teacher-markers, marking policy, assessment guidelines and standardization of errors to minimize extraneous factors due to any individual and institutional differences, as mentioned in Lee's (2008b) study. This will be discussed in the 'Findings & Discussion' sections.

Fifth, this study endeavoured to address the following research questions as earlier stated in Chapter 4.2 'Research Questions', in which the first and second research questions remained the primary focus of the study, whereas the third one was the secondary:

(1) Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a difference on student writing revisions, when comparing to the paper-based feedback?

(2) Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a difference on student writing revisions for each error category, when comparing to the paper-based feedback?

(3) What were the students' perceptions of the effectiveness of teacher written feedback?

Having considered the justifications being put forward above, it is believed the 'Error Count' approach postulated by Ferris (2006) can address the research questions (1) and (2) as the 'Error Count' approach was considered in this study to be the most practical and suitable way of evaluating the language gain/loss based on the actual frequency of errors made by students before and after the intervention. As for the 'questionnaire' modeled on Lee's (2008b) study, it is also believed that the quantity and quality of the questions being fielded in her questionnaire (See Appendix D) was considered to be practical and suitable, in terms of its scope and relevance to the writing feedback environment experienced by the participants, in soliciting views on students' perceptions of the effectiveness of teacher written feedback, which can answer research question (3).

5.5   Data Collection

5.5.1   Student Data

To facilitate access to data, convenience sampling was used in the study. Creswell (2005) refers convenience sampling as a researcher selecting participants who are accessible by and available to the researcher. The advantage is that "the research simply chooses the sample from those to whom she has easy access" (Cohen et al., 2011, p. 156), but "the researcher cannot say with confidence that the individuals are representative of the population" (Creswell, 2005, p. 149). A total of 62 second year engineering students in a local university participated in the study. They all enrolled in the "Year Two English Language for Engineering Students" course in which I was the course coordinator (i.e. Course code: LANG206). They were divided into two groups with 30 of them in the experimental group receiving the electronic feedback and the other 32 students in control group receiving the paper-based feedback. They wrote on a topic about proposing an innovative technology.

All participants in the experimental and control groups were local students whose mother tongue was Cantonese and were educated through the local curriculum of government-subsidized secondary schools. All participants were deemed to be of the comparable level of English proficiency. This is evident by (1) their Hong Kong secondary school A-level 'Use of English' results in which all of them either received a 'D' grade or 'E' grade (i.e. a pass); (2) the fact that the university streamed students into classes of the same proficiency level according to their A-

level 'Use of English' results and their language results obtained in their first academic year (i.e. an average score of 4.5 on a 7-point scale in their language proficiency), and (3) the pre-test results before the intervention of this experimental study which suggested that there was no significant difference between the experimental group and the control group in their language proficiency (Table 6.1 and Table 6.2 in Appendix E).

With the primary purpose of understanding the participants' attitudes towards the written feedback they received in their first year's language experience (i.e. LANG106 Fall Semester Course and Spring Semester Course), a survey which was modeled on Lee's study (2008b) with some modification was distributed to the participants before they took part in this study. The survey questions with each of the corresponding results are summarized in Appendix F.

In summary, based on their past experience in the first year writing classroom, the survey results suggested that 74% of all participants from the experimental and control groups felt that teacher written feedback as a whole is helpful to their long-term writing development. In other words, the majority of them were quite positive with it. However, as far as its effectiveness is concerned, only about 65% of all participants from both groups expressed that they understood their teacher written feedback and were able to correct the errors as indicated by their language instructors. These results may seem to indicate that the comprehensibility and effectiveness of the teacher written feedback is yet to be enhanced. When it came to the questions as to what types of feedback and the focus of error the participants would like to receive and emphasize more, over 50% of the respondents from both groups indicated that they would like to receive more teacher written comments about their

writing rather than just the language errors being identified or highlighted, despite the fact that the 'use of language' remained to be an area they are most concerned about when comparing with other areas such as content and organization. The majority of the respondents were quite error-conscious as 70% of them from both groups demanded that all errors be identified by their language instructors.

5.5.2   Marking Policy

(a) Teacher-marker Data

Two instructors working in the same local university agreed to participate in the study. The teachers were both Cantonese-speakers, with teaching English as a second language experience for 8-10 years and teaching on the same course (i.e. LANG206) course) for five years. They were both subjects-trained (i.e. with an English subject knowledge qualification recognized by the Hong Kong Government). These two selected instructors were named 'Teacher A' and 'Teacher B' in this study.

(b) Selection of teacher-markers in this study

Two markers (namely Teacher A and Teacher B in this study) were trained to be 'reliable' assessors who were fine tuned to the 'institutional' marking standard and thus be more able to

assess student writing fairly against the benchmark. In the Language Centre we are serving, all language teachers are required to go through somehow a center-wide marking reliability test each semester in which teachers' reliability in giving a fair mark in writing and speaking papers is being assessed. Also, an additional procedure was done at the course-level to screen in teaching members to be participating teachers in this study such that both Teacher A and Teacher B demonstrated a pattern of parity in their marking standard. Efforts were geared to avoid selecting teachers (1) whose past records showed a tendency of discrepancy in their grades given to the scripts of similar performance, and (2) whose mean scores and mean standard deviations deviated far away from the overall mean and overall standard deviation. For example, this could avoid the situation that Teacher A tended to be more relatively lenient in grading whereas Teacher B tended to be relatively more stringent. Although some may argue that doing so might not have direct correspondence to ensure that any difference found in their focus of feedback in the end was attributed to different forms of feedback treatment, such additional procedure would do no harm but may do some way to strengthen the validity and reliability of the study. To ensure parity across teachers' marking standard (for both Teacher A and Teacher B):

1. We conducted the marking standardization exercises (moderation meetings and marking based on assessment criteria).

2. Teacher's scripts from the high-med-low bands were selected for checking.

3. The marks of all teachers have been slightly statistically moderated to reflect individual differences in teacher average and teacher standard deviation, so the marks/grades have all been normalized.

According to the internal moderation results conducted by my university over the past five years, Teacher A and Teacher B who were more statistically comparable to the course mean and course standard deviation, and had demonstrated consistent and stable marking standard throughout a considerable period of time. Teacher A and Teacher B who have been teaching on the same course for 5 years were thus selected for this study.

(c) Marker's training

Both Teacher A and Teacher B were teaching the "Year Two English Language for Engineering Students" course, of which I was the coordinator, for over five years. They were experienced markers of the writing assignments being examined in this study. Before all instructors (including Teacher A and Teacher B) started marking the writing assignments of their own classes, the course coordinator (i.e. the researcher of this study) would provide each instructor a marking scheme covering the 'content', 'organization' and 'language' criteria alongside five pieces of sample scripts representing different levels of proficiency. Having marked the sample scripts, all instructors teaching on the same course gathered in the standardization meeting during which the instructors compared and justified their grades given, and discussed the quality of student texts and identifying writing issues under the moderation of the course coordinator who was also the researcher in this study. In the end, all teacher-markers after the thorough discussion agreed to the grades assigned to each of the five pieces of student writing which would finally become the sample benchmark scripts for reference. It is actually through such discussion where instructors teaching on the same course (1) can reconcile each other marking practice with the marking

scheme, (2) build some sort of consensus as to the expected grading standard against various levels of writing performance in the area of 'content', 'organization' and 'language' separately rather than holistically, and (3) develop instructors' abilities for discussing writing issues and giving effective feedback across 'content', 'organization' and 'language' with equal attention. For example, understanding writing problems and common errors identified with year two engineering students in this particular genre of writing, classification of error types, explanation of errors; pedagogical approach to teaching this particular genre of writing, what constitute to good or bad writing practice and etc. This mechanism has become our institutional requirements consecutively for five years. As such, all markers (including Teacher A and Teacher B in this study) were assumed to be reliable and experienced assessors who were fine tuned to the 'institutional' marking standard and thus be more able to assess student writing fairly against the benchmark.

(d) Assessment guidelines

While Lee's studies were testing to what extent a teacher feedback was focusing on 'content', 'organization' and 'language', there seems no clear account from Lee's (2008a & 2008b) studies as to the composition of a grade being given. That is, the weighting across three criteria (i.e. content, organization and language) on an overall grade being given on each student writing are not clearly explained in these studies. For example, whether an assessment policy was in place in Lee's studies that required participating teachers to put equal emphasis across three criteria (i.e. 1/3 weighting for content, 1/3 weighting for organization and 1/3 weighting for language) when considering an overall grade to be given, or there was actually no policy in place at all. In the

absence of clear assessment guidelines for the weighting among these three criteria while conducting research to investigate teacher's emphasis on these three criteria with different feedback forms, the results (i.e. focus of feedback) obtained from those studies to a certain extent might only reflect individual feedback practice which was influenced by a number of extraneous factors.

To test if the focus of feedback to a certain extent reflected the forms of feedback treatment being used in this study (i.e. electronic feedback vs. paper-based feedback), the assessment policy of this study required both Teacher A and Teacher B to put equal emphasis on 'content', 'organization' and 'language' criteria. To achieve this, the assessment guidelines required that participating teachers to give a single grade on each of these criteria and the grade for each criterion carried an equal weighting (i.e. 1/3 for content + 1/3 for organization + 1/3 for language = 100%). It is believed that through this means teachers would tend to make equal effort when assessing and marking student writing against these three criteria, with the knowledge that the final grade on the second draft were very much depending on the frequency of errors reduced in each of the 'content', 'organization' and 'language' criteria. In other words, these three criteria would be assessed independently but not holistically as if in Lee's studies (2008a & 2008b) such that teachers might have a tendency to put unequal weighting on different writing aspects. As a result, the focus of feedback to a certain extent might have reflected individual feedback practice which was influenced by a number of extraneous factors.

(e) Compulsory Marking Principles adopted in this study – Comprehensive Markingn& Indirect Feedback

Regardless of the forms of feedback treatment used (i.e. electronic feedback vs. paper-based feedback), marking must be comprehensive which means all errors needed to be identified by the markers. Also, all markings must be indirect feedback which means teachers indicated students' errors only by means of an underline, circle, code, or other mark but no provision of the correct form, leaving students to solve the language problems that have been called to their attention (Ferris, 2006, p. 83). As part of the efforts in ensuring any statistical difference in language gains was attributed to the different forms of feedback adopted, the following principles were imposed on both Teacher A and Teacher B when marking students' papers:

1. Mark comprehensively

2. Make the fewest possible changes

3. Avoid giving explicit corrections (i.e. error correction; inserting missing
   words; crossing out redundant words; imposing a reformulated syntactic structure; etc.)

4. Provide implicit error feedback (i.e. error codes; any input/simple comment which is
   reasonably comprehensible to students)

Failure to adhere to the above principles would make the subsequent data analysis difficult. For example, if one teacher was adopting comprehensive marking but the other was adopting selective

marking, the effect of different forms of feedback treatment on the frequency of errors reduced would become inaccurate, undermining the validity and reliability of the study.

(f) Standardization of Error Types

The subsequent data analysis stage was foreseen to be complicated if both Teacher A and Teacher B did not adopt the marking codes consistently. To reconcile the error coding system adopted by Teacher B when using the paper-based feedback with the error codes standardized in the electronic feedback system, the error codes used by Teacher B were made reference to with the corresponding error codes standardized in the electronic feedback system. That means, after they had marked the first drafts, I, as a researcher and moderator, looked up to every single error identified in the paper-based feedback with Teacher B, clarified the meaning of his error codes and then re-categorized each of them with reference to the corresponding error codes standardized in the electronic feedback system for the error count purpose (See Appendix B). By doing so, the types and frequency of errors identified in the subject group and control group will be made comparable when conducting the error count. The table in Appendix B shows the one-to-one correspondence between the standardized coding system used in the electronic feedback and the codes marked in the paper-based feedback. It is evident from the table that some errors of the same type were coded inconsistently in the paper-based feedback. For example, 'a wrong word' error type could be coded as '✗', 'wrong word', or simply by underling the word without any code. Some error types like 'Redundant Determiner', 'Redundant Verb/Auxiliary' and 'Redundant Conjunction' were using

the same set of codes in the paper-based feedback which might not have helped student in identifying the error types precisely.

Truscott (1996) argued that practical problems like incomplete, inconsistent, and inaccurate teachers' error feedback have rendered the writing feedback ineffective. Against the above, an attempt has been made on my part to incorporate a consistent system of coded feedback covering content, organization and language into an electronically interactive platform called 'Mark My Words' (MMWs) (See Appendix A).

Despite the fact that the methodology of this study was modeling on the methodologies adopted in Ferris (2006) and Lee's (2008a & 2008b) studies, the most notable difference between the methodology of this study and theirs is that the teachers involved in Ferris's and Lee's studies might have exhibited various teaching practice shaped by their own backgrounds like their own feedback practice, school policy or interpersonal relationship with students, let alone those teachers coming from different schools in Lee's studies (2008a & 2008b). Unfortunately, there seems no indication from their studies if any pre-emptive measures, like what (a) – (f) were implemented in this study, were imposed to minimize those extraneous factors. In such cases, any discrepancy arisen between the experimental groups and control groups might have not been easily attributed to a single factor but multiple ones surrounding the feedback practice in which it took place. For example, Lee (2008b) generalized from her findings that students of higher proficiency in School A relatively welcomed more error feedback than students of lower proficiency did in school B. However, such difference in students' attitudes towards error feedback might have,

wholly or partially, come from other extraneous factors like teachers' marking practice, school culture, interpersonal relationship between students and teachers.

It can be seen from Chapter 5.4 and Chapter 5.5 that reasonable efforts have been made to exclude other extraneous factors that might undermine the validity and reliability of the results. However, some may still challenge the reliability and validity of the research outcomes by arguing that how one can tell if any language gain (i.e. percentage of error reduction by error categories between the experimental group and the control group) after receiving a particular feedback treatment was as a result of the different form of feedback treatment (i.e. electronic feedback vs. paper-based feedback), but not as a result of teachers' individual differences. While it is acknowledged that there were limitations of the study which could hardly exclude all extraneous factors, any teachers' individual feedback practice or feedback preference were to a great extent constrained by the feedback environment (i.e. computer-mediated and paper-based environment) as well as the marking policy (e.g. focus, comprehensiveness, level of explicitness) imposed in this study. As such, it is believed that the effect of individual markers' practice on the validity and reliability of the outcomes has been reasonably minimized.

5.6   *Error Reduction*

The method adopted for counting the number of error reduction was modeled on the one used in Ferris's study (2006). This is deemed to be an objective and fundamental way to evaluate if students have their writing revisions improved when compared with their first draft. It is because the capability to improve the quality of students' texts, which is more easily evident by 'error count', is central to claims made about the effectiveness of teacher written feedback. The 'error count' comparison serves to examine if the electronic feedback, which was employed in the experimental group, makes a statistical difference in improving students' language accuracy in their writing revisions when compared with the control group which received the paper-based feedback. An experimental design was adopted to collect the data in this area. "The basis of the experimental method is the experiment, which can be defined as a test under controlled conditions that is made to demonstrate a known truth, or examine the validity of a hypothesis" (Muijs, 2011, p.11). An experiment design was employed because it can be used to determine causality between variables (Muijs, 2011). The study examines whether an intervention, which is use of the electronic feedback, will improve students' language accuracy in their writing revisions. The two hypotheses are:

(1) Null hypothesis (H0): the electronic feedback will not improve students' language accuracy in their writing revisions.

(2) Alternative hypothesis (H1): the electronic feedback will improve students' language accuracy in their writing revisions.

Before the experiment was carried out, both groups had received a language proficiency pre-test to ensure that students of these two groups were of comparable competence. After the students had submitted their second draft based on the teacher written feedback on the first draft, the same teachers were assigned to mark the papers, and then the frequency of errors by error categories was collated for subsequent analysis at the end.

In line with the rationale of process writing, both experimental group and control group submitted their draft on the same topic twice. The topic was about a proposal for an innovative technology. It was expected that the final draft (i.e. the second draft) should have been revised in response to teacher written feedback on the first draft. Partially modeling on the approach adopted by Ferris in her study titled '*Does error feedback help student writers? New evidence on the short term and long term effects of written error correction*' (2006), the actual effectiveness of the teacher written feedback was evaluated by counting the total number of errors students respectively from the experimental group and the control group made before and after receiving the feedback. For each draft receiving the electronic feedback (i.e. subject group), errors (or feedback points) in each piece of student writing were identified and coded according to the system of the electronic feedback databank (see Appendix C). As for each draft receiving the paper-based feedback (i.e. control group), I looked up to every single error identified by the teacher, then re-categorized each of those identified errors with reference to the corresponding error types standardized in the electronic feedback databank (See Appendix B). By doing so, the types and frequency of errors identified in the subject group and control group were made comparable in the error count. For each batch of the first draft and final draft, the total number of errors 'by category and frequency' will be added

up separately by the experimental group and the control group (See Appendix C). Finally, the percentages of error reduction across error types between the first and final drafts by treatment groups were computed. For the convenience of statistical analysis, these 77 error types were deduced into seven error categories (see Appendix B). The findings, yielding mainly descriptive data and hence tendencies were reported, allowed us to:

(1) Identify and rank the frequency of errors by types/categories in the first draft from each group;

(2) Identify and rank the frequency of errors by types/categories in the second draft from each group;

(3) Evaluate which form of feedback treatment was more effective in reducing particular error categories;

(4) Evaluate which error categories were more responsive to any particular form of feedback treatment in error reduction.

This 'error count' statistical analysis was to be conducted by R program, and the analytical process of which has already been detailed in Chapter 5.3.1 'Statistical Analysis: Two Sample t-tests for Error Count'.

5.7   *Students' Perceptions on Teacher Written Feedback*

The 'questionnaire' method adopted in this part was modeled on the one used in Lee's study (2008b). The purpose of this second set of comparison serves to examine students' perceptions towards the form of feedback they received, as well as to probe into their attitudes as to what may constitute to an effective teacher written feedback. While the quantitative analysis (i.e. 'Chapter 5.6 Error Reduction') remains the core part of the study, results alongside the comments obtained from the questionnaire were used to help provide some details about the statistical results and provide us some insights about students' attitudes towards teacher writing feedback.

A student questionnaire was administered to all 62 students receiving the electronic feedback and paper-based feedback. The questionnaire (See Appendix D) was modeled on the one primarily developed by Lee (2008b). It aimed at investigating students' perceptions of teacher written feedback in improving the overall accuracy and appropriateness of student writing in the areas of the:

(1) Legibility of feedback

(2) Student ability to correct errors

(3) Types of feedback (i.e. grades / error feedback / written comments)

(4) Focus of error feedback (i.e. content / organization / language)

(5) Amount of error feedback

(6) Focus of written comments (i.e. content / organization / language)

(7) Types of error feedback strategy

(8) Students' responsiveness of error types

(9) Types of post-writing activity

(10) Usefulness of feedback

The results and explanation of the questionnaire are detailed in Chapter 7 'Findings & Discussions: Students' Perceptions on Teacher Written Feedback'.

5.8   *Data Analysis & Triangulation*

Other than comparing and reconciling the results of error reduction (see Chapter 6) with students' perceptions on teacher written feedback (see Chapter 7) obtained from the experimental group and the control group, the results gathered within this study were also exclusively compared and reconcile with those of Truscott's study (1996), Ferris's study (2006) and Lee's studies (2008a & 2008b) alongside some common observations derived from some other previous research studies. The reasons for doing so are that this would help put together a fuller, more meaningful and insightful picture as to the impact of teacher written feedback on student writing revisions from various perspectives, and would also help reconcile our results with any contradictory evidence (or vice versa) obtained from Ferris's study (2006) and Lee's studies (2008a & 2008b) on which the methodology of this study was based.

This research aimed at evaluating the actual effectiveness of the electronic feedback on error reduction, when comparing to the traditional paper-based feedback. An explanatory mixed-methods research design was adopted, which was modeled on Ferris's (2006) and Lee's (2008b) methodologies. The data collection was divided into two phrases. The first phase was conducted by 'error count' with a primary objective (i.e. research questions 1 and 2) of comparing student uptake on two different forms of feedback (i.e. electronic feedback vs. paper-based feedback). The second phase was conducted by administering questionnaires with a secondary objective (i.e. research question 3) of understanding the students' attitudes toward the feedback they received. Participants were 62 year-two Engineering students who were enrolled in the same English course with a comparable language proficiency over one semester. Convenience sampling was adopted. A number of measures was implemented to ensure the validity and reliability of the study. Quantitative results of this study were triangulated with the qualitative results of the study, as well as with the results of other studies (e.g. Ferris, 2006; Lee, 2008b; Truscott, 1996).

Chapter 6   Findings and Discussions: Error Reduction

This chapter covers the error reduction by examining the actual effects of the electronic feedback and paper-based feedback in reducing the frequency of errors in student writing revisions. Modeling on the approach adopted by Ferris (2006) in her study titled '*Does error feedback help student writers? New evidence on the short term and long term effects of written error correction*', the actual effectiveness of the teacher written feedback was evaluated by counting the total number of errors, by category and frequency, students respectively from the experimental group and the control group made before and after receiving teacher written feedback. Ferris's study (2006) on the effectiveness of teacher written feedback covered 16 error types whereas the present study extended the number of error types to 77 which were then further divided into 7 error categories (i.e. 'Awkwardness' errors, 'Clausal-level' errors, 'Word-level' errors, 'Collocation' errors, 'Content' errors, 'Organization' errors, and 'Tone & Style' errors). The table in Appendix A shows these 77 error types under 7 categories identified for an investigation in this study. The tables in Appendix C also show how well students in both the experimental group and control group performed in revising the various error types after receiving a particular form of feedback treatment within their own groups.

Other than comparing the results of error reduction obtained from the experimental group and the control group and drawing their implications, the results gathered within this study were also compared with those of Ferris's study (2006) and Truscott's study (1996) alongside some common observations derived from some other previous research (e.g. Lalande, 1982; Zamel, 1985; Robb

et al., 1986; Ferris & Roberts, 2001; Hyland & Hyland, 2006; Lee, 2008a & 2008b; Bitchener & Knoch, 2010b). It is noted that Ferris (2006) and Truscott (1996) took a conflicting position with regard to the effectiveness and values of teacher written feedback. The findings and analysis in this section added on to their substantial discussions not only about the controversy over the effectiveness of teacher written feedback but also about the limitations and constraints faced by teachers and students in the feedback environment. The reasons for doing so are that this would help put together a more complete, meaningful and insightful picture as to the impact of teacher written feedback on student writing revisions from various perspectives, and would also help reconcile our results with any contradictory evidence (or vice versa) obtained from Ferris's study (2006) on which the methodology of this study was partially based.

In summary, it can be concluded that the electronic feedback ('Mark My Words') modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' is relatively more effective in error reduction across all categories, namely (1) 'Awkwardness' errors, (2) 'Clausal-level' errors, (3) 'Word-level' errors, (4) 'Collocation' errors, (5) 'Tone & Style' errors, (6) 'Content' errors and (7) 'Organization' errors, in both 'between-groups' and 'within-groups' comparison (see Appendix G: Summary of the Statistical Results).

It is also noteworthy to highlight that students receiving electronic feedback were more successful than those students receiving paper-based feedback in revising 'Awkwardness' errors and 'Collocation' errors, which are highly characterized as 'L1-L2 interference' errors and regarded as 'untreatable errors' by Ferris (2006). This is evident by the findings that these two error categories were the only categories which made no statistical significant revisions even in the

'within-groups' comparison of the paper-based feedback group (i.e. control group). For all seven error categories, despite the fact that the control group receiving the paper-based feedback was statistically less effective than its counterpart in remedying errors in the 'between-groups' comparison, at least the paper-based feedback still managed to make a statistically significant number of revisions for most error categories within its own treatment group, with the exceptions of 'Awkwardness' errors and 'Collocation' errors. These findings lend weight to Ferris's (2006) argument that non-rule governed 'sentence-level' errors (which is similar to what 'Awkwardness' errors was defined in this study) and 'word choice & idioms' (which is similar to what 'Collocation' errors was defined in this study) are the most difficult error items for students to correct. On the other hand, the results of the questionnaires in Chapter 7.10 and Chapter 7.11 (see Q14-Q17 in Appendix E) also lend support to these findings and Ferris's argument when students were asked to rate all error categories in terms of their level of difficulty for comprehension and correction.

6.1 Pre-test & Post-test

Before carrying out the experiment, both the experimental and control groups received a language proficiency pre-test to ensure that students of these two groups were of comparable language competence. There is no evidence for a statistical difference in the Pre-test and Post-test mean scores for the electronic feedback group (experimental group) and for the paper-based feedback group (control group). The p-value for the pre-test is 0.816, and the p-value for the post-test is 0.763 (Tables 6.1 & 6.2 in Appendix E). Therefore, the overall language competence was the same for both groups.

6.2 Total Errors

'Total Errors' in this study refers to all 7 error categories, comprising 77 error types, identified in student writing. They include (1) 'Awkwardness' errors, (2) 'Clausal-level' errors, (3) 'Word-level' errors, (4) 'Collocation' errors, (5) 'Tone & Style' errors, (6) 'Content' errors and (7) 'Organization' errors. An overview of error count before the statistical analysis for the top ten errors are appended in Appendix C. It is found that 'Awkwardness', 'Singular-Plural Form', 'Missing Determiner', 'Sentence Fragment', 'Run-on Sentence' and 'Wrong Verb Form' are the common error types identified in both experimental group and control group.

When examining the percentage of error reduction across all 77 error types under these 7 error categories in Appendix C, it is observed that the experimental group receiving the electronic

189

feedback achieved the highest percentage of error reduction across all error types when compared to the control group receiving the paper-based feedback. Having said that, statistical analyses (i.e. a two sample t-test) were conducted in this section and the following sections to examine if either form of feedback treatment could offer a more statistically significant difference in the total error reduction (Chapter 6.2), as well as in each of the 7 categorical error reduction (Chapters 6.3 – 6.9)

Findings I (Total Errors between the Treatment Groups)

A Two Sample t-test was conducted between experimental group and control group in order to find out if either group performed better than the other in reducing the total number of errors. The summary statistics of the experiment group and the control group on 'Total Errors' can be seen in Table 6.3 in Appendix E.

For the experimental group receiving the electronic feedback, a total of 615 errors were collected from the 30 students' first drafts marked by Teacher A. There was an average of 20.46 errors per first draft (of 550 words on average – i.e. 1 error per 27 words). A total of 64 errors were collected from the 30 students' second drafts marked by the same teacher. There was an average of 2.13 errors per second draft (of 550 words on average – i.e. 1 error per 258 words).

As for the control group receiving the paper-based feedback, a total of 614 errors were collected from the 32 students' first drafts marked by Teacher B. There was an average of 19.16 errors per first draft (of 550 words on average – i.e. 1 error per 28.66 words). A total of 304 errors were

collected from the 32 students' second drafts marked by the same teacher. There was an average

of 9.5 errors per second drafts (of 550 words on average – i.e. 1 error per 57.89 words).

The average number of total errors per first draft made by experimental group and control group

was 20.46 and 19.16 respectively (Table 6.3 in Appendix E) which were very close. However, the

average number of total errors per second draft made by both groups was 2.13 and 9.5 respectively

(Table 6.3 in Appendix E). The difference in the average number of total errors respectively made

by the experimental group and control group in draft 2 was relatively widened.

The statistical results (between groups) imply that the overall language competence before

treatment was the same for the electronic feedback group and the paper-based feedback group

(Table 6.4 in Appendix E: p-value [draft 1] = 0.629), but the average number of total errors made

by the electronic feedback group in draft 2 was significantly smaller than that made by the paper-

based feedback group in draft 2 (Table 6.4 in Appendix E: p-value [draft 2] = 5.697e-06). Such

difference (between groups) is educationally significant as well as being statistically significant,

because the effect size (= -1.325) is large (Table 6.4 in Appendix E).

In conclusion, these 'between groups' results indicated that students in the experimental group

receiving the electronic feedback were able to make a statistically significant number of revisions

across all 7 error categories (i.e. 'Awkwardness' errors, 'Clausal-level' errors, 'Word-level' errors,

'Collocation' errors, 'Content' errors, 'Organization' errors, 'Tone & Style' errors); whereas

students in the control group receiving paper-based feedback group did not in such 'between

groups' comparison.

Findings II (Total Errors within the Treatment Groups)

Another two Sample t-test was conducted within each of the experimental group and control group in order to find out if each individual treatment group itself helped reduce the total number of errors in draft 2 when comparing to that in draft 1. The summary statistics of the experiment group and the control group on 'Total Errors' can be seen in Table 6.3 in Appendix E.

When testing the effectiveness of each feedback treatment in reducing the total number of errors within its own feedback group, it is found that both the electronic feedback and the paper-based feedback (within its group) were able to make a statistically significant number of revisions in the total number of errors in draft 2 (Table 6.5 in Appendix E: p-value [control group] = 7.496e-05 with an effect size = 1.07, and p-value [experimental group] = 2.736e-10 with an effect size = 2.34). However, with the smaller p-value and the larger effect size, the electronic feedback (within its group) was able to make a more statistically significant number of revisions in the total number of errors when comparing those made by the paper-based feedback.

In conclusion, these "within groups" results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in reducing the total number of errors across all 7 error categories (i.e. 'Awkwardness' errors, 'Clausal-level' errors, 'Word-level' errors, 'Collocation' errors, 'Content' errors, 'Organization' errors, 'Tone & Style' errors).

Discussion

The positive findings obtained from the experimental group contradicted Truscott's (1999) arguments against the effectiveness of error correction on student writing revisions, which were rested on the aspects of (1) teacher's limitations in feedback practice and (2) student's limitations in comprehend sion and adoption of teacher written feedback. In fact, the conflicting results between Truscott's (1996) observations and the findings in this study could be attributed to the pedagogical approach of 'Mark My Words' ('MMWs') which is underpinned by the rationales postulated by Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' (please refer to Chapter 3.1 'Theoretical Framework and Language Theories'). In an attempt to remedy the above limitations which were believed to render teacher written feedback ineffective, this 'three-step approach' pedagogy empowered by its comprehensive and consistent error feedback database is synchronized into the operation of 'Mark My Words' ('MMWs').

For the aspect of teacher's limitations, Truscott and previous researchers (cited in Ferris, 2006) argued that practical problems like incomplete, inconsistent, and inaccurate teachers' error feedback have rendered the writing feedback ineffective. To ensure more detailed, comprehensible and standardized feedback comments to be given for each error type, seventy-seven electronic feedback comments covering content, organization and language are pre-set. This contributes to the electronic feedback database of 'Mark My Words' ('MMWs'). The findings showing that the experimental group made a statistically significant number of reduction in their total number of

errors across all these seven categories between the first draft and the second draft do not support the claims of some previous researchers that teachers give incomplete and inaccurate error feedback and that students ignore teacher feedback or cannot utilize it effectively in their writing revisions (Cohen & Robbin, 1976; Truscott, 1996; Zamel, 1995).

As for the aspects of student's limitations in comprehension and adoption of teacher feedback, Truscott (cited in Hyland, 2003, p.218) attributed the ineffectiveness of teacher feedback to the students' lack of skills to understand and use the feedback. Their lack of skills to comprehend and adopt teacher feedback might possibly due to the theoretical issue raised by Truscott (1996) who argued that different types of linguistic forms may take a different sequence of L2 acquisition. This means a student's interlanguage (i.e. proficiency) level would determine to what extent he or she can comprehend, process and utilize teacher feedback for a particular error type. To put Truscott's theoretical issue into perspective, the sequence of second language acquisition should begin with word-level linguistic forms, and then followed by phrase-level, clausal-level and discourse level. A student who is at a point of struggling with the acquisition of clausal-level linguistic forms in his continuum of interlanguage stands a high possibility of finding teacher feedback on 'Clausal-level' errors (e.g. run-on sentence, sentence fragment, non-standard grammatical structure heavily interfered by L1) incomprehensible if the feedback is not made explicit for the student. Such causal relationship between students' interlanguage level and their comprehension and adoption of teacher feedback for a particular error type lends weight to Ferris's (1996 & 1999) argument in that students' level of progress in error correction varied depending on error types. This implies that different feedback strategies, at various levels of explicitness, should be prescribed for different error types. However, Truscott (cited in Ferris, 2006, p.83-84)

194

argued that teachers often responded to errors of all categories in the same way without sufficient awareness that different types of linguistic forms may take different order and sequence of L2 acquisition, while ignoring variation in students' interlanguage level to process and utilize teacher feedback successfully.

To improve students' comprehensibility and adoption of teacher feedback, 'Mark My Words' ('MMWs') provides a feedback environment where students instead of teachers can exercise their autonomy to determine, according to their interlanguage level and interests, how much information and how explicit the teacher feedback on a particular error type they would like to receive. Hence, with its capacity of reconciling the above teacher's and student's limitations, the electronic feedback has rendered teacher written feedback effective for student writing revisions in this study.

6.3 *'Awkwardness' Errors*

'Awkwardness' errors in this study refers to the "L1 – L2 interference" (or 'Chinglish') – an L2 utterance which is heavily interfered by the mother tongue's (L1) grammatical structure, and this non-standard structure cannot be simply explained by or identified with grammatical rules (i.e. not rule-governed) and specific marking codes, but to a certain extent shares similar syntactic patterns with the Chinese equivalence. Given the proficiency level of the subjects in both experimental group and control group, their utterances produced may still be semantically understandable but they may not be syntactically correct. For example:

1. People will be *more and more to come* (Chinese syntactic structure) to <u>walk behind the trend</u> (Chinese syntactic structure) of iPhone 5.

2. *There are three people go to school*. (Chinese syntactic structure)

3. Recently, *the born of* (Chinese syntactic structure) an advanced communication technology is just around the corner.

4. *The problem will be smaller and smaller …* (Chinese syntactic structure).

<u>Findings I ('Awkwardness' Errors between the Treatment Groups)</u>

A Two Sample t-test was conducted between the experimental group and control group in order to find out if either group performed better than the other in reducing 'Awkwardness' errors. The summary statistics of the experiment group & control group on 'Awkwardness' errors can be seen in Table 6.6 in Appendix E.

The average number of 'Awkwardness' errors per first draft made by the experimental group and control group (between groups) was 1.57 and 2.25 respectively (Table 6.6 in Appendix E) which were very close. However, the average number of 'Awkwardness' errors per second draft made by both groups was 0.167 and 1.66 respectively (Table 6.6 in Appendix E). The difference in the average number of 'Awkwardness' errors respectively made by the experimental group and control group in draft 2 was relatively widened.

The statistical results (between groups) imply that the electronic feedback group and the paper-based feedback group demonstrated the same level of performance with respect to avoiding 'Awkwardness' errors in draft 1 before any intervention took place (Table 6.7 in Appendix E: p-value [draft 1] =0.1237), but the average number of 'Awkwardness' errors made by the electronic feedback group in draft 2 was significantly smaller than that made by the paper-based feedback group in draft 2 (Table 6.7 in Appendix E: p-value [draft 2] = 0.000149). Such difference (between groups) is educationally significant as well as being statistically significant because the effect size (=-1.047) is large (Table 6.7 in Appendix).

In conclusion, the "between groups" results indicated that students in the experimental group receiving the electronic feedback were able to make a statistically significant number of revisions in the average number of 'Awkwardness' errors; whereas students in the control group receiving paper-based feedback group did not in such 'between groups' comparison.

Findings II ('Awkwardness' Errors within the Treatment Groups)

Another two Sample t-test was conducted within the experimental group and control group in order to find out if each individual treatment group itself helped reduce the number of 'Awkwardness' errors between draft 1 and draft 2. The summary statistics of the experiment group & control group on 'Awkwardness' errors can be seen in Table 6.6 in Appendix E.

When testing the effectiveness of each feedback treatment in reducing the number of 'Awkwardness' errors within its own feedback group, it is found that paper-based feedback (within its group) failed to make a statistically significant number of revisions in the average number of 'Awkwardness' errors (Table 6.8 in Appendix E: p-value [control group] = 0.227) when comparing the average number of 'Awkwardness' errors occurred in draft 1 and draft 2. However, the electronic feedback (within its group) was able to make a statistically significant number of revisions in the average number of 'Awkwardness' errors when comparing the average number of 'Awkwardness' errors occurred in draft 1 and draft 2. (Table 6.8 in Appendix E: p-value = 1.32e-05 with an effect size = 1.297). These "within groups" results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in remedying 'Awkwardness' errors.

With as much as 89% of 'Awkwardness' errors being corrected in draft 2 (See Appendix C), the electronic feedback treatment significantly lowered the frequency of 'Awkwardness' errors than the paper-based feedback treatment which could only have 29% of 'Awkwardness' errors' being corrected in draft 2.

Discussion

When comparing to this study which examined the students' revisions of 77 error types in response to the two different forms of feedback treatment, Ferris's study (2006) merely investigated 16 error types in which one of the error types was categorized as 'Sentence Structure'. 'Sentence Structure',

which was referred in Ferris's studies (1999, 2006) as an 'untreatable errors' alongside 'word choice and idioms', shared a very similar error category to 'Awkwardness' errors, which was referred in this study as non-standard forms mainly due to 'L1-L2 Interference' (e.g. Chinglish).

For pedagogical purposes, Ferris (1999 & 2006) suggested a dichotomy between 'treatable errors' and 'untreatable errors'. On the one hand, errors are considered "treatable because they occur in a patterned, rule-governed way" (1999, p.6). Examples of treatable errors "include problems with verb tense or form, subject verb agreement, run-ons, fragments, noun endings, articles, pronouns and spelling" (2006, p.96). On the other hand, errors are considered 'untreatable' when "there is no handbook or set of rules students can consult to avoid or fix those types of errors" (1999, p.6). Examples of untreatable errors include 'word choice and idiom'. Ferris's dichotomy, by coincidence, is very similar to the way this study identified with 'Awkwardness' errors and 'Collocation' errors which are non-rule governed.

With this study falling along a very similar line with Ferris's studies (1999, 2006) in categorizing 'treatable' and 'untreatable' errors, this has allowed me to make a one-to-one parallel comparison of error reduction between 'Awkwardness' errors in this study and 'Sentence Structure (non-rule governed)' errors in Ferris's study (2006).

		% Reduction on 'Awkwardness' error / 'Sentence Structure (non-rule governed)' errors
This study	Paper-based Feedback Group	29%
	Electronic Feedback Group	89%
Ferris's study (2006)	Paper-based Feedback for All Groups	84%

Figure 6.1

According to Ferris's study (2006), the results indicated that as much as 84% of 'Sentence Structure' errors (i.e. non-rule governed sentence-level errors = 'Awkwardness' errors in this study) were corrected in draft 2 (see *Figure 6.1*). Ferris's findings (2006) appeared to have weakened the conclusion which was previously drawn in the present study that the electronic feedback treatment significantly lowered the frequency of 'Awkwardness' errors than the paper-based feedback treatment. The conflicting evidence might have cast the doubts that the apparent success of the electronic feedback in remedying 'Awkwardness' errors could have been attributed to some other extraneous factors but not mainly on a particular form of feedback treatment adopted. On the other hand, Ferris's findings also appeared to have contradicted the proposition Ferris herself put forward in her previous research which stated that students were less able to correct their errors through feedback on form if the errors were not rule-governed (Ferris, 1996 & 1999) when having regard to her study (2006) which indicated that as much as 84% of 'Sentence Structure' errors (identical to 'Awkwardness' errors in this study) were corrected.

However, Ferris's high level of achievement in correcting 84% of 'Sentence Structure' errors was under query because it was reported that as much as 75% of 'Sentence Structure' errors were marked directly. This means students' relative short-term success in correcting 'Sentence Structure' errors in her study might have been merely attributed to the teacher's provision of corrected linguistic forms in 75% of these cases. Hence, the high student uptake rate of paper-based feedback in correcting 'Sentence Structure' errors (i.e. 'Awkwardness' errors in this study) in Ferris's study (2006) cannot in fact make a strong case that her direct intervention proved any substantial value in correcting 'untreatable' errors when comparing to the electronic feedback where only indirect feedback was allowed in the marking policy of this study.

Error Type	Teacher Feedback Strategies		
	Direct	Indirect	Unnecessary
Treatable	36.7%	58.7%	4.6%
Untreatable	65.3%	33.6%	1.1%

Figure 6.2

Moreover, Ferris's (2006) study indicated that "there was a longitudinal difference in student achievement based on whether errors were 'treatable' or 'untreatable', and this distinction may possibly be attributed to widely disparate teacher feedback strategies…" (p. 98). In other words, the long-term success to avoid making the same error, especially 'untreatable' errors, depends on the level of explicitness of individual teacher feedback strategies. In her study examining students' revisions of 16 'treatable and untreatable' error types in response to direct feedback and indirect coded/uncoded feedback between the first draft and the second draft (short-term) of essay 1, and then between first draft of essay 1and the first draft of essay 4(long term), it was found that as much as 65.3% of 'untreatable' errors (i.e. 'sentence structure [non-rule governed], 'word choice'

and 'idioms') were marked directly, whereas 58.7% of 'treatable' errors were marked indirectly (see *Figure 6.2*). This means 'untreatable' errors were often overwhelmingly marked directly while 'treatable' errors tended to be marked indirectly. Ferris's interviews with teachers in the study confirmed that students' proficiency level was the main factor for them to determine the level of explicitness of their feedback strategies for particular error type. For example, "one of the three teachers was heavily biased in favour of direct correction in almost all circumstances, following his instincts that students in this class would be unable to self-edit if the correct forms were not provided" (Ferris, 2006, p. 97). Another reason which might have discouraged teachers from using indirect coded feedback to mark 'untreatable' errors is that the complexity of untreatable errors, heavily characterized by L1 interference, often cannot be easily identified with specific codes and cannot be explained by L2 grammatical rules. The case in point is 'Awkwardness' errors (i.e. Chinglish). This might possibly have resulted in teachers resorting to the use of direct feedback as a quick fix to these language problems.

However, when comparing both the short-term (draft 2 of essay 1 to draft 1 of essay 1) and long-term (from essay 1 to essay 4) effects in error correction with direct feedback and indirect coded/uncoded feedback, it was found that "the direct intervention (on an untreatable 'sentence structure' errors) did not appear to have any lasting effect over time", whereas "the indirect feedback that students received on verb errors (treatable errors) may have helped them more over time because it consistently called this error to their attention…" (Ferris, 2006, p.96).

As far as the above is concerned, Ferris's study (2006) has revealed two major issues which are worth further scrutiny: (1) Should direct/indirect feedback be used as the means to achieve

successful acquisition of the target linguistic forms over time or be used as an end to correct students' linguistic errors in future revisions? (2) Does Ferris make a valid case that indirect feedback is superior to direct feedback in improving students' language accuracy, given what we have seen that indirect feedback was mostly used for 'treatable' errors and direct feedback was mostly used for 'untreatable' errors (see *Figure 6.2*)?

For the first issue, Ferris (2006) reported that teachers in her study often resorted to direct feedback for students of lower proficiency. This situation has posed a dilemma for ESL teachers whether feedback should be used as part of the guided learning process (as that in indirect feedback), or using feedback (as a quick fix as that in direct feedback) to ensure students can edit the erroneous forms correctly in the next draft. Lalande (1982) advocated indirect coded feedback rather direct feedback in the belief that indirect coded feedback can trigger the "guided learning and problem-solving" process (p. 140), which is more conducive to students' long term language development. Lalande's (1982) proposition was confirmed by Ferris (2006) who concluded that the findings in her study "make a strong case for the advantage of indirect feedback over direct feedback for facilitating student improvement over time, at least with this particular population of students" (p. 98).

As for the second issue about validity, Ferris's strong case was merely drawn from her findings that 'verb tense / form' errors which were consistently received indirect coded feedback showed more significant long-term gains in accuracy over time when comparing to 'sentence structure' errors which were consistently received direct feedback. Despite the above, a question remains as to whether it is valid to put on par students' revisions of 'Word-level' errors (i.e. 'verb tense/form'

errors) and students' revisions of the 'Sentence-level' errors ('sentence structure' errors) towards the level of explicitness in error correction; and then use the results to evaluate if the level of explicitness would have any effects on the students' improvement in accuracy. This is because 'Sentence Structure' errors, representing at the later stage of L2 acquisition after 'verb tense/form', is supposed to be more difficult for students to acquire; whereas 'verb tense/form' errors tends to be easier to be identified and corrected. The distinction between the word-level linguistic forms and sentence-level linguistic forms in their sequence of L2 acquisition might have theoretically made students' revisions of the 'verb tense/form' errors less responsiveness to the level of explicitness (i.e. students could have been able to correct 'verb tense/form' errors no matter if direct, indirect coded or indirect uncoded feedback was given), while students' revisions of the 'sentence structure' errors are theoretically more responsive to the level of explicitness in the first place. Hence, this might have possibly weakened the validity of Ferris's case that argued for the advantage of indirect feedback over direct feedback for error correction.

As part of the purpose of this study to evaluate the student uptake of different forms of feedback on different error categories, this study has strictly imposed the marking principles under which only indirect coded / uncoded feedback can be given. Moreover, unlike Ferris's (2006) study which compared the student revisions of different error types towards error feedback in varying levels of explicitness, this study only compared the revisions of the same error types between the experimental group and the control group in which the former received the electronic feedback and the latter received the paper-based feedback. This can ensure that any significant difference noted in the student revisions of specific errors could be attributed to the adoption of a particular feedback form itself, rather than other confounding variables exhibited in Ferris's (2006) study

like teachers' intuitive choice of direct or indirect feedback and the invalid comparison of error types representing different sequence of L2 acquisition (i.e. word-level forms vs. sentence-level forms in Ferris's study). Referring back to the findings of this study where both the experimental group and control group adopted indirect feedback, the indirect feedback received by the experimental group made statistically significant number of revisions in 'Awkwardness' errors, while the indirect feedback received by the control group did not. These conflicting findings between the experimental group and the control group might somehow weaken Ferris's argument for the absolute advantage of indirect feedback in improving students' language accuracy, at least in response to 'Awkwardness' errors. Although both treatment groups were adopting indirect feedback in this study, it is believed that the notable advantage of the experimental group in reducing 'Awkwardness' errors was attributed to the pedagogical operation of 'Mark My Words' (MMWs) which can engender an environment facilitating students to comprehend, process and utilize indirect coded feedback in reducing 'Awkwardness' errors.

6.4   *'Clausal-level' Errors*

'Clausal-level' errors in this study refers to an ungrammatical structure (clause / phrase) which requires a reformulation of the whole or partial syntactic structure. Unlike 'Awkwardness' errors being defined in this study, 'Clausal-level' errors is rule-governed (i.e. 'treatable' errors in Ferris's term). This means these errors can be explained by and identified with particular grammatical rules and specific codes. Examples are run-on sentences, sentence fragments, imbalanced structures, mixed construction and wrong relative pronoun reference. Examples of 'Clausal-level' errors:

1. The automobile industry attempts to invent environmental friendly vehicles, they will use less fossil fuel. (run-on sentence - two independent clauses cannot be joined by a comma)

2. Since people has no awareness on environmental protection, as they are wasting more and more energy. (sentence fragments – two dependent clauses cannot form a sentence)

3. The solar panels can be added easily and no recurring costs. (imbalanced structure: an independent clause + noun phrase = unparallel)

4. By exercising makes you fit. (mixed structure - avoid shifting from one word form, tense, or sentence pattern to another in mid-sentence)

Findings I ('Clausal-level' Errors between the Treatment Groups)

A Two Sample t-test was conducted across both the experimental group and control group in order to find out if either group performed better than the other in reducing 'Clausal-level' errors. The summary statistics of the experiment group & control group on 'Clausal-level' errors can be seen in Table 6.9 in Appendix E.

The average number of 'Clausal-level' errors per first draft made by the experimental group and control group (between groups) was 2.4 and 2.34 respectively (Table 6.9 in Appendix E) which were very close. However, the average number of 'Clausal-level' errors per second draft made by both groups was 0.1 and 1.4 respectively (Table 6.9 in Appendix E). The difference in the average number of 'Clausal-level' errors respectively made by the experimental group and control group in draft 2 was relatively widened.

The statistical results (between groups) imply that the electronic feedback group and the paper-based feedback group demonstrated the same level of performance in avoiding 'Clausal-level' errors in draft 1 before any treatment (Table 6.10 in Appendix E: p-value [draft 1] = 0.913), but the average number of 'Clausal-level' errors made by the electronic feedback group in draft 2 was significantly smaller than that made by the paper-based feedback group in draft 2 (Table 6.10 in Appendix E: p-value [draft 2] = 4.153e-06). Such difference (between groups) is educationally significant as well as being statistically significant because the effect size (=-1.347) is large (Table 6.10 in Appendix E).

In conclusion, the "between groups" results indicated that students in the experimental group receiving the electronic feedback were able to make a statistically significant number of revisions in the average number of 'Clausal-level' errors; whereas students in the control group receiving paper-based feedback group did not in this 'between groups' comparison.

Findings II ('Clausal-level' Errors within the Treatment Groups)

Another two Sample t-test was conducted within each of the experimental group and control group in order to find out if each individual treatment group itself helped reduce the number of 'Clausal-level' errors between draft 1 and draft 2. The summary statistics of the experiment group & control group on 'Clausal-level' errors can be seen in Table 6.9 in Appendix E.

When testing the effectiveness of each feedback treatment in reducing the number of 'Clausal-level' errors within its own feedback group, it is found that both the paper-based feedback and

electronic feedback (within its group) were able to make a statistically significant number of revisions in 'Clausal-level' errors between draft 1 and draft 2 (Table 6.11 in Appendix E: p-value [control group] = 0.0244 with an effect size = 0.578), and p-value [experimental group] = 2e-06 with an effect size = 1.514). However, provided that the p-value of the electronic feedback is smaller than that of the paper-based feedback, and its effect size is larger than that of the paper-based feedback (within its group), these "within group" results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in remedying 'Clausal-level' errors.

Discussion

Unlike 'Awkwardness' errors, 'Clausal-level' errors is rule-governed. This means these errors can be explained by and identified with particular grammatical rules and specific codes. Ferris (1996, 1999) argued that students were more able to correct their errors through feedback on form provided that the errors were rule-governed and that the rules governing concerned errors were taught.

Although Findings I showed that indirect electronic feedback made a statistically significant difference in reducing 'Clausal-level' errors while indirect paper-based feedback did not in the 'cross groups' comparison, the 'within groups' comparison in Finding II showed that both indirect electronic feedback and indirect paper-based feedback made a statistically significant difference in reducing 'Clausal-level' errors within their own groups despite the fact that indirect electronic feedback gave a smaller P-value and larger effect size. These inconclusive results suggested that

indirect paper-based feedback and indirect electronic feedback were both effective to the revisions of 'Clausal-level' errors within their own groups, though indirect electronic feedback was found to be relatively more effective in Findings I and II. Hence, the pedagogical operation of 'Mark My Words' ('MMWs'), unlike its significant difference noted in response to 'Awkwardness' errors, failed to demonstrate an absolute advantage over the paper-based feedback in facilitating students to comprehend, process and utilize indirect coded feedback in reducing 'Clausal-level' errors.

6.5   *'Word-level' Errors*

'Word-level' errors in this study refers to an ungrammatical word, or a missing/redundant word which does not require a reformulation at clausal-level. Examples are spelling mistakes, wrong tenses, wrong prepositions, missing prepositions, wrong verb forms, subject-verb agreement and etc. 'Word-level' errors is rule-governed (i.e. 'treatable' errors in Ferris's term).

Findings I ('Word-level' Errors between the Treatment Groups)

A Two Sample t-test was conducted across both the experimental group and control group in order to find out if either group performed better than the other in reducing 'Word-level' errors. The summary statistics of the experiment group & control group on 'Word-level' errors can be seen in Table 6.12 in Appendix E.

The average number of 'Word-level' errors per first draft made by the experimental group and control group (between groups) was 11.73 and 11.22 respectively (Table 6.12 in Appendix E) which were very close. However, the average number of 'Word-level' errors per second draft made by both groups was 1.5 and 5 respectively (Table 6.12 in Appendix E). The difference in the average number of 'Word-level' errors respectively made by the experimental group and control group in draft 2 was relatively widened.

The statistical results (between groups) imply that the electronic feedback group and the paper-based feedback group demonstrated the same level of competence in avoiding 'Word-level' errors in draft 1 before treatment (Table 6.13 in Appendix E: p-value [draft 1] = 0.796), but the average number of 'Word-level' errors made by the electronic feedback group in draft 2 was significantly smaller than that made by the paper-based feedback group in draft 2 (Table 6.13 in Appendix E: p-value [draft 2] = 0.000577). Such difference (between groups) is educationally significant as well as being statistically significant because the effect size (=-0.927) is large (Table 6.13 in Appendix E).

In conclusion, the "between groups" results indicated that students in the experimental group receiving the electronic feedback were able to make a statistically significant number of revisions in the average number of 'Word-level' errors; whereas students in the control group receiving paper-based feedback group did not in this 'between group' comparison.

<u>Findings II ('Word-level' Errors within the Treatment Groups)</u>

Another two Sample t-test was conducted within each of the experimental group and control group in order to find out if each treatment group itself helped reduce the number of 'Word-level' errors between draft 1 and draft 2. The summary statistics of the experiment group & control group on 'Word-level' errors can be seen in Table 6.12 in Appendix E.

When testing the effectiveness of each treatment group in reducing the number of 'Word-level' errors within its own feedback group, it is found that both paper-based feedback and electronic feedback (within its group) were able to make a statistically significant number of revisions in the average number of 'Word-level' errors between draft 1 and draft 2 (Table 6.14 in Appendix E: p-value [control group] = 0.000251 with an effect size = 0.981, and p-value [experimental group] = 1.217e-07 with an effect size = 1.745). However, provided that the p-value of the electronic feedback is smaller, and its effect size is larger than that of the paper-based feedback (within its group), these "within groups" results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in remedying 'Word-level' errors.

<u>Discussion</u>

Similar to 'Clausal-level' errors, 'Word-level' errors is rule-governed. This means these errors can be explained by and identified with particular grammatical rules and specific codes. Ferris (1996,

1999) argued that students were more able to correct their errors through feedback on form provided that the errors were rule-governed and that the rules governing concerned errors were taught. This might have possibly explained why the student uptake of both feedback forms on the 'Word-level' errors was effective when comparing themselves within their own treatment groups.

Although Findings I showed that indirect electronic feedback made a statistically significant difference in reducing 'Word-level' errors while indirect paper-based feedback did not in the 'between groups' comparison, the 'within groups' comparison in Finding II showed that both indirect electronic feedback and indirect paper-based feedback made a statistically significant difference in reducing 'Word-level' errors within the groups themselves despite the fact that indirect electronic feedback gave a smaller p-value and larger effect size. These mixed results suggested that, despite varying in degrees, indirect paper-based feedback and indirect electronic feedback were both effective in remedying the 'Word-level' errors, though indirect electronic feedback was found to be relatively more effective in Findings I. Hence, the pedagogical operation of 'Mark My Words' (MMWs), unlike its significant difference noted in response to 'Awkwardness' errors, failed to demonstrate an absolute advantage over the paper-based feedback in facilitating students to comprehend, process and utilize indirect coded feedback in reducing 'Word-level' errors.

One of the difficulties a second language learner frequently experiences in writing is the choice of words to achieve native-like naturalness (Hou & Pramoolsook, 2012). 'Collocation' errors refers to 'miscollocation' in this study. Nattinger and DeCarrico (1992) defined collocation as 'a string of specific lexical items that co-occur with mutual expectancy' (p. 36) which means words often come together when they come into use under a specific context. Chang et al. (2008) added that 'the meaning of a collocation can be inferred from the component parts' (p.284). Liu (2002) attributed to such 'miscollocation' from ESL learners to the influence of their first language (L1). Examples of 'Collocation' errors are:

1. …make it impossible to cover a *big* public area. (lexical collocation: 'large' instead of 'big')

2. Also the application of the Internet is not limited *on* browsing webpages. (prepositional collocation: 'to' instead of 'on')

3. …*keep* our living quality… (lexical collocation: 'maintain' instead of 'keep')

Findings I ('Collocation' Errors between the Treatment Groups)

A Two Sample t-test was conducted across both the experimental group and control group in order to find out if either group performed better than the other in reducing 'Collocation' errors. The summary statistics of the experiment group & control group on 'Collocation' errors can be seen in Table 6.15 in Appendix E.

The average number of 'Collocation' errors per first draft made by the experimental group and control group (between groups) was 0.73 and 0.44 respectively (Table 6.15 in Appendix E) which were very close. However, the average number of 'Collocation' errors per second draft made by both groups was 0 and 0.41 respectively (Table 6.15 in Appendix E). The difference in the average number of 'Collocation' errors respectively made by the experimental group and control group in draft 2 was relatively widened.

The statistical results (between groups) imply that the electronic feedback group and the paper-based feedback group demonstrated the same level of competence in avoiding 'Collocation' errors in draft 1 before treatment (Table 6.16 in Appendix E: p-value [draft 1] = 0.150), but the average number of 'Collocation' errors made by the electronic feedback group in draft 2 was significantly smaller than that made by the paper-based feedback group in draft 2 (Table 6.16 in Appendix E: p-value [draft 2] = 0.000271. Such difference (between groups) is educationally significant as well as being statistically significant because the effect size (= -1.00940) is large (Table 6.16 in Appendix E).

In conclusion, the 'between groups' results indicated that students in the experimental group receiving the electronic feedback were able to make a statistically significant number of revisions in the average number of 'Collocation' errors; whereas students in the control group receiving paper-based feedback group did not in this 'between groups' comparison.

Findings II ('Collocation' Errors within the Treatment Groups)

Another two Sample t-test was conducted within each of the experimental group and control group in order to find out if each treatment group itself helped reduce the number of 'Collocation' errors between draft 1 and draft 2. The summary statistics of the experiment group & control group on 'Collocation' errors can be seen in Table 6.15 in Appendix E.

When testing the effectiveness of each feedback treatment in reducing the number of 'Collocation' errors within its own feedback group, it is found that paper-based feedback (within its group) failed to make a statistically significant number of revisions in 'Collocation' errors (Table 6.17 in Appendix E: p-value [control group] = 0.846) when comparing the average number of 'Collocation' errors occurred in draft 1 and draft 2. However, the electronic feedback (within its group) was able to make a statistically significant number of revisions in 'Collocation' errors when comparing with average number of 'Collocation' errors occurred in draft 1 and draft 2. (Table 6.17 in Appendix E: p-value [experimental group] = 7.7172e-05 with an effect size = 1.194). These 'within groups' results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in remedying 'Collocation' errors.

With as much as 100% of 'Collocation' errors being corrected in draft 2 (See Appendix C), the electronic feedback treatment significantly lowered the frequency of 'Collocation' errors than the paper-based feedback treatment which could only have 7% of 'Collocation' errors being corrected in draft 2.

<u>Discussion</u>

English collocation is important in receptive as well as productive language competence (Cowie, 1994) in the sense that it is essential for students to distinguish grammatically well-formed sentences that are "natural" from those that are "unnatural" (Malligamas & Pongpairoj, 2005). Actually, collocation is one of the factors responsible for Chinese EFL learners' inadequate writing competence (Meng & Li, 2005). Students often use unnatural English expressions that have the right word items but improper collocations (Wu, 2003). 'Collocation' errors remains one of the main obstacles for learners to achieve native-like proficiency (Hou & Pramoolsook, 2012).

The following three types of miscollocation were evident in our student writing, be it the experimental group or the control group.

(1) "Verb+Noun" collocation errors

The first error type is "Verb+Noun" collocation errors. For example, "It is important for school children to learn new knowledge." The verb 'learn' does not collocate with the noun 'knowledge' in English, but these two words collocate with each other in Chinese literally. Instead, the verb 'acquire' or 'gain' should co-occur with the noun 'knowledge' in English. This error type is due to "L1-L2 Interference" errors in L1 collocation which is as a result of direct translation from Chinese into English (Hou & Pramoolsook, 2012). One of the example errors occurred in this study are "…keep our living standard…" The verb 'keep' should be replaced by 'maintain' instead. (2) "Prepositional" collocation errors.

'Preposition' does not exist in Chinese. As such, it is sometimes quite difficult for Chinese ESL students to understand what preposition goes with certain words, or even if a preposition should be used after a verb and before an object. For example, "We will adhere the guidelines". The verb 'adhere' is an intransitive verb which should be followed by a preposition "to". One of the example errors occurred in this study is "Also the application of the Internet is not limited *on* browsing webpages." A preposition 'to' instead of 'on' should be used.

(3) Confusion among the usage of the verbs 'make', 'do' and 'take'

Chinese ESL learners might have often regarded that 'make', 'do' and 'take' are de-lexicalized verbs which are interchangeable. For example, "Users can make advantages of the new technology". The verb 'make' does not collocate with the noun 'advantages', and the verb 'take' should be used instead of 'make'.

Unlike 'Word-level' errors and 'Clausal-level' errors which are rule-governed, 'Collocation' errors is not rule-governed. 'Collocation' errors is referred in Ferris's studies (1999, 2006) as "an untreatable errors" alongside 'Awkwardness' errors. This means these errors of non-standard form, partially due to L1-L2 interference, cannot be explained by and identified with particular grammatical rules and specific codes. According to Ferris (1996, 1999), it is argued that students were less able to correct these 'untreatable errors'.

Both Findings I (Table 6.16 in Appendix E) and Findings II (Table 6.17 in Appendix E) suggested that indirect electronic feedback made a statistically significant difference in reducing

'Collocation' errors in both 'between groups' and 'within groups' comparison while indirect paper-based feedback did not. Hence, the pedagogical operation of 'Mark My Words' (MMWs) underpinned by Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' appears to demonstrate an absolute advantage over the paper-based feedback in facilitating students to comprehend, process and utilize indirect coded feedback in reducing 'Collocation' errors. It helps students to be more aware of how words co-occur together and equips students with the skills to search for proper word choice in order to achieve the native-like naturalness.

6.7   'Tone & Style' Errors

'Tone & Style' errors in this study refers to improper genre and/or improper level of formality in the sense that the language being used is not appropriate as far as the context, purpose and audience are concerned. For example, in academic writing, students should avoid using too subjective and extreme words such as:

(1) I think this technology must work!
(2) I am sure this is the best in the market!

Findings I ('Tone & Style' Errors between the Treatment Groups)

A Two Sample t-test was conducted between the experimental group and control group in order to find out if either group performed better than the other in reducing 'Tone & Style' errors. The summary statistics of the experiment group & control group on 'Tone & Style' errors can be seen in Table 6.18 in Appendix E.

The average number of 'Tone & Style' errors per first draft made by the experimental group and control group (between groups) was 0.7 and 0.91 respectively (Table 6.18 in Appendix E). However, the average number of 'Tone & Style' errors per second draft made by both groups was 0.033 and 0.31 respectively (Table 6.18 in Appendix). The difference in the average number of 'Tone & Style' errors respectively made by the experimental group and control group in draft 2 was relatively widened.

The statistical results (between groups) imply that the electronic feedback group and the paper-based feedback group demonstrated the same level of competence in avoiding 'Tone & Style' errors in draft 1 before treatment (Table 6.19 in Appendix E: p-value [draft 1] = 0.524), but the average number of 'Tone & Style' errors made by the electronic feedback group in draft 2 was significantly smaller than that made by the paper-based feedback group in draft 2 (Table 6.19 in Appendix E: p-value [draft 2] = 0.0572). Such difference (between groups) is educationally significant as well as being statistically significant because the effect size (= -0.485) is moderate (Table 6.19 in Appendix E).

In conclusion, the 'between groups' results indicated that students in the experimental group receiving the electronic feedback were able to make a statistically significant number of revisions in 'Tone & Style' errors; whereas students in the control group receiving paper-based feedback group did not in this 'between groups' comparison.

Findings II ('Tone & Style' Errors within the Treatment Groups)

Another two Sample t-test was conducted within each of the experimental group and control group in order to find out if each treatment group itself helped reduce the number of 'Tone & Style' errors between draft 1 and draft 2. The summary statistics of the experiment group & control group on the 'Tone & Style' errors can be seen in Table 6.18 in Appendix E.

When testing the effectiveness of each treatment group in reducing the number of 'Tone & Style' errors within its own feedback group, it is found that both paper-based feedback and electronic feedback (within its group) were able to make a statistically significant number of revisions in the average number of 'Tone & Style' errors between draft 1 and draft 2 (Table 6.20 in Appendix E: p-value [control group] = 0.0594 with an effect size = 0.483 which is moderate; Table 6.20 in Appendix E: p-value [experimental group] = 0.000460 with an effect size = 1.010 which is large). However, provided that the p-value of the electronic feedback is smaller, and its effect size is larger than that of the paper-based feedback (within its group), these "within groups" results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in remedying 'Tone & Style' errors.

<u>Discussion</u>

'Tone & Style' errors in this study refers to improper genre and/or improper level of formality in the sense that the language being used is not appropriate as far as the context, purpose and audience are concerned.

Academic writing tends to be quite formal in style. The word choice that students used in writing should reflect the level of formality in the sense that writers need to know whether the use of a particular word or phrase sound appropriate in a specific context (i.e. writing a proposal to persuade the government officials about your proposed innovation). Possible errors include 'inappropriate register', 'inappropriate tone', and 'inappropriate use of cliché'. Our markers recognized the students' credits if they had made reasonable attempt to make improvement in draft 2. The following are some common 'Tone & Style' error types made by students:

(1) Inappropriate register

For example:

Original version from a student of the experimental group in draft 1:

> What is more, with the improvement of technology, the cost of 4G could be even
> less than before.

The phrase 'what is more' is informal and conversational. The tone is inappropriate for a report proposal.

Improved version from the same student of the experimental group in draft 2:

> Moreover, with the improvement of technology, the cost of 4G could be even less than before.

The student's effort was recognized as he replaced the phrase 'what is more' by an additive connective 'moreover' which sounds more formal.

(2) Cliché

For example:

Original version from a student in draft 1

> It is crystal clear that …

The word 'crystal' is redundant and meaningless.

Improved version from the same student in draft 2:

> It is clear that …

The student's effort was recognized as the word 'crystal' is deleted.

Although Findings I showed that indirect electronic feedback made a statistically significant difference in reducing 'Tone & Style' errors while indirect paper-based feedback did not in the 'between groups' comparison, the 'within groups' comparison in Finding II showed that both indirect electronic feedback and indirect paper-based feedback made a statistically significant difference in reducing 'Tone & Style' errors within the groups themselves despite the fact that

indirect electronic feedback gave a smaller p-value and a larger effect size. These mixed results suggested that, despite varying in degrees, students' revisions on 'Tone & Style' errors were found to be responsive to both indirect paper-based feedback and indirect electronic feedback though indirect electronic feedback was found to be relatively more effective in Findings I. Hence, the pedagogical operation of 'Mark My Words' (MMWs), unlike its significant difference noted in response to the 'Awkwardness' errors and 'Collocation' errors, failed to demonstrate an absolute advantage over the paper-based feedback in facilitating students to comprehend, process and utilize indirect coded feedback in reducing 'Tone & Style' errors.

6.8 'Content' Errors

'Content' errors in this study does not refer to the right or wrong of the subject matter, data or any information provided by the students in their writings. Under this category, markers mainly evaluated the clarity and logic of the statements or arguments in student writings such as the following:

(1) whether the focus of the statement/argument is too general or unclear;

(2) whether there is any logical fallacy;

(3) whether the effect is as a result of the cause;

(4) whether the information provided is irrelevant to the argument/statement;

(5) whether there is sufficient support for any ideas presented or claims made;

(6) whether the statement/argument is ambiguous.

<u>Findings I ('Content' errors between the Treatment Groups)</u>

A Two Sample t-test was conducted between the experimental group and control group in order to find out if either group fared better than the other in reducing 'Content' errors. The summary statistics of the experiment group & control group on 'Content' errors can be seen in Table 6.21 in Appendix E.

The average number of 'Content' errors per first draft made by the experimental group and control group (between groups) was 0.9 and 1.31 respectively (Table 6.21 in Appendix E). However, the average number of 'Content' errors per second draft made by both groups was 0.1 and 0.44 respectively (Table 6.21 in Appendix E). The difference in the average number of 'Content' errors respectively made by the experimental group and control group in draft 2 was relatively widened.

The statistical results (between groups) imply that the electronic feedback group and the paper-based feedback group demonstrated the same level of competence in avoiding 'Content' errors in draft 1 before treatment (Table 6.22 in Appendix E: p-value [draft 1] = 0.237) but the average number of 'Content' errors made by the electronic feedback group in draft 2 was significantly smaller than that made by the paper-based feedback group in draft 2 (Table 6.22 in Appendix E: p-value [draft 2] = 0.0253). Such difference (between groups) is educationally significant as well as being statistically significant because the effect size (=-0.576) is moderate (Table 6.22 in Appendix E).

In conclusion, the 'between groups' results indicated that students in the experimental group receiving the electronic feedback were able to make a statistically significant number of revisions in 'Content' errors; whereas students in the control group receiving paper-based feedback group did not in this 'between groups' comparison.

Findings II ('Content' Errors within the Treatment Groups)

Another two Sample t-test was conducted within each of the experimental group and control group in order to find out if each treatment group itself helped reduce the number of 'Content' errors between draft 1 and draft 2. The summary statistics of the experiment group & control group on 'Content' errors can be seen in Table 6.21 in Appendix E.

When testing the effectiveness of each treatment group in reducing the number of 'Content' errors within its own feedback group, it is found that both paper-based feedback and electronic feedback (within its group) were able to make a statistically significant number of revisions in Content' errors between draft 1 and draft 2 (Table 6.23 in Appendix E: p-value [control group] = 0.00761 with an effect size = 0.700; Table 6.23 in Appendix E: p-value [experimental group] = 0.000439 with an effect size = 0.996). However, provided that the p-value of the electronic feedback is smaller, and its effect size is larger than that of the paper-based feedback (within its group), these 'within groups' results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in remedying 'Content' errors.

<u>Discussion</u>

The so-called 'Content' errors in this study is in fact more like suggestions for improving the clarity of any statements, arguments or propositions. Possible errors include 'unclear or too general', 'illogical', cause-and-effect problem', 'irrelevance', 'claims or ideas without adequate support', and 'ambiguity'. The markers adopted a "minimalist" approach to check if any improvement or rectification was made. That is we recognized the students' credits if they had made reasonable attempt to make improvement. The following are some common 'Content' error types made by students:

(1) Unclear or too general

For example:

Original version from a student of the experimental group in draft 1:

Secondly, the transmitting rate may reach the theoretical value.

The student did not provide any information about what the theoretical value and its bearing on the transmitting rate.

Improved version from the same student of the experimental group in draft 2:

Secondly, the transmitting rate may reach the theoretical value, especially when there are a lot of users accessing the service at the same time.

The student's effort was recognized as he had apparently made a reasonable attempt to clarify what the 'transmitting rate' is and the bearing of the 'theoretical value' on the 'transmitting rate'.

226

(2) Ineffective Introduction

For example:

Original version from a student of the experimental group in draft 1:

> Genetically modified (GM) foods discussed these days are foods that have been genetically modified by mixing the genes of animals and plants in a laboratory. GM foods always link up with the words such as "gene", "deoxyribonucleic acid" and "cloning". Genes are a segment of deoxyribonucleic acid (DNA), which records a function. For example a gene of an organism decides whether it has tail or not. Genetic modification is a process in which inherited genes of creatures and viruses are modified. GM foods mean using genetic technology to modified foods and agricultural products.

Besides missing a clear thesis statement, the above introduction failed to provide an adequate scene setting to orient audience to the background of the subject matter, that is, how GM foods improve our living quality.

Improved version from the same student of the experimental group in draft 2:

> On average, 36 million people die either directly or indirectly of hunger every year which is 58% of all deaths in 2001-2004 estimates. It is a simple calculation that 1 person dies every second owing to lack of food. 180 people die from famine, when we are waiting a cup noodle to be ready. Hopefully, food crisis can be solved easily by using genetically modified technology to improve food quality and increase the quantity of food.

The student's effort was recognized as he had apparently made a reasonable attempt to rewrite the introduction such that an adequate scene setting and thesis statement were included.

(3) Cause-and-Effect Problem (or Logical Problem)

For example:

Original version from a student of the experimental group in draft 1:

> However, misuses of the technology can lead to serious consequences. To show the power of a country, people choose to use this technology for weapon. Because of the small size of nano particles and its ability of carrying different chemical functions, bio-weapons can be made.

The misuse of the nanotechnology seems to depend more on if the technology falls on the ethical hands of people rather than how small the size of nano particles is. The student made no change to the original version so the error count remained.

Although Findings I showed that indirect electronic feedback made a statistically significant difference in reducing 'Content' errors while indirect paper-based feedback did not in the 'between groups' comparison, the 'within groups' comparison in Finding II showed that both indirect electronic feedback and indirect paper-based feedback made a statistically significant difference in reducing 'Content' errors within the groups themselves despite the fact that indirect electronic feedback gave a smaller p-value and larger effect size. These mixed results suggested that, despite varying in degrees, students' revisions of 'Content' errors was found to be responsive to both indirect paper-based feedback and indirect electronic feedback though indirect electronic feedback was found to be relatively more effective in Findings I. Hence, the pedagogical operation of 'Mark My Words' (MMWs), unlike its significant difference noted in response to 'Awkwardness' errors and 'Collocation' errors, failed to demonstrate an absolute advantage over the paper-based

feedback in facilitating students to comprehend, process and utilize indirect coded feedback in reducing 'Content' errors.

6.9   *'Organization' Errors*

'Organization' errors in this study refers to the problems about cohesion and coherence of the text. For instance, how logically ideas / propositions are connected within a paragraph or across paragraph; do the 'introduction' and 'conclusion' serve their purpose effectively; is the line of reasoning easy to follow; do the supporting details follow the topic sentence, etc. More specifically, possible problems related to 'Organization' include any of the following:

(1) One or more required moves or sections (i.e. standard information) is missing;

(2) Lack of focus in the paragraph;

(3) Lack of transitions within the paragraph;

(4) Unclear / Inadequate topic sentence;

(5) Supporting details do not follow the topic sentence;

(6) Claims or Ideas without adequate support;

(7) Ideas are choppy or jumpy.

(8) Lines of reasoning difficult to follow

(9) New Paragraph / Sentence

Findings I ('Organization' Errors between Treatment Groups)

A Two Sample t-test was conducted between the experimental group and control group in order to find out if either group fared better than the other in reducing 'Organization' errors. The summary statistics of the experiment group & control group on 'Organization' errors can be seen in Table 6.24 in Appendix E.

The average number of 'Organization' errors per first draft made by the experimental group and control group (between groups) was 2.43 and 0.69 respectively (Table 6.24 in Appendix E) However, the average number of 'Organization' errors per second draft made by both groups was 0.23 and 0.28 respectively (Table 6.24 in Appendix E). The difference in the average number of 'Organization' errors respectively made by the experimental group and control group in draft 2 was relatively narrowed.

The statistical results (between groups) imply that the paper-based feedback group initially performed better than the electronic feedback group in avoiding 'Organization' errors in draft 1 (Table 6.25 in Appendix E: p-value [draft 1] = 0.000322). Such difference (between groups) is educationally significant as well as being statistically significant because the effect size (= 1.0364) is large (Table 6.25 in Appendix E). However, there is no significant difference between the average number of 'Organization' errors made by the electronic feedback group in draft 2 and that made by the paper-based group in draft 2. (Table 6.25 in Appendix E: p-value [draft 2] = 0.730). This implies that that the electronic feedback group and the paper-based feedback group demonstrated the same level of competence in avoiding 'Organization' errors in draft 2 after treatment. This signifies a relative improvement for the electronic feedback group.

In conclusion, the 'between groups' results indicated that students in the experimental group receiving the electronic feedback were able to make a statistically significant number of revisions in 'Organization' errors.

Findings II ('Organization' Errors within Treatment Groups)

Another two Sample t-test was conducted within each of the experimental group and control group in order to find out if each treatment group itself helped reduce the number of 'Organization' errors between draft 1 and draft 2. The summary statistics of the experiment group & control group on 'Organization' errors can be seen in Table 6.24 in Appendix E.

When testing the effectiveness of each treatment group in reducing the number of 'Organization' errors within its own feedback group, it is found that both paper-based feedback and electronic feedback (within its group) were able to make a statistically significant number of revisions in 'Organization' errors between draft 1 and draft 2 (Table 6.26 in Appendix E: p-value [control group] = 0.0261 with an effect size = 0.571; Table 6.26 in Appendix E: p-value [experiment group] = 1.181e-05 with an effect size = 1.338). However, provided that the p-value of the electronic feedback is smaller and its effect size is larger than that of the paper-based feedback (within its group), these 'within group' results further support the conclusion that the electronic feedback treatment is more effective than the paper-based feedback treatment in remedying 'Organization' errors.

<u>Discussion</u>

The so-called 'Organization' errors in this study are in fact more like suggestions for improving the coherence and cohesion of the paragraphs. Achieving coherence and cohesion can help guide readers moving from one sentence (idea) to the next sentence (idea) without efforts, with a goal of getting across the intended messages effectively and facilitating readers' train of thought in understanding the logic.

Possible errors include 'missing a particular section', 'lack of focus in the paragraph', 'lack of transitions within the paragraph', 'unclear/inadequate topic sentence', 'supporting details do not follow the topic sentence', 'claims or ideas without adequate support', 'ideas are choppy or jumpy', 'lines of reasoning difficult to follow', and 'new paragraph / sentence'. The markers adopted a "minimalist" approach to check if any improvement or rectification was made. That is we recognized the students' credits if they had made reasonable attempt to make improvement in draft 2. The following are some common 'Organization' error types made by students:

(1) Ideas are choppy / Lines of reasoning is difficult to follow

For example:

Original version from a student in draft 1:

> Although the GM food still has some unsolved and unpredictable problems, human can still have a gain on it. The plant can be greatly reproduced by just changing the DNA. The crop yield can be greatly increased.

The development of ideas from the sentence 1 to sentence 2 appears disconnected. This might have undermined readers' train of thought.

Improved version from the same student in draft 2:

> Although the GM food may still have some unsolved and unpredictable problems, its benefits outweigh its perils. With its modification of its DNA (<-putting the 'old info' / 'cause' first), plants can be grew more healthily (<- putting the result after), yielding a higher production." (<- using a participle phrase to illustrate the result)

The student's effort was recognized as he had apparently made a reasonable attempt to develop the connectivity and logical relationship of ideas by applying the theme-rheme principle and sentence variety (i.e. the present participle).

(2) Lack of transitions within the paragraph

For example:

Original version from a student of the experimental group in draft 1:

> Using genetically modified technology can increase the pest resistance of plants. By importing pest resistant genes, the modified plant can resist certain pests and farmers can use fewer pesticides. It can also reduce the pollution and side effect caused by pesticides.
>
> The productivity can be increased by adding genes which can increase the release of growth hormone in poultry and fish in order to increase their growth rate and shorten their consumable time.

> Genetic modification can improve the growth in unfavourable environments. For example applying drought resistance genes to tomatoes makes it allows growing in dry areas.
>
> The quality of food can be improved by GM technology for example improving nutrients in food and removing allergens in food such as crab, crabs and peanuts.

The ideas of the original paragraph are exhibited in chunks with little connectivity.

Improved version from the same student of the experimental group in draft 2:

> By the technology of importing pest resistant genes, the modified plant can resist certain pests and farmers can use fewer pesticides. It can also reduce the pollution and side effect caused by pesticides. The food quality is improved as fewer pesticides residue. More healthy food can be produced so the amount of food can be increased.
>
> There are other methods to increase the amount of food. One is adding genes which can increase the release of growth hormone in poultry and fish in order to increase their growth rate and shorten their consumable time. Another one is adding genes to improve the growth in unfavourable environments. For example applying drought resistance genes to tomatoes makes it grow in dry areas.

The student's effort was recognized as he had apparently made an attempt to join the ideas by using some discourse markers like 'there are other methods to increase the amount of food', 'one is …', ad 'another one is…'.

In the 'between groups' comparison, Findings I showed that the control group was originally doing better in avoiding 'Organization' errors in draft 1. It was until draft 2 in which an intervention had taken place that there is no significant difference between the average number of 'Organization'

errors made by the electronic feedback group in draft 2 and that made by the paper-based group in draft 2. In other words, the intervention of the indirect electronic feedback in draft 1 had put the experimental group on par with the control group in avoiding 'Organization' errors in draft 2. On the other hand, the 'within group' comparison in Finding II showed that both indirect electronic feedback and indirect paper-based feedback made a statistically significant difference in reducing 'Organization' errors within the groups themselves despite the fact that indirect electronic feedback gave a smaller p-value and larger effect size. These mixed results suggested that, despite varying in degrees, students' revisions on 'Organization' errors were found to be responsive to both indirect paper-based feedback and indirect electronic feedback, though indirect electronic feedback was found to be relatively more effective in Findings II. Hence, the pedagogical operation of 'Mark My Words' ('MMWs'), unlike its significant difference noted in response to 'Awkwardness' errors and 'Collocation' errors, failed to demonstrate an absolute advantage over the paper-based feedback in facilitating students to comprehend, process and utilize indirect coded feedback in reducing 'Organization' errors.

All in all, the statistical results in Appendix E suggest that the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy and Krashen's (1985) 'Input Hypothesis' is relatively more effective in error reduction across all error categories, namely (1) 'Awkwardness' errors, (2) 'Clausal-level' errors, (3) 'Word-level' errors, (4) 'Collocation' errors, (5) 'Tone & Style' errors, (6) 'Content' errors, and (7) 'Organization' errors. In particular, student uptake of the electronic feedback on 'Awkwardness' errors and 'Collocation' errors was very prominent in the sense that the paper-based feedback failed to make any statistically significant number of reduction in these two categories within its own group at all. These results

coincide with Ferris's (2006) study that students found non-rule governed errors which cannot be explained by and associated with any grammatical rules particularly difficult to comprehend and correct, possibly due to 'L1-L2 Interference'.

Chapter 7   Findings and Discussions: Students' Perceptions on Teacher Written Feedback

Even though students have a very central role to play in the feedback process, much of the research on teacher written feedback has been focusing on the teachers' perspective. For example, the strategies teachers use in giving feedback, their stances towards error correction and the impact of their feedback on student writing (e.g. Ferris, 1997; Ferris, Pezone, Tade, & Tinti, 1997; Hyland & Hyland, 2001; Stern & Solomon, 2006, Lee, 2008b). However, rather than being a mere recipient under the feedback process (Lee, 2008b), they should be viewed as active and proactive agents (Hyland & Hyland 2006a). Hence, understanding how students respond to and act on teacher written feedback is essential to understand how students perceive what would constitute a good feedback practice, which in turns allows ESL teachers better understand how they may adjust their feedback in order to not only to reduce errors but also to engender positive learning responses.

A student questionnaire (see Appendix D) was administered to all 62 students receiving the electronic feedback and paper-based feedback after they had submitted their final drafts. The questionnaire was modeled on the one primarily developed by Lee (2008b) with more detailed modification on the number of items given for participants' selection. The table below (see *Table 7.1*) shows the terms which were used in the questionnaire:

Table 7.1 Terms used in the Questionnaire

	Terms	Remarks
1.	Error Feedback	Error identified + Error symbol assigned
2.	Written Comments	Overall comments on Content, Organization & Language
3.	Grades	Good ← Satisfactory → Inadequate

The following (see *Table 7.2*) lists out the categories of question fielded in the questionnaire alongside the summary of findings:

Table 7.2 Summary of Findings

	Category	Section #	Question #	Findings
1.	Comprehensibility of teacher feedback	7.1	Q1	Experimental group better comprehended
2.	Ability of error correction	7.2	Q2	Experimental group more able to correct
3.	Types of feedback preferred	7.3	Q3 & Q6	Comments > Error feedback > Grade
4.	Preference between Error Feedback and Teacher Commentary	7.4	Q4-Q5	Teacher commentary > Error feedback
5.	Focus of error feedback preferred	7.5	Q7-Q8	Language > Organization > Content
6.	The Amount of error indicated	7.6	Q9	All
7.	Focus of written comments preferred	7.7	Q10-Q11	Organization > Language > Content
8.	Error Feedback strategies preferred	7.8	Q12	At least (1) Error identification; (2) Error categorization and; (3) Error explanation
9.	Student autonomy on explicitness of error feedback	7.9	Q13	Feedback autonomy preferred
10.	Comprehensibility of Error Types	7.10	Q14-Q15	Awkwardness errors being the most difficult to understand; Word-level error the easiest
11.	Errors Types being Most Difficult for Correction	7.11	Q16-Q17	Awkwardness errors being the most difficult to understand; Word-level errors being the easiest

12.	Post-writing activities preferred after receiving the marked essay	7.12	Q18	Asking teachers for clarifications & explanation, and error correction are most preferred
13.	Teacher written feedback helps with students' long term writing	7.13	Q19	Over 80% (experimental group) – very helpful vs. over 70% (control group) – very helpful

The results of the questionnaires were not only compared between the experimental group and the control group, they were also compared and reconciled with the corresponding statistical results of this study and/or with the results of some previous studies, if applicable. It is hoped that the qualitative part of the study can supplement the explanation of the statistical results with a greater depth and provide us more insights about students' attitudes towards teacher writing feedback, despite the fact that the quantitative part remains the backbone of the analysis in this study.

7.1 *Comprehensibility of Teacher Feedback (Q1)*

The findings in *Table 7.3* show that 21% of the participants in the experimental group indicated that they totally understood their instructor's feedback whereas only 7% of the participants in the control group expressed that they totally understood their instructor's feedback. 72% of the participants in the experimental group expressed that they partially understood their instructor's feedback whereas 73% of the participants in the control group expressed the same.

Table 7.3: Q1 - Was your language instructor's feedback understandable?

	Experimental Group	Control Group
1 - Not understandable	0%	0%
2	0%	3%
3 - roughly half-half	7%	13%
4	72%	73%
5- Totally understandable	21%	7%

The result in *Table 7.3* indicates that the experimental feedback group receiving the electronic feedback was relatively more able to understand their teacher written feedback. This matches the statistical results of the overall error reduction (see Appendix E and Appendix G) which also conclude that electronic feedback was performing relatively better in reducing errors. Most comments left by the participants in the experimental group have confirmed these positive findings of the electronic feedback in terms of its comprehensibility. Among those are "As there are Google search link, it is easier to follow", "The feedback is clear and understandable", "detailed explanation and provide a Google link for us to go for and know more about the error, "yes, [the teacher] can give clear explanation on error and it is clearly that [the teacher] has paid great effort on the comments to us", "can be read easily", "good to have Google search", "supported with explanation on error and example can make me understand more", "Google search is especially useful", "quite easy to understand the problems I made. But sometimes it is not very clear why it is wrong" and "That's clear and understandable. I can easily recognize the mistake". When compared to the comments left by the participants in the experimental group, fewer positive comments were made by the control group in terms of the comprehensibility of the paper-based feedback. Among those are "The instructor explained clearly", "The instructor tried to explain clearly, "Yes, I think this is acceptable and suitable for us to learn", "The teacher commented the wrong verbs/phrases in detail", "Quite good".

Instead of comparing the quantity of positive comments respectively made by both groups, it is found that the positive comments made by the participants in the electronic feedback carried much more details such that the students were able to point out the features of the electronic feedback (e.g. 'Google Link' or 'Google Search'), explained how particular features helped them identify and correct errors (e.g. "detailed explanation"; "supported with explanation on error and example can make me understand more"), and how they benefit from the teacher feedback (e.g. "know more about errors", "can be read easily", "quite easy to understand the problem I made" and "I can recognize the mistake") . However, the positive comments made by the participants receiving paper-based feedback were relatively simple and general without insights. This may imply that hardly could the participants in the control group identify any peculiar merits from the papered-based feedback.

7.2   *Ability of Error Correction (Q2)*

The findings in *Table 7.4* show that 17% of the participants in the experimental group indicated that they were totally able to correct the error accurately according to their instructor's feedback whereas only 7% of the participants in the control group expressed that they were totally able to act on the feedback. 62% of the participants in the experimental group expressed that they were quite able to correct the error whereas 53% of the participants in the control group expressed that they were quite able to do so.

Table 7.4: Q2 - To what extent were you able to correct the error accurately according to the language instructor's feedback?

	Experimental Group	Control Group
1 - Not able	0%	3%
2 – Quite unable	0%	3%
3 – Sometimes able	21%	37%
4 – Quite able	62%	53%
5- Totally able	17%	7%

With very close to 80% of the participants from the experimental group reporting that they were totally able or quite able to correct teacher written feedback as comparing with 60% from the control group, the result in *Table 7.4* indicates that the experimental feedback group receiving the electronic feedback felt that they were relatively more able to process, utilize and act on teacher written feedback. This matches the statistical results of the overall error reduction (see Appendix E and Appendix G) which also conclude that the electronic feedback was performing relatively better in reducing errors. About 50% of the comments left by the participants in the experimental group have confirmed these positive findings of the electronic feedback in terms of its effectiveness in helping them correct errors. Among these positive comments are "For awkward/unclear expressions, I can make corrections due to the examples the (feedback system) given", "grammatical mistakes are more easily to be corrected according to the feedback", "the google search is useful to correct the errors but afraid student may resist to learn as the answer is provided", and "[it] is easy to correct especially those with google search". However, the effectiveness of the electronic feedback is not omnipotent and absolute. This is reflected from few comments left by students from the experimental group which said "For most errors I made there are clear instructions of how to correct the errors. But sometimes I can only know what is the problem and have no idea how to correct it", "there are some comments that I can't understand,

run-on sentence", "some errors about the organization are difficult to correct", and "I can understand almost all the errors, and know how to correct it, except some problem[s] about the organization, or content". These comments may not seem to reject the effectiveness of the electronic feedback as participants concerned still indicated that they were just sometimes not able to correct most errors, where errors related to 'content' and 'organization' were highlighted the most.

When compared with the comments left by the participants in the experimental group, 70% of the comments left by the participants in the control group were negative in terms of the effectiveness of the paper-based feedback in helping students correct errors. Among these negative comments are "It is especially hard for the error such as awkward (expression)', 'For the category of awkward, I usually do not know how to correct", "some vocabulary and phrases (are) hard to correct on my own", I fully understood his feedback only after personally asking him about his comments", and "Some of the errors are very common for me and I am not able to correct it".

Instead of comparing the quantity of positive and negative comments respectively made by both groups, it is also very worthwhile to scrutinize the nature and dimensions of these qualitative comments made by students and then reconcile them with the quantitative data collected in this study and other studies, despite the fact that the majority of them from both groups claimed that they are at least quite able to correct the identified errors. On the one hand, the qualitative comments from the experimental group suggest that the electronic feedback may not seem as equally effective as helping student correct grammatical errors when it comes to deal with 'Content' errors and 'Organization' errors. These qualitative findings to a certain degree lend

support to the statistical findings that the difference between the experimental group and the control group in improving 'Content' errors in the final draft is just marginally significant (Table 6.22 in Appendix E: p-value = 0.0253), whereas there is no significant difference between the two groups in improving the 'Organization' errors in the final draft (Table 6.25 in Appendix E: p-value = 0.730). On the other hand, the qualitative comments from the control group suggest that the paper-based feedback may seem less effective in helping students correct 'Awkwardness' errors and 'Collocation' errors. This is supported by the statistical findings that the experimental group receiving the electronic feedback were much more able to make statistically significant number of revisions in 'Awkwardness' errors (Table 6.7 in Appendix E: p-value = 0.000149) and 'Collocation' errors (Table 6.16 in Appendix E: p-value = 0.000271) in the final draft. Also, the 'within groups' statistical results suggest that the paper-based feedback within its own group made no statistically significant number of error reduction in 'Awkwardness' Errors (Table 6.8 in Appendix E: p-value = 0.227) and 'Collocation' Errors (Table 6.17 in Appendix E: p-value = 0.8464). Hence, both statistical results and qualitative comments from the control group lend support to each other.

Both quantitative and qualitative results were triangulated with Ferris's observation in her studies (1999, 2006) that 'sentence-level' errors (which is equivalent to 'Awkwardness' errors in this study) and 'word choice' problem (which is equivalent to 'Collocation' errors in this study) are untreatable errors which are relatively more difficult for ESL learners to correct.

The findings in *Table 7.5* show that as much as 88% of the participants in the experimental group and 77% of the participants in the control group indicated that they would like to receive 'Grade + Error Feedback + Written Comments' on their writings.

Table 7.5: Q3 - In future compositions, which of the following combinations of feedback do you prefer to receive from the language instructor?

		Experimental Group	Control Group
	Grade + Error Feedback + Written Comments	83%	77%
	Error Feedback + Written Comments	10%	3%
	Grade + Written Comments	3%	3%
	Grade + Error Feedback	3%	23%
	Only Written Comments	0%	0%
	Only Error Feedback	0%	3%
	Only Grade	0%	0%

The following are all qualitative comments gathered from Q3 which are categorized by the emphasis respectively placed by the respondents from the experimental group and the control group:

(1) Grade, Written Comments & Error Feedback are all important	
Experimental Group (83% preferred)	Control Group (77% preferred)
a. "Grades help me to understand my language level within my class. Error feedback let me know what mistakes (use of words / grammar) I made. Written comments let me know how can I improve my writing when I need to write the same kind of writing in the future. b. "These three are important to me."	a. "The grade is to know our overall performance. Error feedback and written comments are also important for us to fully understand what we have done wrong. It's not like math, where there's only one correct solution and no ambiguity. Plus, sometimes the instructor might make mistakes also, and if there are no written comments, the student will keep on wondering where he has gone wrong."

	b. "I know giving all (grade/error feedback/comment) is time consuming, but students can learn much from that."

(2) Written Comments & Error Feedback are important

Experimental Group (10% preferred)	Control Group (3% preferred)
a. "Comments are very important since it can let student(s) know what the instructor think(s) about their article(s). Students know different kind(s) of feedback including grammar and comments. b. 'I prefer that my writing should be well examined so that I can learn more during in that writing. Comments and error feedback would provide a comprehensive analysis on my writing and therefore (this is a) better way to learn." c. "I think the error feedback and comments are needed so we can understand what we get wrong and know how to correct.	a. "More feedback and comments can improve my writing."

(3) Grade + Written Comments are important

Experimental Group (3% preferred)	Control Group (3% preferred)
NIL	NIL

(4) Grade + Error Feedback are important

Experimental Group (3% preferred)	Control Group (23% preferred)
a. "Grade can help me know more about my language standard with comparing with the other students. Certainly, error is useful for understanding more about what I made a mistakes."	NIL

(5) Written Comments are more important

Experimental Group (0% preferred)	Control Group (0% preferred)
a. "If my work is given with grade, it is also fine with me, but I don't really care about the grade. To reduce unnecessary work for the instructor and to have more time to write comments on my errors, I suggest without grading."	a. "Written comments (are) important to improve the content (of) the composition." b. More information provided for my improvement of composition can let me comprehend mistakes in details. Especially written comments can improve not only

(6) Only Error Feedback is important	
Experimental Group (0% preferred)	Control Group (0% preferred)
NIL	*NIL*

(7) Only Grades	
Experimental Group (0% preferred)	Control Group (3% preferred)
NIL	*NIL*

The above qualitative comments in general reflect that students regarded that the grade is to allow them to understand their overall performance relative to others, error feedback allows them to understand and correct the identified errors, and written comments give them an idea how to improve the overall writing. Holistically speaking, participants irrespective of groups prefer to receive as much feedback as possible, and students generally appreciated the efforts and hard work made by their teachers. This also lends support to some previous studies which concluded that students expected feedback on their written errors from their teachers (Ferris, 1995; Ferris et al., 2000; Hedgecock & Lefkowitz, 1994; Leki, 1991) and 'no student said they did not want errors corrected by their teacher' (Ferris, 200, p.177). Another study also concluded that 'feedback provided by the teacher is still highly valued by the student-writers because of his/her traditional role as "evaluator"' (Enginarlar, 2001, p.7). To further understand the participants' preference towards each type of feedback, Q6 (see *Table 7.6*) in the questionnaire was fielded:

Table 7.6: Q6 - In future compositions, which of the following types of feedback would you be most interested in receiving from the language instructor?

		Experimental Group	Control Group
	Teacher's comments on my writing	62%	47%
	The errors I have made	34%	33%
	The grade	0%	20%
	Others	3%	0%

The results in *Table 7.6* show that the majority of participants in both groups seem to attach more importance to receiving written comments, and then followed by error feedback and the grade. This is evident from the qualitative comments given by participants from both groups. The majority of the respondents expressed their preference on teacher's comments: "Teacher's comments are helpful in the overall writing style and organization", "I would like to know how to write with good structure, e.g. how to plan, how to organize, etc.", "It would be helpful if the teacher can give more comments on how to improve", "Teacher's comments on my writing can help me do some overall improvement of my writing skills", "Teacher's comments can directly reflect the capability of the students, and thus they can review their corresponding weaknesses", and "It is good to justify my grade with more qualitative comments".

7.4 *Preference between Error Feedback and Teacher Commentary (Q4 & Q5)*

Q4 & Q5 of the questionnaire were fielded to narrow down the focus where only 'Error Feedback' and 'Written Comments' (and 'None') were made available for the respondents' selection. Question 4 asked "Which of the following types of feedback would you like your language instructor to give more in future?", and Question 5 asked "Which of the following types of feedback would you like your instructor to give less in future?"

The findings in *Table 7.7* (Q4) show that as much as 66% of the participants in the experimental group and 63% of the participants in the control group indicated that they would like to receive 'Written Comments', whereas 31% of the participants in the experiment group and 27% of the participants in the control group chose "Error Feedback". 3% of the participants in the

experimental group and 10% of the participants in the control group expressed that they would not like to receive any 'Written Comments' and 'Error Feedback'. The findings in *Table 7.8* (Q5) show that close to 100% of the participants from both groups expressed that they did not want to receive any less from their language instructors.

Table 7.7: Q4 – Which of the following types of feedback would you like your language instructor to give more in future?

		Experimental Group	Control Group
	Written Comments	66%	63%
	Error Feedback	31%	27%
	None of the above	3%	1%

Table 7.8: Q5 – Which of the following types of feedback would you like your language instructor to give less in future?

		Experimental Group	Control Group
	Written Comments	3%	0%
	Error Feedback	0%	3%
	None of the above	97%	97%

The findings from Q4 and Q5 coincide with those from Q3 and Q6 of the questionnaire (see Appendix D) in which the order of student preference on the three types of feedback is 'Written Comments", and then followed by "Error Feedback" and "Grade". These qualitative results agree with the study conducted by Lee (2008b) that students, irrespective of their proficiency level, wanted more written comments from teachers when comparing to the amount of errors they wished to be pointed out. However, a very high proportion of teacher written feedback focused on error feedback (Lee, 2008a, 2008b). This implies that students would like to receive more qualitative comments about what their teachers thought and felt about their writing holistically than seeing their texts awash in red ink of error identification and correction (Lee, 2008b). The following are the comments respectively made by both groups with respect to Q4 and Q5. Student preference

over 'Written Comments" is evident by the relatively higher number count of comments related to the importance of receiving written comments emphasized by participants from both groups. In summary, students generally believed that written comments can better help improve the quality of their writing style and content more holistically for future writing rather than error feedback which mainly focuses on error correction.

(1) Written Comments are preferred
Experimental Group
a.   "Written comments are, comparatively, less than, error feedback. It is hoped that I can improve the writing style other than errors."   b.   "I think written comments are more important because they indicate the impression of my writing from others."   c.   "My suggestion is to write the comment right next to the error. It is much better for us to understand and can save time. We just need to leave the right half of our submitted work blank for more spaces."   d.   "Error feedback only gives the explanation of the error I made; written comments can further improve my writing style and content."   e.   "I think most errors are based on careless and grammar mistakes, so the comments are useful for us."   f.   "Written comments give an overall view of the whole passage."   g.   "Written comments are more important that they can help me improve not only about the error, but also the structure of the whole article."
Control Group
a. "Too little written comments is difficult to improve the content. Why not error feedback is that I can still figure it out y myself by revising it more times."   b. "In my opinion, content is more important for a composition. Written comments can focus more on content."   c. "Giving comments to us to improve."   d. "Written comments help me a lot for future writing."

(2) Error Feedback is preferred
Experimental Group
a. "I choose 'error feedback' because I think 'written comments', overall comments on content / organization/ language, sometimes cannot help me to indicate the errors in the essay."   b. "To correct the fatal mistakes or wrong sentences."   c. It depends on whether the writing has problems on grammar and sentence structure mistakes. If it has, the error feedback should be given more. Otherwise, instructors should focus on written comments."
Control Group
NIL

7.5   *Preference on the Focus of Error Feedback (Q7 & Q8)*

With the primary attention given to 'Error Feedback' alone, Q7 & Q8 in the questionnaire (see Appendix D) aimed to find out what aspects students would like language instructors to focus on more in error feedback. Question 7 asked "Which of the following focus of error feedback would you like your instructor to give more in future?", and Question 8 asked "Which of the following focus of error feedback would you like your instructor to give less in future?"

The findings in *Table 7.9* (Q7) show that 41% of the participants in the experimental group and 60% of the participants in the control group indicated that they would like their instructors to emphasize more on 'Language' in future. 41% of the participants in the experiment group and 27% of the participants in the control group preferred 'Organization'. 14% of the participants from the experimental group and 13% of the participants from the control group chose 'Content'. The findings in *Table 7.10* (Q8) show that about 60% of the participants from the experimental and control groups expressed that they did not want instructors to attach less emphasis on any items in future. About 24% of the participants from both groups expressed that instructors could place less emphasis on 'Content' in future.

Table 7.9: Q7 – Which of the following focus of error feedback would you like your language instructor to emphasize more in future?

		Experimental Group	Control Group
	Language (e.g. grammar, vocab.)	41%	60%
	Organization (e.g. connectivity)	41%	27%
	Content	14%	13%
	None of the above	3%	0%

251

Table 7.10: Q8 – Which of the following focus of error feedback would you like your language instructor to emphasize less in future?

		Experimental Group	Control Group
	Language (e.g. grammar, vocab.)	7%	10%
	Organization (e.g. connectivity)	3%	3%
	Content	24%	23%
	None of the above	63%	60%

Scrutinizing the qualitative results from Q7 and Q8 together (see Appendix D), it is found that participants from both groups preferred their instructors to put more emphasis on 'Language' in their error feedback, and then it is followed by "Organization" and "Content". Apparently, students from the control group seem to attach more importance to 'Language' when comparing to those from the experimental group. Such disparity might have been explained by both statistical results (see Chapter 6) and the qualitative results (Q1 & Q2 in Appendix D) that students from the control group were relatively less able to comprehend, process and utilize their teacher feedback in responding to their errors. Thus they might feel more immediate needs from their instructors to offer more help on the language aspect. Another observation noted is that respondents from both groups gave the least attention to 'Content' as some students expressed that teachers may not have subject knowledge (i.e. engineering discipline) to help them with their content. The following are the qualitative comments made by students:

Focus of Error Feedback Preferred More
Experimental Group
a. "It is very important to know how to do the organization better."
b. "Organization is most important for others to understand my writing."
c. "Introduce more phrases and idioms will make our writing become sophisticated and natural."
d. "It is the most difficult part in a writing and rather need more time to learn."
e. "It is appropriate for the material that I choose."
f. "Organization is one of the weakest part in my article."
g. "Because the organization of a paragraph determines whether it is understandable or not, and whether it can be easy followed, which I think is most important for a compensation."
h. "There is lack of feedback about organization."
Control Group
a. "Since English is not my mother language, I cannot accurately use the words."
i. "Maybe this is because my organization skill is not good enough. More comments may be needed for my further improvement."
j. "It is useful for us to improve in the future."
k. "I think the most important part of writing is the flow of ideas."
l. "It is the problems that all students have."
m. "Language is the part that I want the instructor to focus on, but organization is also important."

Focus of Error Feedback Preferred Less
Experimental Group
a. "I think any comments would be useful as they can help improve my writing."
b. "Content I not very useful considering engineering composition. Teacher only has little idea about the content of the engineering composition."
c. "No need to write comments for every missing 's' or simple present / past tense errors. Just circle the error for identification is enough."
d. "Content may be the main part of the writing."
e. "All are useful items."
f. "All are important to me."
g. "Good to include more information."
Control Group
a. "All kinds of feedback are usually good for improving."
b. "Error categorization, because I don't remember the English language by categories."
c. "It is not really useful as we have enough technical knowledge already."
d. "Students should know how to build up the content."

7.6 Preference on the Amount of Errors Indicated (Q9)

Q9 of the questionnaire (see Appendix D) asked the students about the amount of errors they would like their instructor to respond to. The options are: "All errors", "Some errors" or "None".

The findings in *Table 7.11* (Q9) show that as much as 76% of the participants in the experimental group and 63% of the participants in the control group indicated that they would like to their instructors to respond to all errors, whereas only 21% and 33% respectively from the experimental group and control group expressed that they would like their instructors to respond to some errors only.

Table 7.11: Q9 – Click ONE option below to indicate the amount of error you would like your language instructor to respond to.

	Experimental Group	Control Group
All	76%	63%
Some	21%	33%
None	0%	3%

The results in *Table 7.11* show that the majority of participants in both groups would like their instructors to identify and explain all their errors. This is evident from the qualitative comments given by participants from the experimental groups such as "I think comments would be useful as they can help improve my writing", "As I would not know the errors that I have made, I will be fully understanding to the error given by tutors", "know more about what error I have made" . Despite the fact that over 60% of the respondents from the control group also indicated that they preferred to have all errors to be addressed, almost one-third of them expressed that they would like their instructors to address their errors selectively. The qualitative comments given by the participants of the control group might shed light on what they meant by addressing some errors

but not all. Their comments included "as some of the errors are careless mistakes", "some errors are obvious. These are not needed to be explained too much", Maybe there is some mistakes, but it is not frequent", and "As the error can be repeated many times in the same article". In summary, some students believed that some errors were probably repetitive and merely due to carelessness which could be corrected easily, so that these errors might not be worth instructors' effort to respond so often.

7.7   *Preference on the Focus of Teacher Commentary (Q10 & Q11)*

Q10 & Q11 of the questionnaire (see Appendix D) were fielded to find out what focus of written comments students would like their instructors to emphasize more (Q10) and less (Q11) in future. The available options for respondents' selection were 'Content', 'Organization', 'Language', and 'None of the above'.

The findings in *Table 7.12* (Q10) show that 45% of the participants in the experimental group and 37% of the participants in the control group indicated that they would like their instructors to emphasize more on 'Organization'. Around 30% of the participants from both groups indicated that they would like their instructors to emphasize more on 'Language', and 17% and 23% of the participants respectively from the experimental and control groups indicated that they would like to receive more comments on 'Content'. The findings in *Table 7.13* (Q11) show that about close to 70% of the participants from the experimental group and 60% of the participants from control group expressed that they did not want instructors to attach less emphasis on any items in future.

About 20% of the participants from both groups expressed that instructors could place less emphasis on 'Content' in future.

Table 7.12: Q10 – Which of the following focus of written comments would you like your language instructor to emphasize more in future?

		Experimental Group	Control Group
	Organization (e.g. flow of ideas)	45%	37%
	Language (e.g. grammar)	31%	33%
	Content	17%	23%
	None of the above	6%	3%

Table 7.13: Q11 – Which of the following focus of written comments would you like your language instructor to emphasize less in future?

		Experimental Group	Control Group
	Organization (e.g. flow of ideas)	7%	7%
	Language (e.g. grammar)	7%	17%
	Content	17%	20%
	None of the above	69%	57%

Scrutinizing the qualitative results from Q10 and Q11 together (see Appendix D), it is found that participants from both groups preferred their instructors to put more emphasis on 'Organization' in their written comments, and then it is followed by "Language" and "Content". One notable difference is identified when comparing the focus the majority of students from both groups preferred in written comments (i.e. 'Organization' in Q10 & Q 11) with the focus they preferred in error feedback (i.e. 'Language' Q7 & Q8). It is observed that there seems to be an association between students' preferred focus and the type of feedback they received. The majority of the respondents would like their instructors to give more attention to 'Language' in error feedback while they preferred more attention to 'Organization' in written comments. Such difference could have been attributed to the reason that written comments in sentences may provide better and

clearer explanation in logical flow and connectivity, while error feedback my serve better in

labeling grammatical errors. The following are the qualitative comments made by students:

Focus of Written Comments Preferred More
Experimental Group
a. "Organization is most important for others to understand my writing."   b. "It is the most difficult part in writing and rather need more time to learn."   c. "All are important."   d. "Organization can help to improve our writing very much."
Control Group
a. "I think it is hard to have good content in my composition."   b. "This is the area where I would like to improve."   c. "The grammar and sentence pattern information is useful for us to improve in the future."

Focus of Written Comments Preferred less
Experimental Group
a. "I think any comments would be useful as they can help improve my writing."   b. "All are useful items."   c. "Because content's comments are little subjective."   d. "All more important."   e. "The more the better."
Control Group
a. "Everyone can have their own style, except their compositions are really messy."   b. "It is not really useful as we have enough technical knowledge already."

7.8 Preference on Error Feedback Strategies (Q12)

Q12 of the questionnaire (see Appendix D) was fielded to find out what error feedback strategies

students would like their instructors to use more in future. The available options for respondents'

selection alongside the findings are shown in *Table 7.14* below.

Table 7.14: Q12 – Which of the following error feedback strategies would you like your language instructor to use more in future?

	Feedback Strategies	Experimental	Control
1.	Underline/Circle errors	0%	0%
2.	Underline / Circle errors + Categorize errors	0%	13%
3.	Underline/Circle errors + Provide correction	3%	10%
4.	Underline/Circle errors + Provide examples	3%	3%
5.	Underline/Circle errors + Categorize errors + Explain errors	7%	30%
6.	Underline/Circle errors + Categorize errors + Provide correction	0%	0%
7.	Underline/Circle errors + Categorize errors + Provide examples	7%	0%
8.	Underline/Circle errors + Categorize errors + Explain errors + Provide correction	13%	13%
9.	Underline/Circle errors + Categorize errors + Explain errors + Provide examples	52%	17%
10.	Underline/Circle errors + Categorize errors + Explain errors + Provide examples + Provide correction	17%	13%
11.	None of the above	0%	0%

The findings in *Table 7.14* (Q12) show that 52% and 17% of the participants in the experimental group preferred receiving Option 9 (i.e. Underline/Circle errors + Categorize errors + Explain errors + Provide examples) and Option 10 (i.e. Underline/Circle errors + Categorize errors + Explain errors + Provide examples + Provide correction) respectively. As for the control group, 30% and 17% of the participants preferred receiving Option 5 (i.e. Underline/Circle errors + Categorize errors + Explain errors) and Option 9 (i.e. Underline/Circle errors + Categorize errors + Explain errors + Provide examples) respectively. The two most well-received options respectively selected by the majority of the respondents from both groups, which are Options 9 & 10 for the experimental group and Options 5 & 9 for the control group, share one common characteristic. Students generally demand the error feedback comprising at least (1) error identification, (2) error categorization, and (3) error explanation. It is noted from the qualitative

comments (see below) that students from both groups might not feel there was a need for error correction by their instructors as this was part of students' learning responsibilities, and the majority of the respondents in the experimental group seem to be quite satisfied with the electronic feedback characterized by Option 9 (i.e. Underline/Circle errors + Categorize errors + Explain errors + Provide examples). It is noteworthy to point out that the majority of the respondents from the experimental group preferred Option 9. This implies that the provision of a recommended lexico-grammatical form or an example sentence in a new context through the web-based resources (i.e. Google link) is quite well-received by the experimental group receiving the electronic feedback as one of the feedback options. The following are the qualitative comments made by students with respect to Q12 (i.e. *Table 7.14*):

Focus of Written Comments Preferred less
Experimental Group
a. "It is important to try to make corrections first before giving them answers" b. "The best way to learn is to let the students find the correction. Teachers only need to indicate the errors they have made but not correct all errors." c. "[Option 9] would be the best, however, frankly speaking, most example sentences for every error is reluctant[?]. Example sentences for major error or common usages of phrases are always good for studies." d. "I think no correction is needed to be provided as students need to think by themselves to learn." e. "I can correct the error." f. "The error should be clear and understandable. Too much information is included may make students feel troublesome."
Control Group
a. "This method can let me know what my mistakes are, and it is the suitable way for me." b. "This option can let me know more about my mistakes made. Also, I can observe other sentences expressing similar meaning." c. "More information is better for us and help us in self-learning." d. "Explanations are examples are quite useful for students to learn from errors. Also, correct the errors ourselves is a good practice."

7.9   *Student Autonomy on Explicitness of Error Feedback (Q13)*

Q13 (see *Table 7.15*) was fielded to find out if students prefer making their own decision on the explicitness of error feedback (i.e. the types of error feedback strategy offered in Q12 in *Table 7.14*) they would like to receive on various error types from their instructors.

The findings in *Table 7.15* (Q13) show that as much as 83% of the participants in the experimental group and 64% of the participants in the control group indicated that they preferred or quite preferred to have autonomy over the choice of error feedback strategy received from their instructor. The relatively higher percentage of participants from the experimental group indicating their favour toward feedback autonomy might have been attributed to the fact that the experimental group had already benefit from the electronic feedback in which they could decide the level of explicitness of error explanation. The relatively lower percentage of preference on feedback autonomy exhibited by the control group might have been possibly due to the fact that they were not yet aware of the possibilities of making their own choice by the electronic means.

Table 7.15: Q13 – Based on your answer in Q12, do you prefer making your own decision on the explicitness of error feedback (i.e. the types of error feedback strategy offered in Q12) you would like to receive on various error types from your language instructor?

		Experimental Group	Control Group
	Most preferred	31%	7%
	Quite preferred	52%	57%
	Neutral	17%	33%
	Not quite preferred	0%	3%
	Not preferred	0%	0%

The following are the qualitative comments made by the participants from both groups.

Student Autonomy over the Explicitness of Error Feedback
Experimental Group
a. "It allows me to understand more about the errors I made."   b. "No need to ask for individual preference everytime ones submit their work. Perhaps every student will make a preference record on instructor's note on the first lesson."   c. "No, it may cause a lot of time to deal with."   d. "I would like to make my own decision."   e. "The choices is enough for me to choose and I can find the most preferred one."
Control Group
a. "It would be better, but I think it is useful when the case that we made a wrong word mistake rather than a grammar error."   b. "The understanding of each error types are different."   c. "If the technology can achieve this function, it will bring convenience to me because some errors are the careless mistakes, the feedback can be skipped."   d. "This is because different students have different needs for improving composition. Some may be weak in organization, such as me, but some may be weak in grammar."   e. "More information is better for us to do the improvement in the future. It is more flexible to know solution in detail for us to choose."   f. "Different student may have different needs."

Some of the above comments expressed by the respondents in the control group are quite insightful and encouraging in the sense that what they were expecting in the future was very similar to what the electronic feedback was offering to the experimental group in this study. Some respondents from the control group demonstrated some broad-brush awareness that different error types should be treated by different feedback strategies, and different learners might require different levels of explicitness in error feedback due to their different level of language competence.

7.10   Comprehensibility of Error Types (Q14-Q15)

Q14 and Q15 (see *Table 7.16* and *Table 7.17*) were fielded to find out which error types are more difficult (Q14) and easier (Q15) to understand by the respondents. Q14 is "Which of the following types of error indicated in the feedback is/are more difficult to understand?"; and Q15 is "Which of the following types of error indicated in the feedback is/are easier to understand?". The language items for selection were (1) Wrong Word/Wrong Word Choice, (2) Missing/Redundant Word, (3) 'Word-level' errors, (4) 'Clausal-level' errors, (5) 'Awkwardness' errors, (6) 'Tone & Style' errors, and (7) 'Organization-related' errors. Participants were allowed to select more than one language item.

The findings in *Table 7.16* (Q14) show that as much as 71% of the participants in the experimental group indicated that 'Awkwardness' errors was the language item most difficult to correct, whereas 57% of the participants in the control group indicated that both 'Awkwardness' errors and 'Clausal-level' errors were the most equally difficult to understand. On the other hand, the findings in *Table 7.16* (Q15) show that 'Awkwardness' errors was not the most easiest to understand, and this is then followed by 'Clausal-level' errors.

Table 7.16: Q14 – Which of the following types of error indicated in the feedback is/are more difficult to understand?

		Experimental Group	Control Group
1.	Wrong Word/Wrong Word Choice	21%	7%
2.	Missing/Redundant Word	0%	3%
3.	Word-level errors	0%	7%
4.	Clausal-level errors	29%	57%
5.	Awkwardness errors	71%	57%
6.	Tone & Style errors	25%	37%
7.	Organization-related errors	29%	17%

262

Table 7.17: Q15 – Which of the following types of error indicated in the feedback is/are easier to understand?

		Experimental Group	Control Group
1.	Wrong Word/Wrong Word Choice	36%	50%
2.	Missing/Redundant Word	71%	67%
3.	Word-level errors	50%	57%
4.	Clausal-level errors	7%	0%
5.	Awkwardness errors	0%	3%
6.	Tone & Style errors	0%	0%
7.	Organization-related errors	4%	0%

Scrutinizing the qualitative results from Q14 and Q15 together (i.e. *Table 7.16* and *Table 7.17*), it is noted that most participants from the experimental group and control group regarded 'Awkwardness' errors being most difficult to understand, and 'Clausal-level' errors being the second most difficult. These findings coincide with the theory of second language acquisition in which the Natural Order Hypothesis (Krashen & Terrel, 1983) postulates that the acquisition of grammatical structures follows a natural order which is predictable. Along similar lines, Truscott (1996) argued that different types of linguistic forms may take a different sequence of L2 acquisition. This means a student's interlanguage (i.e. proficiency) level would determine to what extent he or she can comprehend, process and utilize teacher feedback for a particular error type. As such, it appears reasonable that 'Awkwardness' errors and 'Clausal-level' errors, which are at the sentence-level, are deemed to be more difficult to be comprehended by our participants when comparing to 'Word-level' errors and others in the study. Second, these qualitative findings also coincide with the statistical results that the control group failed to make statistically significant number of revisions in 'Awkwardness' errors in draft 2 (see *Table 6.7 & Table 6.8*). The following are the qualitative comments made by students:

Error Types Most Difficult to Understand
Experimental Group
a. "Most are easy to understand, so I am choosing any."
b. "I may not understand what mistakes I made and how can I correct it."
c. "Due to <u>chinglish</u>, student may find it difficult to identify the unclear expression they have made."
d. "Tone and style error may also becomes very confusing sometimes."
e. "Difficult to understand as we can't not know the error exactly. May not know how to correct it."
f. "Because it usually involves <u>several sentences</u>."
g. "It is difficult to understand the <u>error in a sentence</u>."
h. "Sometimes, I don't know why the <u>sentence is treated as chinglish</u>."
i. "They are kinds of sterotype."
j. "The <u>mistakes of a whole sentence</u> is more difficult to correct."
Control Group
a. "I think this type of errors cannot be figured out by myself."
b. "Because if a student writes an <u>"awkward" sentence</u>, it means that it doesn't seem awkward for the student. So since it makes sense for the student, he won't know what the error is."
c. "Sometimes I cannot distinguish what kind of words should I use for informal writing. Also, as English is not my first language, I just tried my best to correct <u>awkward expression</u>, but I found it a bit difficult."
d. "It is because grammar and tone are important parts for us to use. It is difficult to improve in the short-term."
e. "It's hard for me to change these mistakes because I even don't know why I made mistakes there."

Error Types Easiest to Understand
Experimental Group
a. "Wrong use of tense are the easiest one to be identify."
b. "It is easier to understand error related to word."
c. "The singular-plural error is obvious to know."
Control Group
a. "They can be detected once I revise the composition."
b. "Obvious grammatical mistakes are easy to be corrected for me."
c. "Words are easier to understand than others."

7.11 Error Types Being More Difficult / Easier for Correction (Q16-Q17)

Q16 and Q17 (see *Table 7.18* and *Table 7.19*) were fielded to find out which error types are more difficult (Q16) and easier (Q17) to correct by the respondents. Q16 is "Which of the following types of error indicated in the feedback is/are more difficult to correct?"; and Q17 is "Which of the following types of error indicated in the feedback is/are easier to correct?". The language items for selection were (1) Wrong Word/Wrong Word Choice, (2) Missing/Redundant Word, (3) 'Word-level' errors, (4) 'Clausal-level' errors, (5) 'Awkwardness' errors, (6) 'Tone & Style' errors, and (7) Organization-related errors. Participants were allowed to select more than one language item.

The findings in *Table 7.18* (Q16) show that over 60% of the participants in the experimental group and the control group indicated that 'Awkwardness' errors was the language item most difficult to correct. On the other hand, the findings in Table 7.19 (Q17) show that 'Awkwardness' errors was not the least easiest to correct, and this is then followed by 'Clausal-level' errors.

Table 7.18: Q16 – Which of the following types of error indicated in the feedback is/are more difficult to correct?

		Experimental Group	Control Group
1.	Wrong Word/Wrong Word Choice	17%	3%
2.	Missing/Redundant Word	0%	0%
3.	Word-level errors	0%	10%
4.	Clausal-level errors	31%	43%
5.	Awkwardness errors	69%	63%
6.	Tone & Style errors	24%	40%
7.	Organization-related errors	28%	13%

Table 7.19: Q17 – Which of the following types of error indicated in the feedback is/are easier to correct?

		Experimental Group	Control Group
1.	Wrong Word/Wrong Word Choice	34%	47%
2.	Missing/Redundant Word	76%	57%
3.	Word-level errors	52%	63%
4.	Clausal-level errors	7%	3%
5.	Awkwardness errors	0%	0%
6.	Tone & Style errors	3%	0%
7.	Organization-related errors	3%	3%

Scrutinizing the qualitative results from Q16 and Q17 (*Table 7.18 and Table 7.19*) alongside Q14 and Q15 (*Table 7.16 and Table 7.17*), it is further confirmed that most participants from the experimental group and control group regarded 'Awkwardness' errors being most difficult to understand and correct. This is also supported by students' qualitative comments from Q14-Q17, where 'Awkwardness' errors was being named the most concerned grammatical error, whereas 'Word-level' errors was being named as the easiest. These findings also coincide with the theory of second language acquisition in which the Natural Order Hypothesis (Krashen & Terrel, 1983) postulates that the acquisition of grammatical structures follows a natural order which is predictable. Along similar lines, Truscott (1996) argued that different types of linguistic forms may take a different sequence of L2 acquisition. This means a student's interlanguage level would determine to what extent he or she can comprehend, process and utilize teacher feedback for a particular error type. As such, it also appears reasonable that 'Awkwardness' errors, which is at the sentence-level and defined as 'untreatable errors', is deemed to be more difficult to be comprehended and corrected by our participants when comparing to 'Word-level' errors and others in the study. Second, these qualitative findings also coincide with the statistical results that the control group failed to make statistically significant number of revisions in Awkwardness' errors

in draft 2 (see *Table 6.7* & *Table 6.8*). The following are the qualitative comments made by students:

Error Types Most Difficult to Correct
Experimental Group
a. "There are many ways to correct these errors, it is not easy to know which is the best."
b. "Correct <u>awkward/unclear expression</u> usually need to correct the whole sentence. Like chinglish."
c. "Sometimes I just couldn't figure out a formal phrase to replace an informal phrase."
d. "Not enough explanation are given and students may not understand the tutor's mind easily."
e. "It is difficult for me to understand the tone and style requirement when I am writing."
f. "It is difficult to correct a <u>sentence</u>."
g. "Without comprehensive explanations, these errors would probably make the students confuse."
h. "Sterotype. Hard to identify. Hard to find the right one."
Control Group
a. "As this type of errors are hard to be figured out by myself, I think if a correction provided is easier."
b. "Because it might imply correcting a whole paragraph or even more."
c. "It is because grammar and tone is important part for us to use. It is difficult to improve in the short-term."

Error Types Easiest to Understand
Experimental Group
a. "They can be corrected through grammar book."
b. " 'My words' can help and the 'google link' provided can help too."
c. "It is easier to correct error related to word."
Control Group
d. "It is because of careless."
e. "I think concept of words are easier to understand then others, so it is easy for us to understand."

7.12   *Preference on Post-writing Activities (Q18)*

Q18 (see *Table 7.20*) was fielded to find out which post-writing activities students think their instructors should ask them to do more often after they have received the marked composition. Participants were allowed to select more than one option item.

The findings in *Table 7.20* (Q18) show that the top three preferences indicated by the experimental group follows this order: (1) "Ask the teacher for clarifications, explanations or help in class", (2) Show and explain the most common errors/mistakes made by students, and (3) Correct all errors. On the other hand, the top three preferences indicated by the control group are: (1) Read the written comments, (2) Correct all errors, and (3) "Ask the teacher for clarifications, explanations or help in class".

Table 7.20: Q18 – Which of the following types of error indicated in the feedback is/are more difficult to correct?

		Experimental Group	Control Group
1.	Read the grade	3%	7%
2.	Read the written comments	31%	43%
3.	Correct all errors	34%	37%
4.	Correct some of the errors	10%	23%
5.	Rewrite the whole composition	3%	7%
6.	Ask the teacher for clarifications, explanations or help in class	48%	30%
7.	Consult dictionaries, grammar books or writing text books	7%	13%
8.	Refer back to previous compositions	10%	10%
9.	Work with a partner to help each other improve the composition	7%	13%
10.	Work on a proofreading exercise	14%	20%
11.	Read aloud some good sentences in class	3%	7%
12.	Show and explain the most common errors/mistakes made by students	38%	17%
13.	Hold an individual consultation with the instructor to get his/her advice	28%	17%
14.	None of the above	0%	3%
15.	Others	3%	0%

The findings in *Table 7.20* (Q18) suggest that participants from both groups are most interested in 'asking the teacher for clarifications, explanations or help in class' and 'correcting all errors'. They

seem to be least interested in 'checking 'reading the grade' and 'reading aloud some sentences in class'.

7.13 Usefulness of Teacher Written Feedback over Long-term Writing (Q19)

Q19 (see *Table 7.21*) of the questionnaire aimed to find out, to what extent, students believe teacher written feedback (as a whole) helps with their long-term writing.

The findings in *Table 7.21* (Q19) show that as much as 86% of the participants in the experimental group believed that the electronic feedback as a whole could help them with their long-term writing. Among these, close to 25% of the participants believed that the electronic feedback was very helpful. On the other hand, 73% of the participants in the control group believed that the paper-based feedback as a whole could help them with their long-term writing. Of these, 10% of the participants believed that the paper-based feedback was very helpful.

Table 7.21: Q19 – To what extent do you think the teacher written feedback as a whole helps you with your long-term writing?

		Experimental Group	Control Group
1.	Very helpful	24%	10%
2.	Quite helpful	62%	63%
3.	Neutral	14%	23%
4.	Quite unhelpful	0%	0%
5.	Unhelpful at all	0%	3%

The findings in *Table 7.21* (Q19) show that the majority of the respondents in both groups regard their teacher written feedback as helpful to their long-term writing development. These findings echo the statistical results that both the electronic feedback and the paper-based feedback (within

their own groups) were able to make statistically significant number of revisions in all errors in draft 2 (see *Table 6.5* in Appendix D: p-value [control group] = 7.496e-05; p-value [experimental group] = 2.736e-10). However, with a smaller p-value and larger effect size, the electronic feedback (within its group) was able to make more statistically significant number of revisions in all errors when comparing those made by the paper-based feedback.

Most qualitative comments remarked by the participants in the experimental group have confirmed these positive findings of the electronic feedback in terms of its helpfulness for the long-term writing development. Among those are "Thank you!", "Detailed explanation is provided and there are google link provided which is different from the previous online feedback I received before. Give me motivation to find out more by myself", "It is helpful for me", and "It can help us to be careful about the grammar. Also, it helps us to know more about how to improve the content of writing". When compared to the comments left by the participants in the experimental group, fewer positive comments were made by the control group in terms of the helpfulness of the paper-based feedback over the long-term writing development. The only positive comment remarked by the control group is "Feedback can help me to elaborate some points of my compositions."

Chapter 8   Conclusions

The effectiveness of teacher written feedback has been a subject of debate in second language writing for decades. The most basic debate in this area among ESL writing researchers and teachers is whether teacher written feedback in various forms of treatment (e.g. pen-and-paper form, computer-generated feedback or peer feedback), of various types (i.e. grade, error feedback or written commentary), with different focus (i.e. content, language or organization), with various feedback strategies (i.e. level of explicitness), and in varying levels of comprehensiveness (i.e. detailed or selective error identification) has any positive effects on student writing development. Ferris (1999, 2002, 2003, 2004, 2006) and Lee (2008a, 2008b) argued that while the effectiveness of error feedback was not conclusive, more research should be done to explore in what ways error feedback, as an integral part of teacher written feedback, can be improved. On the other hand, a number of research studies (e.g. (e.g. Willetts; 1992; Egbert, 2002; Milton, 2006; Lee, 2008a; Stevenson & Phakiti, 2013; Ene & Upton, 2014) suggested that the use of technology might offer an alternative way of giving teacher written feedback in student writing more effectively and efficiently. However, the critical review of 'computer-facilitated feedback' (Ene & Upton, 2014) as well as 'computer-generated feedback' (Stevenson & Phakiti, 2013) suggested that the effectiveness of these 'computer-mediated feedback' on student writing is also not conclusive (see Chapter 3.5).

As has been seen so far in the literature review of teacher written feedback, both the effectiveness of the traditional 'paper-and-pen feedback' and 'computer-mediated feedback' is inconclusive.

Indeed, a number of methodological flaws and misinterpretation of findings have been identified in my critical review of the 'teacher written feedback' and 'computer-mediated feedback' studies (which are further divided into 'computer-facilitated feedback' and 'computer-generated feedback'). The heterogeneity of these studies characterized by different focus, research designs, institutional and instructional contexts, and participant backgrounds, alongside these methodological flaws and misinterpretation of findings, which have possibly neglected the impact of extraneous variables on the validity and reliability of the studies, might have given rise to these mixed results for both paper-and-pen feedback and computer-mediated feedback. In other words, those studies might not have possibly minimized or controlled these extraneous variables but simply attributed the effectiveness or ineffectiveness of a particular feedback treatment to the treatment itself. In other words, the objectives of these studies were actually to test if a particular form of feedback treatment or feedback practice more effective than others in error reduction under an experimental condition. However, it was found in my critical review that the causality between these feedback treatments and their outcomes of error reduction might have been blurred or influenced by some extraneous factors and even methodological flaws in both 'paper-and-pen feedback' studies and 'computer-mediated feedback' studies. Hence, any positive or negative claims identified in error reduction should not be merely attributed to the feedback treatment being tested under the experiment condition alone.

For example, in those 'paper-and-pen feedback' studies, these extraneous factors are the (1) institutional requirements (e.g. teacher autonomy, teacher trainings, detailed or selective marking, product writing or process writing), (2) instructional contexts (e.g. teachers' beliefs and practice like focus on meaning or forms; preference on direct feedback or indirect feedback for various

error types), (3) students' language proficiency, (4) students' level of motivation, as pointed out by Goldstein (2004, 2005), Hyland & Hyland (2006b) and Lee (2008a, 2008b), and (5) student uptake of different forms of feedback on different error categories, as pointed out by Truscott (1996) who argued that most teachers responded to all error types in the same way, and as being neglected by Ferris (2006) in her study which has rendered her findings about the superiority of indirect feedback invalid. Details of the extraneous factors (1) to (4) were discussed thoroughly in Chapter 3.2.3(H) 'Contextualization of Feedback Studies'. Details of the extraneous factor (5) in Ferris's (2006) study are explained below.

Ferris's (2006) inadequate comparison of different error types towards the testing of different feedback treatments in the absence of a controlled condition under which teachers were free to use either direct feedback or indirect feedback to respond to all error types based on their intuitive choices. It was reported in Ferris's (2006) study that teachers mostly used indirect feedback (59%) for 'treatable errors' (i.e. 'Word-level' errors) and direct feedback (65%) for 'untreatable errors' (i.e. 'Awkwardness' errors and 'Collocation' errors). Ferris concluded in her study (2006) that 85% of 'untreatable errors' were successfully corrected by students, and further claimed that teacher feedback on writing revisions was helpful. In my view, the improvement in reducing 'untreatable errors' by students might have not been possibly attributed to the language gain rendered by the feedback itself, but its overt and explicit provision of the correct linguistic forms for the students 'as a quick fix' in the short term. This was subsequently recognized by the teachers in the post-interview of the study, As such, the question being raised in this study is if one of the purposes of Ferris's (2006) study is to compare the effectiveness of different feedback forms in remedying 'treatable errors' and 'untreatable errors', would it be more methodologically

appropriate to impose some conditions in which, for example, only direct feedback should have been used for all errors in one condition and then only indirect feedback should have been used for all errors in another condition.

Another methodological flaws arisen in Ferris's (2006) study is her inadequate comparison of different error types towards the testing of different feedback treatments in the absence of the consideration for the sequence of second language acquisition in her research design. It is reported in her study that direct feedback on 'non-rule governed sentence-level errors' (i.e. 'untreatable errors') did not show more significant long-term gains in accuracy over time (from Essay 1 to Essay 4), but indirect feedback on verb errors (i.e. 'treatable errors') did show improvement in error reduction over time. As such, Ferris claimed in her study (2006) that indirect feedback is superior than direct feedback in improving students' language accuracy in the long term, given what we were told that indirect feedback was mostly (59%) used for 'treatable errors' whereas direct feedback (65%) for 'untreatable errors'. The concerns being raised here is if it is adequate and valid to put on par the student uptake of direct feedback on 'untreatable errors' (i.e. 'Awkwardness' errors) and the student uptake of indirect feedback on 'treatable errors' (i.e. verb-form) for the comparison of their effectiveness in improving the long-term language gain, and in the end to conclude that indirect feedback is more superior than direct feedback in remedying errors in the long term. The above methodological flaws identified in Ferris's (2006) study were addressed and avoided in my study, which are already detailed in Chapter 5.4.1 'Measures for Minimizing Methodological Flaws'.

For the 'computer-facilitated feedback' studies, their results are susceptible to (a) logical flaws, (b) methodological flaws, and (c) lack of insightfulness. As for those 'computer-generated feedback' studies, the inconclusive results are due to 'paucity of research, heterogeneity of existing research, the mixed nature of research findings, and the methodological issues in some of the existing research' (Stevenson, M & Phakiti, A., 2013, p.62). This was discussed thoroughly in Chapter 3.5 'The Use of Technology in Teacher Feedback on Writing'.

Extraneous factors are variables that influence the outcome of an experiment. They are not the variables of interest in this study. Against this, a number of measures were implemented in the research design of this study (see Chapter 5.4 'Validity and Reliability of the Methodology') to minimize or control the influence of extraneous variables as much as possible such that any language gain could have been more likely attributed to the form of feedback treatment adopted (as opposed to Ferris's study (2006)), thus enhancing the reliability and validity of the study.

This study focused on examining the effectiveness of the electronic feedback system named 'Mark My Words' ('MMWs'), as a form of feedback treatment adopted in this experimental study, in improving students' error reduction in their writing revisions. This focus premised on the ground that the capability of a particular form of feedback treatment in improving the quality of students' revisions is central to the claims made about the effectiveness of 'MMWs' feedback, as a form of feedback treatment, in error reduction.

It is noteworthy to point out that 'Mark My Words' ('MMWs') is not a fully automated writing evaluation tool. Unlike those 'Automated Writing Evaluation' ('AWE') software programs (the so-called computer-generated feedback) which can automatically detect and identify language errors based on techniques such as artificial intelligence, natural language processing and latent semantic analysis (Stevenson & Phakiti, 2013), the pedagogical functions of 'MMWs' is solely operated by teacher-users via an add-on tool bar specially designed for Microsoft Word. This add-on tool bar enables teacher-users to manually insert pre-set comments alongside boilerplate 'resource-rich' feedback into students' written assignments via the web-based platform (please refer to Chapter 4.4.2 'Pedagogical Operation of 'Mark My Words''). Despite the difference between 'MMWs' and 'AWE' in terms of their technological level of advancement, the reason for the inclusion of 'AWE' into the literature review and discussion of this study are that, (1) if we take a broader scope and a looser definition in examining the use of technology in teacher writing feedback, both 'MMWs' and 'AWEs' can be viewed as under the same category of 'computer-assisted language learning' (CALL) on writing, in which both of them involve the integration of technology in supporting teacher feedback on writing; and (2) the methodological designs of those 'AWE' research studies and their research outcomes provide this study much insights as to how to improve the quality of the electronic feedback under limited resources (unlike those 'AWE' programs which are mostly institutionally funded or commercially funded for the profit-making purposes), as well as to enhance the validity and reliability of the research design in this experimental study (which was already explained in greater detail in Chapter 3.5, Chapter 5.4 and Chapter 5.5).

There were three research questions in this experimental study. Whether teacher written feedback, in the form of the electronic feedback, can lead to error reduction in student writing revisions is what the first two research questions in this study aimed to investigate by primarily using the 'error-count' approach. The third question aimed to explore students' attitudes towards teacher written feedback by administering questionnaires, and the qualitative results of which (Chapter 7) were triangulated with the statistical results of the error count (Chapter 6). Both quantitative results alongside qualitative results were triangulated with the results of some previous studies (e.g. Truscott, 1996; Ferris, 1999, 2002, 2003, 2004, 2006; Lee, 2008a, 2008b) in order to obtain a fuller, more insightful and meaningful understanding of the role of teacher written feedback, modeled on Nunan's (1997) Model of Framework for Developing Learner Autonomy (1997) and Krashen's (1985) Input Hypothesis in error correction.

The three research questions, which were stated in Chapter 4.2 'Research Questions', are as follows:

1. Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a difference on student writing revisions, when comparing to the paper-based feedback?

2. Did the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' make a difference on student writing revisions for each error category, when comparing to the paper-based feedback?

3. What were the students' perceptions of the effectiveness of teacher written feedback?

This study began with examining the first draft and second draft of an essay written by two treatment groups comprising sixty-two Year 2 engineering students over six weeks. It was hoped that by comparing the total number of errors identified in each of the seven error categories between the first draft and the second draft across and within these two treatment groups, some kind of pattern would appear revealing the effectiveness of teacher written feedback in use. Efforts were made to prevent the results to be affected by extraneous variables such as the selection of markers, the marking principles imposed on the markers as well as the homogeneity of both treatment groups with respect to the academic background and the level of language competence.

8.1   *Controversy over the Effectiveness of Teacher Written Feedback*

Controversy continues as to whether feedback, or particularly which form of feedback, can help improve accuracy and overall quality of student writing (Robb et al., 1986; Truscott, J., 1996, Ferris, D.R., 2001, Lalande, J.F., 1982; Semke, 1984; Ferris, 1996, 1999, 2006; Lee, 1997, 2008a, 2008b; Stevenson & Phakiti, 2014; Ene & Upton, 2014). In fact, a number of studies have been conducted to examine how students responded to and acted upon various types of teacher written feedback underpinned by different feedback strategies or forms. For example, direct feedback vs. indirect feedback, coded indirect coded feedback vs. indirect uncoded feedback, teacher feedback vs. peer feedback, and paper-based feedback vs. computer-mediated feedback. However, no conclusive evidence could be drawn to support which form of feedback is relatively more effective than the others in improving accuracy in student writing revisions. Such controversy over the

effectiveness of teacher written feedback was found on two limitations exhibited by teachers and students, which are mainly characterized by the lack of feedback strategies in remedying errors of different categories as well as the lack of the understanding on students' variations in their interlanguage level to comprehend, process and utilize teacher feedback (see Chapter 3.2.3 'Effectiveness of Teacher Written Feedback' and Chapter 3.5 'The Use of Technology in Teacher Feedback on Writing').

With the 'pedagogical operation of 'Mark My Words' (MMWs) underpinned by Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis', participants in the experimental group could opt for the type of feedback strategies (i.e. varying in terms of their levels of explicitness) in response to their interlanguage level and learning needs. It is believed that the positive results of this study can contribute some sort of concrete evidence to the growing body of literature of the 'effectiveness of teacher written feedback' and 'second language writing', thus clarifying some mixed results of the previous research. The significance of the findings is as follows:

1. Students in this study could utilize the electronic feedback (see Table 6.4 and Table 6.5 in Appendix E) based on Nunan's (1997) 'Model of Framework for Developing Autonomy' and Krashen's (1985) 'Input Hypothesis', as opposed to some previous studies which claimed that students ignored teacher feedback or failed to utilize it effectively in their writing revisions (Cohen & Robbin, 1976; Truscott, 1996; Zamel, 1985).

2. Students in this study welcomed teacher written feedback and regarded teacher written feedback as being helpful to their language development, according to the student questionnaire results.

3. 'Awkwardness' errors (i.e. non-rule governed and untreatable sentence-level errors) is the most difficult error type whereas 'Word-level' error is the easiest error type perceived by students in the questionnaire results. These findings lend support to the theory of second language acquisition (i.e. The Natural Order Hypothesis) in which different types of linguistic form follow a different sequence of L2 acquisition, and to the fact that students at different interlanguage levels have different levels of progression in error correction. These results from the questionnaires are consistent with the quantitative results which suggest that students in general made less statistically significant improvement in 'Awkwardness' errors whereas made more statistically significant improvement in 'Word-level' errors after receiving teacher feedback (see Appendix C and Appendix E).

4. Students in the control group made no statistically significant number of reduction in 'Awkwardness' errors and 'Collocation' errors, even in its within-group comparison, according to the 'error count' results (Tables 6.8 & 6.17 in Appendix E).

5. The electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Autonomy' and Krahsen's (1985) 'Input Hypothesis', which was able to facilitate students in comprehending, processing and utilizing indirect coded feedback, made statistically significant number of revisions in all error types, especially errors highly characterized by 'L1-L2 Interference' and regarded as 'untreatable', such as 'Awkwardness' errors and 'Collocation' errors. Both statistical results from the 'error count' (see Appendix E and Appendix G) and positive results from the questionnaires (see Chapter 7 'Findings &

Discussions: Students' Perceptions on Teacher Written Feedback') lend support to each other.

6. The electronic feedback with its interactive features has allowed students to play an active and autonomous role in error correction, such that students can exercise their autonomy to determine the level of explicitness they need in response to various error types.

7. The majority of the students from both groups preferred student autonomy in error correction when they were asked if they would prefer making their own choice on the feedback strategies they would like to receive on various error types. This is evident by the questionnaire results which indicated that 83% and 64% of the students from both groups would prefer having autonomy on feedback strategies (see Chapter 7.9). For the experimental group, 4 out of 5 qualitative comments indicated that they would prefer autonomy on feedback options. For the control group, 6 out of 6 qualitative comments indicated that they preferred autonomy on feedback options (see Q13 of Chapter 7.9). This resonates with Lee (2008a, 2008b) who argues for student autonomy in error correction.

8. The electronic feedback with its three-step feedback approach, characterized by the provision of a recommended lexcio-grammatical form via Google link, has catered to variations in students' interlanguage level and the student uptake of different forms of feedback treatment on different error types, in such a way that it can more effectively address students' individual learning needs and respond to those errors highly characterized by 'L1 interference' and 'regarded as 'untreatable' errors.

9. The provision of a recommended lexico-grammatical form in authentic texts via 'Google Link', which is approximated to the subtle process of hypothesis making and testing in Krashen's (1985) 'Input Hypothesis', was quite well received by students receiving the

electronic feedback. This is evident by, out of 10 positive qualitative comments expressed by the respondents in the experimental group about the comprehensibility of feedback, 'Google link' was mentioned 4 times respectively in these 4 out of 10 comments (see Chapter 7.1). In addition, this is also evident by, out of a total of 4 positive qualitative comments expressed by the respondents in the experimental group about their ability in error corrections, 'Google link/search/example' was mentioned 3 times respectively in these 3 out of 4 comments (see Chapter 7.2) Also, the 'error count' results suggested that 'Awkwardness' errors received significant reduction for students receiving the electronic feedback with this distinguished feature when compared to the control group.

10. Students in this study generally regarded error feedback as more suitable for addressing 'language' errors (see Chapters 7.3-7.7 (Qs6-11)) while teacher written commentary as more suitable for addressing 'Organization' errors and 'Tone & Style' errors, according to the questionnaire results.

11. A number of studies reported that teacher feedback is not always understood due to its illegibility (Zamel, 1985; Lee, 2008b). Truscott (1996) argued that practical problems like incomplete, inconsistent and inaccurate teachers' error feedback have rendered the writing feedback ineffective. Also, Truscott (1996) attributed such ineffectiveness to the teachers' lack of skills to analyze and explain language problems, as well as students' lack of skills to understand and use feedback. In response to these concerns and the subsequent suggestions made by Ferris (2006), a consistent system of coded feedback with grammar instructions alongside web-based resources was incorporated in 'Mark My Words' ('MMWs').

12. Bitchener & Knoch (2010b) concluded that the "provision of clear, simple metalinguistic explanation with examples" on a separate sheet of attached paper was the best type of written corrective feedback for long-term accuracy (p.216). The question I raised here is the practicality of the provision of grammatical explanation on a separate piece of paper if the types of error teachers attended add up. In response to such space constraint posed by the paper-based feedback, the electronic feedback can solve this problem.

13. In terms of the types of teacher feedback, the majority of students' preference was teacher commentary, and then followed by error feedback and grade. By scrutinizing the qualitative comments they made, the following tendency of students' attitudes is noted:

	Teacher comments	Quite a number of students' comments indicated teacher comments were more helpful in focusing on 'content', organization' and 'writing style (tone & style)'.
		Some comments indicated that teachers' commentary could help provide them the overall performance of their writing.
		Some comments indicated that teacher commentary could tell students what teachers thought about their writing.
		Some comments indicated that teacher commentary could justify the grades.
	Error feedback	Quite a number of students indicated that error feedback should focus on 'language', 'organization' and then 'content'.

8.2   *Responsiveness of Teacher Written Feedback on Error Correction*

The statistical results in both the 'between groups' and 'within groups' comparisons (see Appendix E) concluded that the electronic feedback treatment was more effective than the paper-based

feedback treatment in remedying all error types, especially 'Awkwardness' errors and 'Collocation' errors which are highly characterized by 'L1 interference' and regarded as 'untreatable errors'.

Hence, the pedagogical operation of 'Mark My Words' (MMWs) underpinned by Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' appears to demonstrate a more effective means over the paper-based feedback in facilitating students to comprehend, process and utilize indirect coded feedback in error reduction. The statistical results in Chapter 6 ('Findings and Discussions: Error Reduction') have answered the first research question that writing feedback modeled on Nunan's (1997) Model of Framework for Developing Learner Autonomy and Krashen's (1985) Input Hypothesis has made a difference on student writing revisions, and the second research questions that student uptake of the electronic feedback, which is modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis', on all error categories has been relatively more successful than student uptake of the paper-based feedback, in particular for those error categories highly characterized by 'L1-L2 Interference' and regarded as 'untreatable errors', namely 'Awkwardness' errors and 'Collocation' errors. The varying levels of explicitness of error feedback offered by the electronic feedback, to a great extent, has responded to Ferris's subsequent recommendation that a judicious combination of direct & indirect feedback, varying according to error types, may be most helpful to students (Ferris, 1999, 2006). And it can be seen in this study that, on the one hand, hardly could the 'paper-and-pen' feedback, in the absence of the electronic- and interactive environment, allow students to exercise their autonomy to determine the level of explicitness they need in response to various error types; on the other hand, hardly could it allow

teacher-makers to respond to errors judiciously in the absence of such electronic and interactive environment. This means when the time a teacher-marker has underlined, circled or coded any erroneous form, the teacher-marker himself or herself has already determined the kind of feedback the students should receive (e.g. Type: grade, error feedback or written commentary? Focus: language-oriented or content-oriented? Level of explicitness? Detailed marking or selective marking?), which is quite teacher-centered.

8.3 *Students' Perceptions on Teacher Written Feedback*

Qualitative results obtained the questionnaire (see Chapter 7) which cover the students' perceptions on teacher written feedback were triangulated with the quantitative results obtained from the 'error count' (Chapter 6) in order to supplement the explanation of the statistical results with a greater depth and provide us more insights about students' attitudes towards teacher written feedback (see *Table 8.1* below). It is found that not only the qualitative comments (in terms of 'feedback comprehensibility' in Chapter 7 .1, 'ability for error correction' in Chapter 7.2, 'the error type being the most difficult / the easiest for comprehension' in Chapter 7.10 and 'the error type being most difficult / the easiest for correction' in Chapter 7.11) coincided with the corresponding statistical differences noted between the experimental group and the control group, but also the results of which lent support to research question 1 and research question 2, where the electronic feedback made a statistical difference in the total number of student writing revisions and it was evident that student uptake of the electronic feedback are relatively more successful than student uptake of the paper-based feedback on all error categories.

Table 8.1 Summary of the Questionnaire Responses

Related Chapter (Question #)	Summary of the Questionnaire Responses
Chapters 7.1 & 7.2 (Q1 & Q2)	Experimental group was more able to comprehend, process and utilize the electronic feedback for error correction. This coincides with the 'error count' results that the experimental group was able to make statistically significant number of revisions across all 7 error categories in their writing revisions.
Chapters 7.3 & 7.4 (Q3 – Q6)	Both groups prefer 'written comments' > 'error feedback' > 'grade'.
Chapter 7.6 (Q9)	Both groups prefer to have all errors identified.
Chapter 7.8 (Q12)	Both groups prefer feedback strategies with (1) Error identification; (2) Error categorization and; (3) Error explanation. This suggests that participants, irrespective of treatment groups, attached more value on comprehension of their errors rather than merely having their errors corrected by their teachers.
Chapter 7.5 (Q7 & Q8)	Both groups preferred to have the error feedback focusing on language-related errors.
Chapter 7.7 (Q10 & Q11)	Both groups preferred to have the written comments focusing on organization-related errors. This implies that the students might have realized that the usual practice of error identification by means of 'underlining, circling and coding' might not work equally effective for organization-related problems. Rather, it can be inferred that they might think written commentary was deemed to be more effective and suitable for commenting on the logical development aspect of an essay.
Chapter 7.9 (Q13)	Both groups preferred to have more control of their own on the level of explicitness in error feedback as most respondents believed that they wanted to take a more active role in the feedback process and were able to make good judgment on their learning needs. This suggests that the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' could possibly address favourably to such students' preference.
Chapters 7.10 & 7.11 (Q14 – Q17)	Both groups agreed that 'Awkwardness' errors was the most difficult language item to comprehend and correct, whereas 'Word-level' errors was the easiest. This coincides with the statistical results that the control group failed to make any statistical difference in the reduction of 'Awkwardness' error, while 'Word-level' errors seemed to be relatively indifferent to the two forms of feedback treatment. On the contrary, despite the same level of relative difficulty perceived by the experimental group, students receiving the electronic feedback were much more able to correct 'Awkwardness' errors. This

	suggests that the electronic feedback modeled on Krashen's (1985) 'Input Hypothesis' could possibly respond more effectively to errors highly characterized by 'L1 interference' and regarded as untreatable errors.
Chapter 7.12 (Q18)	Both groups preferred 'asking teachers for clarifications, explanations and error correction' after they have received their marked essay.
Chapter 7.13 (Q19)	Over 80% and 70% of the respondents respectively from both groups believed that the feedback treatment they received had helped them with their long term writing.

According to the qualitative results obtained from the questionnaire, respondents in the experimental group believed that they were more able to comprehend, process and utilize the electronic feedback for error correction (Chapters 7.1 & 7.2). This coincides with the 'error count' results that the experimental group was able to make statistically significant revisions across all seven error categories in their writing revisions. Both the experimental group and the control group shared the same order of preference (Chapters 7.3 & 7.4) with 'written comments' being the most preferred item, followed by 'error feedback' and then 'grade'. The majority of the participants in both treatment groups would like to have all errors indicated by teachers (Chapter 7.6). Most respondents from both treatment groups shared a very similar preference on the error feedback strategy comprising (1) Error identification; (2) Error categorization and; (3) Error explanation (Chapter 7.8). This suggests that participants, irrespective of treatment groups, attached more value on comprehension of their errors rather than merely having their errors corrected by their teachers. Both treatment groups would prefer to have the error feedback focusing on language-related errors (Chapter 7.5) whereas they would prefer to have the written comments focusing on organization-related errors (Chapter 7.7). This implies that the students might have realized that the usual practice of error identification by means of 'underlining, circling and coding' might not work equally effective for organization-related problems. Rather, according to student preference

indicated in Chapter 7.7, it can be inferred that they might think written commentary was deemed to be more effective and suitable for commenting on the logical development aspect of an essay. Despite the fact that participants from the control group was excluded from receiving the electronic feedback, respondents from both groups would prefer to have more control of their own on the level of explicitness in error feedback (Chapter 7.9) as most respondents believed that they wanted to take a more active role in the feedback process and were able to make good judgment on their learning needs. This suggests that the electronic feedback modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' could possibly address favourably to students' preference. It is also found that respondents in the experimental group were quite satisfied with the electronic treatment they received. Both groups agreed that 'Awkwardness' errors was the most difficult language item to comprehend and correct, whereas 'Word-level' errors was the easiest (Chapter 7.10 & 7.11). This coincides with the statistical results that the control group failed to make any statistical difference in the reduction of 'Awkwardness' errors, while 'Word-level' errors seemed to be relatively indifferent to the two forms of feedback treatment. On the contrary, despite the same level of relative difficulty for 'Awkwardness' errors perceived by the experimental group, students receiving the electronic feedback were much more able to correct 'Awkwardness' errors. This suggests that the electronic feedback modeled on Krashen's (1985) 'Input Hypothesis' could possibly respond more effectively to errors highly characterized by 'L1 interference' and regarded as untreatable errors. Both groups would prefer 'asking teachers for clarifications, explanations and error correction' after they have received their marked essay. Over 80% and 70% of the respondents respectively from both groups believed that the feedback treatment they received had helped them with their long term writing.

8.4   *The Role of L1 in ESL Writing*

A number of previous studies (Ying, 1987; Chen, 1998; Cai, 1998; Chen, 2000; Cai, 2002; Hsin, 2003; Chan, 2004) have contributed to our understanding on how L1 interference gives rise to negative transfer in ESL writing from Chinese students. However, the focus of this study is not to add any evidence if and in what ways L1 interference has taken place but to investigate if there is any alternative feedback strategy (that is the electronic feedback in this case) which can more effectively remedy the 7 error categories, in which L1 interference might possibly have a negative role to play. Indeed, quite a number of errors, irrespective of 'treatable errors' or 'untreatable errors' as defined by Ferris (2006), identified in the student writing of this study revealed that such negative transfer were indeed taking place as a result of the dichotomy between TPL in Chinese and SPL in English, as well as, students' reliance on their L1 linguistic repertoire as reference when they were probably thinking in Chinese but writing in English simultaneously. This is particularly evident in two of the seven error categories, namely 'Awkwardness' errors and 'Collocation' errors, which the students found most difficult to understand and correct according to the quantitative findings of the control group (see Chapter 6) where no intervention was made, and the qualitative findings of both treatment groups (see Chapter 7). The reason why 'Awkwardness' errors and 'Collocation' errors were relatively more difficult for our participants for comprehension and correction could be explained by Ferris (1996, 1999) who argued that students were less able to correct 'untreatable' errors such as 'sentence-level' errors (which is equivalent to 'Awkwardness' errors in this study) and word collocation (which is equivalent to 'Word-level' errors) through feedback on form because these errors cannot be explained by and identified with particular grammatical rules and specific codes.

Despite the findings of these previous studies suggested that L1 Interference might have posed difficulty for Chinese ESL learners in writing 'standard' English, the results of this study seemed to suggest that L1 interference (or negative transfer) might possibly be overcome at the 'revising' stage as long as an effective feedback strategy was provided. The electronic feedback ('MMWs') might offer an alternative solution to this. In the final analysis, it may not be 'L1-L2 Interference' that matters, but whether our feedback strategy can to what extent remedy errors due to 'L1-L2 Interference' and 'untreatable errors' that does.

8.5   *Implications of the Source of Errors on Teacher Written Feedback*

Following what was earlier discussed in Chapter 3.3 'Linguistic Differences between English and Chinese', the findings indicated that the sources of most grammatical errors identified in student texts could be also traced to (1) 'L1-L2 Interference'; (2) the 'Theory of Language Typology', and (3) the 'Theory of Markedness'. These sources of errors would possibly have an impact on the interlanguage of individual ESL learners, where 'Interlanguage' refers to a developmental process ESL learners undergo to become competent in L2 (Brown, 1994). This section first discusses the relationship between the sources of errors and some of their corresponding key error types, and then explores if the source of errors has any relevance to the feedback practice which should be adopted.

8.5.1 L1-L2 Interference

'L1-L2 interference' is a process under which ESL writers constructed their own L2 interim rules, more or less approximated to the rules of L2 structure, with the help of their L1 knowledge (Ellis, 1997).

In their production and comprehension of L2 during such development language stage, ESL learners often either directly refer to their L1 rules or construct their own perceived L2 interim rules based on their approximation to L1 rules. This process is called 'L1-L2 interference'. Such level of L2 competence, characterized by varying degrees of reliance on L1 as reference, is known as 'Interlanguage'. 'Interlanguage' should be viewed as a continuum with one end being viewed as nearly-native competence in L2 and the other end being otherwise.

Error types possibly affected by, but not limited to, 'L1-L2' interference included [1] 'Awkwardness' errors, [2] 'Sentence Fragment', [3] 'Run-on Sentence', [4] 'Mixed structure', [5] 'Imbalanced Structure' and [6] 'Wrong Word Collocation', and [7] 'Finite and Non-finite words'. Except [6] 'Wrong Word Collocation', the common linguistic features of these errors is that they are mostly at sentence-level, clausal-level or phrasal-level, and more significantly, their linguistic patterns to a certain extent bear resemblance to L1 syntactic features. The 'Error Count' results (see Appendix C) provided some good evidence that all these 7 error types identified in the students' first draft from both treatment groups fall within the top 20 error types among the total of 77 error types on the list. Some example errors are shown below:

[1] 'Awkwardness' errors	
Student's original sentence	And for solar energy, only at daytime energy could be generate.
Chinese	Same grammatical structure as above (L1-L2 interference)
Revised version after treatment	And for solar energy, energy could only be generated at daytime.

[2] Sentence fragment	
Student's original sentence	Since people has little awareness on environmental protection.
Chinese	Same grammatical structure as above (L1-L2 interference) No conception of a dependent clause or an independent clause in Chinese
Revised version after treatment	Since people has little awareness on environment protection, the government should make more efforts to educate the public.

[3] Run-on sentence	
Student's original sentence	Switching to renewable energy echoes the government policy in energy saving, it improves the air quality and extend the life of our earth.
Chinese	Same grammatical structure as above (L1-L2 interference) No conception of a dependent clause or an independent clause in Chinese
Revised version after treatment	Switching to renewable energy echoes the government policy in energy saving which can improve the air quality and extend the life of our earth.

[4] Mixed Structure	
Student's original sentence	By exercising can make you fit (teacher interpreted it as 'L1-L2 interference' rather than a careless mistake of omitting a dummy subject)
Chinese	Same grammatical structure as above (L1-L2 interference) Subject is not grammatically obligatory in Chinese
Revised version after treatment	By exercising, it can make you fit.

[5] Imbalanced Structure	
Student's original sentence	Nuclear power is cost effective but high risk.
Chinese	Same grammatical structure as above (L1-L2 interference) High risk could be a noun or an adjective in Chinese
Revised version after treatment	Nuclear power is cost effective but highly risky.

[6] Wrong Word Collocation	
Student's original sentence	We are doing the comparison between hydroelectricity and some other renewable energy.
Chinese	'Doing the comparison' is identical to the equivalent Chinese literal expression
Revised version after treatment	We are making the comparison between hydroelectricity and some other renewable energy.

[7] Finite vs. Non-finite Verb	
Student's original sentence	We can build more wind mills can save more energy so protect the environment.
Chinese	Same grammatical structure as above (L1-L2 interference) Chinese language can have more than one main verb in a sentence, and there is no conception of a dependent clause or independent clause in the Chinese language
Revised version after treatment	We can build more wind mills to save more energy and protect the environment.

8.5.2 Theory of Language Typology

According to Li & Thompson (1976), the 'Theory of Language Typology' suggests a dichotomy between the 'subject-prominent' structure of English language (SPL) represented by the S-V-O pattern and the 'topic-prominent' structure of Chinese language (TPL), which possibly gives rise to 'L1-L2 Interference'. Li & Thompson (1976) stated that English is the 'subject-prominent' language in which subject is syntactically and/or semantically obligatory, whereas Chinese is the 'topic-prominent' language in which subject may be syntactically optional. Some example errors include [4] Mixed Structure and [7] Finite vs. Non-finite Verbs above, plus the following:

[8] Missing Subject in the S-V-O Pattern	
Example	By exercising can make you fit (teacher interpreted it as the omission of a subject which was seen as a careless mistake rather than 'L1-L2 interference'
Chinese	Same grammatical structure as above (Language Typology and L1-L2 interference) Subject is not grammatically obligatory in Chinese
English	By exercising, it can make you fit.

[7] Main Verb Reference	
Student's original sentence	She die one horse then this much cry no stop
Chinese	Very similar to the grammatical structure of the Chinese language (Language Typology and L1-L2 interference) The main verb in English often refers to the subject (i.e. the doer) in SPL, but this may not be always the case in TPL
Revised version after treatment	She had a horse die on her, and she cried a lot without stopping.

8.5.3 Theory of Markedness

The 'Theory of Markedness' argues that some language features (i.e. marked features like 'complementizer', 'tenses', 'articles', 'singular/plural form', 'verb form', 'preposition' etc.) are specific and unique to English but not present in Chinese, whereas some language features (i.e. unmarked features like noun and pronoun) are common to both English and Chinese (Li, 2010). Except 'Complementizer' which involves the use of dependent clauses and independent clauses which are absent in Chinese, the common linguistic features of these errors is that they are mostly

at word-level. What is worth noticing is that, according to the 'Error Count' summary results (see Appendix C), students from both treatment groups made very minimal 'noun'/'pronoun' errors, where its error frequency in their first draft was only 7 times and 3 times respectively in the experimental group and the control group. These findings suggested that these word-level unmarked features did not cause much difficulties to the students to handle in their first place, attributed to the fact that nouns and pronouns are semantically and syntactically obligatory in both Chinese and English as well. As for the marked features of English, students in this study made a range of errors mostly at 'Word-level' (see Appendix C). However, irrespective of the form of feedback they received, most 'Word-level' errors were successfully remedied in the subsequent draft (see Appendix C). These results also suggested that these word-level marked features appeared to be not causing much difficulties for students to comprehend, process and correct with indirect coded feedback.

Despite the observation above suggesting that 'L1-L2 Interference' and 'Linguistic Typology' seem to be highly contributed to 'Awkwardness' errors and 'Clausal-level' errors, while the 'Marked and Unmarked Features' of an linguistic item seem to be more highly related to 'Word-level' errors. It appears to me that the source of errors has little relevance as to what feedback practice should be adopted. Instead, it seems to be more relevant to question if a particular error type is treatable or untreatable according to Ferris's dichotomy of error types (2006). That is, if an error can be identified with and explained by grammatical rules, then indirect coded feedback is often a desirable option to go with; if an error cannot be identified with and explicitly explained by grammatical rules like 'Awkwardness' errors and 'Collocation' errors, then indirect coded feedback may not be equally effective in remedying these error types, which was statistically

supported by our findings. In the event of this, an alternative like the electronic feedback may probably be the solution for its provision of a recommended lexico-grammatical form in some new contexts through a web-based and interactive platform where ESL learners can deduce the subtle grammatical rule alongside the function of the target form in different contexts and then incorporate it into the writing of their own context. This subtle process of hypothesis making and testing approximated to Krashen's (1985) 'Input Hypothesis' resonates with Myhill, Lines and Watson (Making Meaning with Grammar, Metaphor, 2011) who argued that "studying how language works can make children more alert to the infinite possibilities of the English language, allow them to evaluate others' language use, and be in a better position to use it for themselves" (D., Waugh, 2014). Putting the key phrase 'Infinite possibilities of the English language' in the context of this study, this may imply there are always linguistics forms or patterns which are beyond any grammatical explanation. As such, this would pose some difficulties to ESL learners whose sense of security in their English repertoire stemming from their understanding and application of grammatical rules, but unfortunately not all linguistic forms or patterns are explainable.

8.6 *Pedagogical Outcomes: The Three-Step Approach to Error Correction within the Model of Framework for Developing Learner Autonomy*

With the primary interest in improving the effectiveness of teacher written feedback in error correction while cultivating students' autonomy in their language development, 'Mark My Words' ('MMWs'), the electronic feedback system, was designed in such a way to accommodate individual learners' language needs (by incorporating Nunan's (1997) 'Model of Framework for

Developing Learner Autonomy') and to be more responsive to various error types (by incorporating Krashen's (1985) 'Input Hypothesis'). This accommodation also extends to whether they are learning or acquiring a particular language item.

To improve students' comprehensibility and adoption of teacher feedback, 'Mark My Words' (MMWs) provides a feedback environment where students instead of teachers can exercise their autonomy to determine, according to their interlanguage level and learning needs, how much information and how explicit the teacher feedback on a particular error type they would like to receive. Hence, with its capacity of reconciling the above teacher's and student's limitations (see Truscott's (1996) arguments in Chapter 2.2), the electronic feedback has rendered teacher written feedback effective for student writing revisions.

It is already evident in the summary of Chapters 8.1, 8.2 and 8.3 that the pedagogical operation of Mark My Words ('MMWs') modeled on Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' can effectively (and statistically significantly) remedy errors across all categories.

8.6.1 Pedagogical Justifications: The Three-step Approach to Error Correction within the Model of Framework for Developing Learner Autonomy

The theoretical framework scaffolding the pedagogical operation of 'Mark My Words' ('MMWs') is a model postulated by Nunan (1997) in promoting learner autonomy, in the belief that 'developing some degree of autonomy is essential if learners are to become effective language

users' (p.192). By taking learners through level one to level three of Nunan's model, 'Mark My Words' ('MMWs') partially incorporates Nunan's notion of achieving autonomy into its pedagogical operation in the following fashion (see *Figure 8.1*):

Pedagogical Operation of 'Mark My Words' ('MMWs')		
Nunan's Theoretical Framework for cultivating learner autonomy		The 3-step Feedback Approach underpinned by the beliefs: (1) Role of Error Correction & (2) Krashen's Input Hypothesis
Level One:	Inform learners of the available feedback types	1. Error identification with a brief explanation of the error type shown in a pop-up window. 2. Provision of a recommended lexico-grammatical form or an example sentence in a new context through the web-based resources (e.g. Google link). 3. An electronic English Grammar Guide provides a more detailed explanation of the identified error type, followed by some practice
Level Two:	Learners make their own choice of the feedback type	
Level Three:	Learners act on the writing feedback	
Student can exercise their autonomy to follow or randomly select the step(s) according to their needs and interests		

Figure 8.1   Pedagogical Operation of 'Mark My Words' ('MMWs')

Within this three-step pedagogical framework, the 'Role of Error Correction' and 'Krashen's (1985) 'Input Hypothesis' are the second language theories which form the guiding principles in error correction, such as the types and the level of explicitness of error identification and explanation.

Allwright (1988) associates the idea of learner autonomy with a 'radical restructuring of language pedagogy' that involved 'the rejection of the traditional classroom and the introduction of wholly new ways of working' (p.35). According to Benson (2006), Allwright's idea of learner autonomy

refers to the 'deconstruction of conventional language learning classrooms' (Benson, 2006, p.22). Along similar lines, the advent of 'Mark My Words' ('MMWs') is therefore made analogous to such 'deconstruction' of conventional writing feedback environment (i.e. paper-based writing feedback), creating a more autonomous electronic feedback environment under which students can make their own decision as to the types and level of explicitness of the teacher written feedback they need. And such electronic feedback environment, with its interactive features, is also analogous to the situation defined by Dickinson (1987) where he described learner autonomy as a kind of 'situation where students are given an opportunity to decide on the learning options that best fit their needs, for example, their interlanguage level.

The merits of the pedagogical approach of 'Mark My Words' lies with its provision of the web-based interactive features where not only can learner autonomy help play down teachers' overt control over students' writing revision process by explicitly pointing out or even correcting their errors, but also it can help eliminate some possible teacher's 'blind spots' in ascertaining the students' level of interlanguage in the continuum in which teachers may encounter difficulties in deciding the appropriate level of explicitness to be given to individual students in their writing feedback. If students are to be given a choice as to the level of explicitness of error feedback they need, it is believed that student awareness of their own interlanguage level would inform them of the best judgement. With the electronic environment provided by 'Mark My Words' ('MMWs'), the onus of determining the kind of feedback and one's suitable level of explicitness can therefore lie with the learners themselves. The pedagogical operation of 'Mark My Words' ('MMWs') can facilitate this subtle process of hypothesis-making, as postulated by Krashen's (1985) 'Input Hypothesis', such that ESL learners can make sense of the new comprehensible input (i.e.

Exemplary sentence shown on a pop-up window and/or the recommended lexcio-grammatical form shown on Google Link), and then derive the correct rule governing its usage.

This study has no pre-determined stance concerning Krashen's (1985) view or his opponents' view about the level of explicitness of error feedback. The question this study has explored is how Krashen and his opponents evaluate if an error feedback is too explicit or implicit for an ESL learner as the level of explicitness in error feedback is a relative term to individuals. This is to say error feedback which seems clear and comprehensible to one learner might not be equally clear and comprehensible to others. This is due to individual variations on their interlanguage levels which suggest that individual ESL learners have their own unique interim rule-based linguistic system before their near-native mastery of English proficiency. This interim system determines how well individual ESL learners can comprehend, write, read, listen and speak the second language. As such, the conventional paper-based feedback might not be possibly cater to such individual variations on interlanguage as its feedback environment is unable to allow ESL learners to make their own choice but can only have their teachers to make the decisions on the kind of feedback and the level of explicitness of the error feedback for their students.

However, some may argue why error explanation and recommended lexcio-grammatical form could not be given on the paper-based feedback; and if it does, then this would make no difference from providing error explanation and exemplary sentences on the electronic feedback or on the paper-based feedback. However, the edge of the electronic feedback is that it can create a feedback environment which is more authentic, autonomous, institutive and interactive. These are the basic elements cultivating a situation where students are given learning options. As postulated by Nunan

(1997) in one of his assumptions underlying his model of framework for developing learner autonomy, 'developing some degree of autonomy is essential if learners are to become effective language users' (p.192). Also, a similar point has already been mentioned in the previous paragraph is that the onus of determining the kind of feedback and one's suitable level of explicitness lies with the teachers when the conventional paper-based feedback is adopted, but not the students.

8.6.2 *Pedagogical Operation: The Three-Step Approach to Error Correction within the Model of Framework for Developing Learner Autonomy*

Based on this research, the pedagogical operation of 'Mark My Words' (MMWs) underpinned by Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' was introduced and tested out for its effectiveness in remedying errors in student writing revisions. Such pedagogical operation is exhibited through the so called 'Three-Step Approach to Error Correction within the Model of Framework for Developing Learner Autonomy' which follows the sequence below:

Step 1: Error identification and a brief explanation in the pop-up window:

> A pop-up window which identifies the lexcio-grammatical error (i.e. error type) and suggests ways for error correction or improvement is shown when students move the cursor to the highlighted item on their received essay through the MS-word file.
>
> Example: I ***concern*** my grammar. ← Verb-Adjective Confusion: Is this a verb or adjective?

Step 2: Provision of a recommended lexico-grammatical form or an example sentence in new contexts through the web-based resources:

A link to 'Google Search' / 'Word Neighbor' is posted at the bottom of the pop-up window. Click on the 'Google Search' link (or 'Word Neighbors' link), then a recommended usage of the lexico-grammatical form or an example sentence will be presented in a different context. Students are encouraged to make a hypothesis about the recommended language structure / item in the new context and then may incorporate the recommended structure / item into their own context of writing.

Corpus from the web-based resources:
(1) In the long run, however, denying it or minimizing it, will be more damaging to you, other family members, and the person you are concerned about.
(2) Conservatives today are finding more reasons than ever to be concerned about the death penalty…
(3) The parents were concerned about the safety of the park because the sun shines into the eyes of the pitcher.

Step 3: An electronic English Grammar Guide provides a more detailed explanation of the identified error type, followed by some practice:

If students need more detailed explanation and practice on the recommended lexico-grammatical form, simply just click on "Click here for more advice and practice" and they will be diverted to the electronic "English Grammar Guide" (EGG) where it will provide the full and relevant details along with practice.

Students can exercise their autonomy to follow or intuitively select the step(s) according to their interlanguage level and learning needs. They need not take all 3 steps. In this way, a particular feedback strategy is not imposed on students, and students are given their own choice to determine how much information and how explicit of the teacher written feedback they would like to receive. As Ferris (1996, 1999) argued that students' level of progress in error correction varies depending on error types. This implies that, for example, less proficient students might take through all three steps above to tackle more difficult errors like 'Awkwardness' errors and 'Collocation' errors which are highly marked in English. Conversely, students can just take step one to tackle some simple errors mainly due to carelessness like 'missing article' or 'wrong preposition'.

Unlike the conventional paper-and-pen marking which sees the provision of teacher written feedback as an end in itself along the continuum of assessment, 'Mark My Words' ('MMWs') can be viewed as a cyclical or complementary practice in a teaching-learning-assessment cycle. To achieve this, not only does the electronic feedback guide students through error categorization, explanation and correction, but also it opens the door for an online tutorial where a student who finds himself in need of further explanation or practice in a particular error can choose to go over the corresponding page of the interactive 'English Grammar Guide' by just simply a click at the bottom of the feedback window (i.e. Step 3: "Click here for more advice and practice").

In addition, the electronic feedback can generate a statistical report diagnosing the category and frequency of errors. The statistic report about students' error pattern can be used as reference informing teachers and students about the effectiveness of the current teaching and learning practice, shedding lights on what further needs to be done in future teaching and learning. Hence,

this cyclical and complementary feedback practice is conducive to enhancing teaching and learning.

8.6.3   Second Language Theories vs. Learning Theory

The pedagogical operation of 'Mark My Words' (MMWs) underpinned by Nunan's (1997) 'Model of Framework for Developing Learner Autonomy' and Krashen's (1985) 'Input Hypothesis' synchronizes with the learning theory of Constructivism. Constructivism is based on the notion that learners construct their own knowledge based on the existing knowledge (Driscoll, 2005; Sewell, 2002). In a constructivist learning environment, the conditions for learning may include: (a) complex, realistic and relevant environments that incorporate authentic activity; (b) social negotiation; (c) multiple perspectives and multiple modes of learning; (d) ownership in learning; and (e) self-awareness in knowledge construction (Driscoll, 2005). Such conditions for learning to a certain extent resonates the feedback environment of 'MMWs' where students are (a) exposed to comprehensible input from authentic texts (via corpus from online resources), (b) tasked with negotiating the level of explicitness of the input themselves by (c) selecting their preferred feedback strategy. Students instead of teachers are expected (d) to take charge of their learning goals and (e) can devise their learning plan with the help of a statistical report diagnosing the category and frequency of errors made by students.

8.7 Limitations of the Study

This study has posed some limitations. The first one is the small number of participants in which a total of 62 second year engineering students in a local university participated in the study. The second one is the small number of essays analyzed. It is possible that with a much larger number of participants and essays, results might have been different and varied. Third, only two language teachers were tasked with marking the essays. Findings might have been different had another teacher marked the same essays. Fourth, the possible impact of markers using different coding. It can be seen in Appendix B that the coding for each error type adopted by Teacher A of the experimental group was used quite consistently, whereas the coding for each error type adopted by Teacher B of the control group was sometimes used differently (e.g. Teacher B either put the code 'form' above the erroneous 'verb form' or simply underlined the erroneous 'verb form' without any code in some situations). Despite the fact that efforts were geared to avoid individual factors by imposing marking principles (e.g. comprehensive marking, no explicit feedback, etc.), findings might have been different had Teacher B provided the indirect feedback for the same error types more consistently and systematically.

Due to the small sample size of the study, it might not be sound to make a generalization about the students' perceptions of the teacher written feedback, as well as the effectiveness of the electronic feedback when it is applied to other local university students in Hong Kong. As it is pointed out by Goldstein (2004, 2005), Hyland & Hyland (2006b) and Lee (2008a, 2008b), contextual factors like overall context of teachers' work, school culture, examination culture, teachers' beliefs and

practice, students' level of proficiency and motivation are all dynamic to students' responsiveness to teacher written feedback in use.

For further research, it is recommended that the effectiveness of the electronic feedback can be tested on students of different proficiency levels or different disciplines while minimizing other extraneous variables. It is also recommended that this study can be extended to a longitudinal one in which the effectiveness of the electronic feedback on various error categories can be tested out over a longer period of time.

Bibliography

Allwright, D. (1988). *Observation in the Language Classroom*. London: Longman.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, p. 369-406.

Anson, C. M. (2006). Can't touch this: Reflections on the servitude of computers as readers. Machine scoring of student essays. In P. Freitag Ericsson & R. Haswell (Eds.), *Machine Scoring of student essays* (p. 38-56). Logan, Utah: Utah State University Press.

Antwi, S. & Hamza, K. (2015). Qualitative and Quantitative Research Paradigms in Business Research: A Philosophical Reflection. *European Journal of Business and Management, 7(3),* 217-225.

Arndt, V. (1993). Response to writing: Using feedback to inform the writing process. In M. Brook & L. Walters (Eds.), *Teaching composition around the Pacific rim: Politics and pedagogy* (p. 90-116). Clevedon, UK: Multilingual Matters.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100*(3), p. 603-617.

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

Belcher, D. & Liu, J. (2004). Conceptualizing discourse/responding to text. *Journal of Second Language Writing, 13*, p.3-6.

Bell, J. & Waters, S. (2014). *Doing your research project: a guide for first-time researchers*. New York: McGraw Hill.

Benson, P. (2006). Autonomy in language teaching and learning. *Language Teaching, 40*, p. 21-40.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17* (2), p. 69-124.

Bitchener, J. & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31* (2), p.193-214.

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

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

Black et al. (2003). *Assessment for Learning- putting it into practice.* Maidenhead, U.K.: Open University Press.

Brannon, L. & Knoblauch, C. (1982). On students' rights to their own texts: A model of teacher response. *College Composition and Communication, 33*, p. 157-166.

Brewer, J. & Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage.

Brophy, J. E. (1981). Teacher Praise: A Functional Analysis. *Review of Educational Research, 51*(1), p. 5-32.

Brown, J. D. (1991). Statistics as a foreign language – Part 1: What to look for in reading statistical language studies. *TESOL Quarterly, 25*, p. 569-586.

Brown, J. D. (1998). *Understanding Research in Second Language Learning: A Teacher's Guide to Statistics and Research Design*. Cambridge Language Teaching Library.

Brown, H. D. (1994). Teaching by Principles: An Interactive Approach to Language Pedagogy. Englewood Cliffs, NJ: Prentice Hall Regents.

Brown, H. D. (2000). Principles of Language Learning and Teaching. England: Longman.

Bryman, A. (1998). Quantitative and Qualitative Research Strategies in Knowing the Social World. In T. May & M. Williams (eds.) *Knowing the Social World.* Open University Press, 138-156.

Bryman, A. (2001). *Social Research Methods*. Oxford: Oxford University Press.

Bulter, C. 1985. *Statistics in Linguistics*. Oxford. Blackwell.

Cai, J. T. (1998). The influence of Chinese topic prominent features on Chinese EFL learners' compositions. *Foreign Language Teaching and Research,* 30(4), 17-21.

Cai, Y. (2002). Topic structure transfer: An analysis of subject loss in English output]. *Foreign Language Education,* 23(4), 1-6.

Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly, 28*, p. 181-188.

Chan, A. Y. W. (2004). Syntactic transfer. Evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal, 88*(1), 56-74.

Chaney, S. J. (1999). *The effect of error types on error correction and revision*. Sacramento, CA: California State University, Department of English, M.A. thesis.

Chang et al. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning*, 21(3), 283-299.

Cheng, C. E. & Cheng, W. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology, 12*(2), 94-112.

Chen, J. F. (1997). Computer generated error feedback and writing process: A link [Electronic Version]. *TESL-EJ, 2*. Retrieved from http://tesl-ej.org/ej07/a1.html

Chen, Shu-hui Eileen. (1998). The development of referential strategies in the English narratives by Chinese children. Paper presented at the International Conference on Trends in Second Language Teaching and Learning, May 20-23. Ottawa, Canada.

Chen, Shu-hui Eileen. (2000). Children's L2 referential skills and its implication for the EFL classroom. Taipei: Taipei 臺北: National Taipei University of Education.

Chen, W. X. (2002). Collocation errors in the writings of Chinese learners of English. *Journal of PLA University of Foreign Languages*, *25*(1), 60-62.

Chen, Ying (1987). Han ying yin gao tixi bijiao (Contrastive study of Chinese-English sound pitch). *Xiandai Waiyu* (Modern Foreign Languages), no.1.

Chodorow, M., Gamon, M. & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing, 27*(3), 419-436.

Cogie, J., Strain, K., & Lorinskas, S. (1999). Avoiding the proofreading trap: The value of the error correction process. *The Writing Center Journal, 19*(2), p. 7-32.

Cohen, A. (1987). Student processing of feedback on their compositions. In A.L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (p. 57-69). Englewood Cliffs, N.J.: Prentice-Hall.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, A. D. & Cavalcanti, M.C. (1990). Feedback on written compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (p. 155-177). Cambridge: Cambridge University Press.

Cohen, A. D. & Robbins, M. (1976). Toward assessing interlanguage performance: The relationship between selected errors, learners' characteristics, and learners' expectations. *Language Learning, 26*(1), p.45–66.

Cohen et al. (2011). *Research Methods in Education*. London: Routledge.

Connors, R. J. & Lunsford, A.A. (1993). Teachers' Rhetorical Comments on Student Papers. *College Composition and Communication, 44*(2), p.200-223.

Conrad, S. & Goldstein, L. (1999). ESL student revision after teacher written comments: Texts, contexts and individuals. *Journal of Second Language Writing, 8*, p. 147-180.

Cowie, A. P. (1994). Phraseology. In Asher, R.E. & Simpson, J. (Eds.), *The encyclopedia of language and linguistics* (pp. 3168-71). Oxford: Pergamon Press.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. London: Sage.

Creswell J. W. (2005). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. New Jersey: Pearson Education Inc.

Cumming, A. (1985). Responding to the writing of ESL students. *Highway One, 8*, p. 58-78.

Deetz, S., (1996). Describing Differences in Approaches to Organization Science: Rethinking Burrell and Morgan and Their Legacy; *Organization Science, 7* (2), p. 191-207.

Diab, R.L. (2005). Teachers' and students' beliefs about responding to ESL writing: A case study. *TESL Canada Journal, 23*(1), p. 28-42.

Dickinson, L. (1987). *Self-instruction in language learning.* Cambridge: Cambridge University Press.

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment, 5*(1), p.1-35.

Dornyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal, 78*(3), p.273-284.

Driscoll, M. (2005). *Psychology of learning for instruction (3rd. ed.)*. Boston: Pearson.

Dulay, H. Burt, M. & Krashen, S. 1982. *Language Two.* New York: Oxford University Press.

Eginarlar, H. (1993). Student response to teacher feedback in EFL writing. *System, 21*(2), p. 193-204.

Elliot, S. & Mikulas, C. (2004). *The impact of MY Access!! Use on student writing performance: A technology overview and four studies. Paper presented at the Annual Meeting of the American Educational Research Association.*

Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press, 1997.

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

Evan, N., Hartshorn, J., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university matriculated ESL learners. *System, 39*, p. 229-239.

Fathman, A. K., Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second Language Writing* (pp. 178-190). Cambridge, UK: Cambridge University Press.

Farmer, R. & Sweeney, E. (1994). Self-access in Hong Kong: A square peg in a round hole? *Occasional Papers in Language Teaching 4* (ELT Unit: Chinese University of Hong Kong), p. 24-30.

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

Ferris, D. R. (1996). Teaching ESL composition students to become independent editors. *TESOL Journal, 4*(4), p. 18-22.

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

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*, p. 1-10.

Ferris, D. R (2002). *Treatment of error in second language student writing*. Ann Arbor: University of Michigan Press.

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

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

Ferris, D. R. (2006). Does error feedback help student writers? In Hyland, K. & Hyland, F. (Eds), *Feedback in second language writing: Contexts and issues* (p. 81-104). Cambridge: Cambridge University Press.

Ferris, D. R. (2015). Written corrective feedback in L2 writing: Connors & Lunsford (1988); Lunford & Lunsford (2008); Lalande (1982). *Language Teaching, 48*(4), p. 531-544.

Ferris, D. R. & Roberts, B. (2001). Error feedback in L2 writing class. How explicit does it need to be? *Journal of Second Language Writing, 10*(3), p. 161-184.

Ferris, D. R. & Hedgocok, J. S. (2005). Teaching ESL composition: Purpose, process, and practice (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Ferris, D.R., Pezone, S., Trade, C.R., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions and implications. *Journal of Second Language Writing, 6*, p. 155-182.

Frantzen, D. (1995). The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *Modern Language Journal, 79*, p. 244-329.

Frantzen D. & Rissell, D. (1987). Learner self-correction of written compositions: What does it show us? In B. VanPatten, T.R. Dvorak, & J.F. Lee (Eds.), *Foreign language learning: A research perspective* (p. 92-107). Cambridge: Newbury House.

Frodesen, J. (1991). Grammar in writing. In M. Celce-Murcia (Ed.), *Teaching English as a second language* (2nd ed., p. 264-276). Boston: Heinle & Heinle.

Gee, T.C. (1972). Students' responses to teacher comments. *Research in the Teaching of English, 6*, p. 212-221.

Genesee, F. & J. Upshur. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.

Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing, 13*(1), p. 63-80.

Goldstein, L. M. (2006). Feedback and revision in second language writing: Contextual, teacher, and student variables. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (p. 185-205). Cambridge: Cambridge University Press.

Goldstein, L. & Kohls, R. (2002). Writing, commenting and revising: The relationship between teacher feedback and student revision online. Paper presented at the American Association of Applied Linguistics Conference, Salt Lake City, UT.

Green et al. (2006). *Handbook of Complementary Methods in Education Research.* New Jersey: Lawrence Erlbaum Associates.

Grimes, D. (2008). Automated Writing assessment in the classroom. Pedagogies: An International Journal, 3, p.22-36.

Grimes, D. & Warschauer, M (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Language, and Assessment, 8*(6), 1-43.

Guba, E., & Lincoln, Y. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage.

Guenette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of

feedback on writing. *Journal of Second Language Writing, 16*(1), p.40-53.

Hammersley, M. (2012). *What is Qualitative Research?* London: Bloomsbury.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), p.81-112.

Hedge, T. (2000). Teaching and learning in the language classroom. Oxford: Oxford University Press.

Hedgcock, J. & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity in second language writing. *Journal of Second Language Writing, 3*, p. 141-163.

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

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

Henning, E., Van Rensburg, W. and Smit, B. (2004) *Finding Your Way in Qualitative Research.* Van Schaik Publishers, Pretoria.

Higgs, T. & Clifford, R. (1982). The push toward communication. In T. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (p. 57-79). Skokie, IL: National Textbook Company.

Hitchcock, G. and Hughes, D (1995). *Research and the Teacher: A Qualitative Introduction to School-based Research*. London: Routledge.

Holec, H. (1981). *Autonomy and foreign language learning*. Oxford: Pergamon.

Hou, Q. & Pramoolsook, I. (2012). Non-English Major EFL Learners' Collocation Errors in a Chinese Context. *The Journal of Teaching English with Technology, 12(3)*.

Hsin, Ai-li. (2003). *Studies on Contrastive Linguistics Between Chinese and English*. Taipei: Ying-jian Publisher.

Hung, T. (2000). '*Interlanguage analysis and remedial grammar teaching*'. Papers in Applied Language Studies, Vol.5, pp.155-168. Hong Kong Baptist University.

Huberty, C. (2002). *A History of Effect Size Indices*. London: Sage.

Hyland, K. (1993). ESL computer writers: What can we do to help? *System, 21*(1), 21-30.

Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing, 7*(3), p. 255-286.

Hyland, F. (2001). Providing effective support: Investigating feedback to distance language learner. *Open Learning, 16*(3), p. 233-247.

Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System, 31*, p. 217-230.

Hyland, F. (2010). Future directions in feedback on second language writing: overview and research agenda. *International Journal of English Studies, 10*(2), p. 171-182.

Hyland, F. & Hyland, K. (2001). Suggesting the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*, p. 185-212.

Hyland, K. & Hyland, F. (2006a). State-of-the-art review on "Feedback in second language students' writing". *Language Teaching, 39(2)*, p. 83-101.

Hyland, K. & Hyland, F. (2006b). Interpersonal aspects of response: Constructing and interpreting teacher written feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in ESL writing: Contexts and issues* (p. 206-224). Cambridge: Cambridge University Press.

Kepner, C. G. (1991). An Experiment in the Relationship of Types of Written Feedback to the Development of Second-Language Writing Skills. *The Modern Language Journal, 75*(iii).

Keh, C. (1990). A design for a process-approach to writing course. *English Teaching Forum, 28*(1), p. 10-12.

Kirk, J. & Miller, L. (1986) *Reliability and validity in qualitative research*. London: Sage.

Krashen, S. & Kendall, T. (1983). *The Natural Approach*. Pergamon.

Krashen, S. (1985). *The Input Hypothesis: Issues & Implications*. New York: Longman.

Komura, K. (1999). *Student response to error correction in ESL classrooms*. Unpublished master's thesis. Sacramento: California State University.

Krishnamurthy, S. (2005). A demonstration of the futility of using Microsoft Word's spelling and grammar check. faculty.washingston.edu/sandeep/check/. Retrieved September 4, 2005.

Lai, Y. H. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology, 41*(3), p. 432-454.

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

Lee, C. et al. (2009). Web-based essay critiquing system and EFL students' writing: A quantitative and qualitative investigation. *Computer Assisted Language Learning, 22*(1), p. 57-72.

Lee, C. et al. (2013). Immediate web-based essay critiquing system feedback and teacher follow up feedback on young second language learners' writings: An experimental study in a Hong Kong secondary school. *Computer Assisted Language Learning, 26,* 39-60.

Lee, I. (1997). Peer reviews in a Hong Kong tertiary classroom. *TESL Canada Journal, 15*(1), p. 58-69.

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

Lee, I. (2005). Error correction in the L2 writing classroom: What do students think? *TESL Canada Journal, 22*(2), p. 1-16.

Lee, I. (2007). Assessment for Learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *Canadian Modern Language Review, 64(*1), p. 199-214.

Lee, I. (2008a). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. Journal of Second Language Writing, 17(2), p. 69-85.

Lee, I. (2008b). Student reactions to teacher feedback in two Hong Kong secondary classrooms. Journal of Second Language Writing, 17(3), p. 144-164.

Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second Language Writing* (p. 57-68). Cambridge, UK: Cambridge University Press.

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

Leki, I. (2006). ''You cannot ignore'': Graduate L2 students' experience of and response to written feedback practices within their disciplines. In K. Hyland & F. Hyland (Eds.), *Feedback in ESL writing: Contexts and issues*. Cambridge: Cambridge University Press.

Li, C. N. & Thompson, S. A. (1976). Subject and topic: A new typology of language, 459-471. http://angli02.kgw.tu-berlin.de/Korean/Artikel03/

Li, X. (2010). The Influence of Chinese Topic-Prominent Features on the Acquisition of Four Types of English Subjects. *Chinese Journal of Applied Linguistics, Vol. 33*(5), p. 80-94.

Lightbown, Patsy M. and Nina Spada. (1999). *How Languages Are Learned (Revised edition). Oxford Handbooks for Language Teachers*. Oxford University Press.

Liu, A. (2002). A Corpus-based Lexical Semantic Investigation of VN Miscollocations in Taiwan Learners' English. Master Thesis, Tamkang University, Taiwan.

Loewen, S. & Erlam, R. 2006. Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning 19*(1), p. 1-14.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (p. 438-468). San Diego, CA: Academic Press.

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition, 28*, p. 269-300.

Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19*(1), p. 37-66.

MacDonald, R. B. (1991). Developmental students' processing of teacher feedback in composition instruction. *Review of Research in Development Education, 8*(5), p.1-5.

Malligamas, P. & Pongpairoj, N. (2005). Thai learners' knowledge of English collocations. *HKBU Papers in Applied Language Studies*. 9, 1-28.

Mantello, M. (1997). Error correction in the L2 classroom. *Canadian Modern Language Review, 54*, p. 127-131.

Merriam, S. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass Publishers.

Martin-Beltran, M. & Chen, P. J. (2013). From monologue to dialogue: a case study on mediated feedback in a transitional asynchronous online writing tutorial. *Academic Exchange Quarterly, 17*(1), p. 145-150.

Maxwell, J. (2013). *Qualitative Research Design: An Interactive Approach (3rd ed.).* London: Sage.

Milton, J. (2006). Resource-rich Web-based feedback: Helping learners become independent writers. In Hyland, K. & Hyland, F. (Eds), *Feedback in second language writing: Contexts and issues* (p. 123-139). Cambridge: Cambridge University Press.

Muijs, D. (2011). *Doing Quantitative Research in Education with SPSS* (2nd Ed.). London: Sage.

Nunan, D. (1997). Designing and adapting materials to encourage learner autonomy. In P.Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (p. 192-203). London: Longman.

Neuman, W. L. (2003) *Social Research Methods: Qualitative and Quantitative Approaches (5th ed.)*. Boston: Allyn and Bacon.

Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33*(7), 14-26.

Oladejo, J. A. (1993). Error correction in ESL: Learners' preferences. *TESL Canada Journal,*

10(2), p. 71-89.

Paulus, T. (1999).The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing, 8*(3), p. 265–289.

Philips, S. M. (2007). *Automated essay scoring: A literature review (SAEE research series #30).* Kelowna, BC: Society for the Advancement of Excellence Education.

Pennington, M. C. (1992). Beyond off-the-shelf computer remedies for student writers: Alternatives to canned feedback. *System, 20*(4), p. 423-447.

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

Reid, J. (1994). Responding to ESL students' text: The myth of appropriation. *TESOL Quarterly, 28*, p. 272-292.

Rennie, C., 2000. *Error feedback in ESL writing classes: What do students really want?* Unpublished master's thesis. Sacramento: California State University.

Riazi, A. (1997). Acquiring disciplinary literacy: A socio-cognitive analysis of text production and learning among Iranian graduate students of education. *Journal of Second Language Writing, 6*(2), p. 105-137.

Robb, T., Ross, S., & Shorteerd, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*, p. 83-93.

Roberts, B. J. (1999). *Can error logs raise more than consciousness? The effects of error logs and grammar feedback on ESL students' final drafts*. Unpublished master's thesis, California State University, Sacramento.

Roos, B., & Hamilton, D. (2005). Formative Assessment: A cybernetic viewpoint. *Assessment in Education, 12*(1), p. 7-20.

Saito, H. (1994). Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal, 11*(2), p. 46-70.

Samburskiy, D. & Quah, J. (2014). Corrective feedback in asynchronous online interaction: developing novice online language instructors. *CALICO Journal, 31*(2), p. 158-178. http://dx.doi.org/10.11139/cj.31.2.158-178.

Sarantakos, S. (2005). *Social Research (3rd ed.).* Melbourne: Macmillan Education.

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

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

Schneider, D & McCoy, K.F. (1998). Recognizing syntactic errors in the writing of second language learner. *Proceedings of the 36th Conference on Association for Computational Linguistics*, Vol. 2. Montreal, Quebec.

Schroeder, J. A., Grohe, B., & Pogue, R. (2008). The impact of criterion writing evaluation technology on criminal justice student writing skills. *Journal of Criminal Justice Education, 19*(3), 432-445.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagn & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, p. 39-83. Chicago, IL: Rand McNally.

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

Sewell, A. (2002). Constructivism and student misconceptions: why every teacher needs to know about them. *Australian Science Teachers' Journal 48*(4). 24-28. Retrieved May 21, 2006, from ProQuest database.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*, p. 103-110.

Sommers, N. (1996). Revision strategies of student writers and experienced adult writers. *College Composition and Communication, 31*(4), p.378-388.

Stern, L. A. & Solomon, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing, 11*, p. 22-41.

Stevenson, M. & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, p.51-65.

Silva, T. (1990). Second language composition instruction: developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language Writing Research: Insights for the classroom* (p. 11-17). New York: Cambridge University Press.

Storch, N. & Wigglesworth, G. (2010). Learner' processing, uptake, and retention of corrective feedback on writing: case studies. *Studies in Second Language Acquisition, 32*(2), p. 303-304.

Straub, R. (2000). The student, the text, and the classroom context: A case study of teacher response. *Assessing Writing, 7*, p.23-55.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (p. 63-81). Cambridge: Cambridge University Press.

Terre Blanche, M. & Durrheim, K.(Eds.) (1999). *Research in Practice. Applied methods for the social sciences.* Cape Town: University of Cape Town Press.

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

Truscott, J., & Hsu, Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*, 292-305. doi:10.1016/j.jslw.2008.05.003

Ulin, R., Robinson, T. & Tolley, E. (2004). *Qualitative Methods in Public Heath: A Field Guide for Applied Research*. Wiley.

Varnosfadrani, A. & Basturkmen, H. (2009). The Effectiveness of Implicit and Explicit Error Correction on Learners' Performance. System, 37(1), p. 82-98.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Ene, E., & Upton, T. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System, 46*, p. 80-95. http://dx.doi.org/10.1016/j.system.

Wand, Y. & Weber, R. (1993) On the ontological expressiveness of information system analysis and design grammars. *Journal of Information Systems, (3):217–237.*

Wang, F. & Wang, S. (2012). A comparative study on the influence of automated evaluation system and teacher grading on students' English writing. Procedia *Engineering, 29*, 993-997.

Warden, C. A. (2000). EFL business writing behavior in differing feedback environments. *Language Learning, 50*(4), 573-616.

Ware, P. & Warschauer, M. (2006). Electronic feedback and second language writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts & issues* (p. 105-122). Cambridge: Cambridge University Press.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*, 57-71.

White, R. and Arndt, V. (1991). *Process Writing*. London: Longman.

Wiliam, D. (2001). An overview of the relationship between assessment and the curriculum. In D. Scott (Ed.), *Curriculum and assessment* (p. 165-181). Westport, CT: Ablex Publishing.

Willetts, K. (1992). Technology and second language learning (Information analysis-ERIC Clearinghouse Products (071) EDO-FL-92-07). Washington, D. C.: Eric Clearinghouse on Languages and Linguistics.

Winne, P. H., & Butler, D. L. (1994). Student cognition in learning from teaching. In T. Husen & T. Postlewaite (Eds.), *International encyclopaedia of education* (2nd ed., p. 5738–5745). Oxford, UK: Pergamon.

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, p.179–200.

Yorke, M. (2003). *Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice.* Higher education. Springer.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 21*, p. 697-715.

Zourou, K. (2011). Towards a typology of web-based error correction moves in an asynchronous distance language learning environment. In W.M. Chan, K. N. Chin, & M. Nagami (Eds.), Studies in second and foreign language education: *Media in foreign language teaching and learning* (p. 651-668). Berlin, DEU: Walter De Gruyter.

Appendix A: 77 Error Types under 7 Error Categories

	Error Categories	Error Types
1.	"Awkwardness" error	Non-treatable clausal-level error (e.g. Chinglish)
2.	"Clausal-level" error	Mixed construction; Imbalanced Structure; Sentence fragment; Run-on sentence; Wrong relative pronoun reference
3.	"Word-level" error	Redundant determiner; Missing determiner; Misuse determiner; Subject-verb agreement; Redundant word; Tense error; Singular-Plural form; Wrong Preposition; Voice; Punctuation; Wrong word/expression; Wrong verb form; Adjective-adverb confusion; Noun-Adjective confusion; Verb-Noun confusion; Adjective-verb confusion; Missing auxiliary verb; Misuse auxiliary verb; Redundant preposition; Missing preposition; Missing noun/pronoun; Missing possessive; Finite vs. Non-finite verbs; Misuse of 'besides'; 'Comparing' vs. 'Compare'; Redundant conjunction; Wrong adverb/connective; Wrong adverb form; Intransitive verb; Missing adverb/connective; Misuse 'concern'; Missing verb; Wrong form; Missing word; Misuse of 'on the other hand'; Redundant adjective; 'Be'-'Do' confusion; 'Be'-'Have' confusion; Auxiliary or Non-finite; Capitalization; Wrong relative pronoun; Misuse of 'consist'; Subordinate clause/conjunction; Missing infinitive-'to'; 'Due to' vs. 'Because'; 'Word-order in question'; 'Negative form'; Misuse of 'on the contrary'; Superlative confusion; Spelling.
4.	"Collocation" error	Wrong Word Collocation (including wrong preposition).
5.	"Tone & Style" error	Inappropriate register; Level of courtesy/ confidence; Cliché.
6.	"Content" error	Too general / unclear; Logic flaw; Cause & Effect problem; Irrelevance; Inadequate support; Ambiguous.
7.	"Organization" error	Choppy ideas (cohesion); Choppy ideas across paragraphs (coherence); Missing / Inadequate topic sentence; New sentence; New paragraph; Inadequate / Missing a thesis statement; Unclear topic sentence; Missing / Inadequate scene setting; Supporting details do not follow the topic sentence; Reasoning is difficult to follow; Missing/Ineffective conclusion.

Appendix B:
Standardization of Electronic Error Codes and Paper-based Error Codes

	Electronic Error Code	**Paper-based Error Code**		
C	Effective Introduction	Good opening		
C	Decent Introduction	'OK Intro'		
C	Ineffective Introduction (thesis statement)	'No purpose statement'	Weak opening	
C	Claims lack adequate support	'could be more informative'		
C	Unclear / Incomprehensible	Underline the problematic chunk, and place a superscript '?' above the chunk	'Unclear?'	'No specific focus'	'What do you mean'
C	Logic (Cause-Effect)?	'?'	'wrong word'	
C	Good Point	✓		
L	Wrong Form	'form'	'wrong form'	
L	Wrong Verb Form	'form'	'wrong form'	'Underline' the word'
L	A verb cannot be a subject of a sentence	'underline the 'word'		
L	Wrong Word/Unclear Word	'✗'	'wrong word'	'Underline' the word'
L	Wrong Lexical Collocation	'✗'	'wrong word'	'?'
L	Wrong Preposition Collocation	'✗ prep'	'wrong prep'	'prep'
L	Wrong Preposition	'✗'	'Underline' the word'	
L	Wrong Article	'✗'		
L	Redundant Word(s)	'✗'	'Underline' the word'	
L	Redundant Preposition	'✗ prep'	'✗'	'Underline' the word'
L	Redundant Determiner Redundant Verb/Auxiliary	'✗'	'Underline' the word'	
L	Redundant Verb/Auxiliary	'✗'	'Underline' the word'	
L	Redundant Conjunction	'✗'	'Underline' the word'	
L	Spelling	'sp'		

L	Capitalization	'Circle' the problematic letter'		
L	Missing Word	∧		
L	Missing Article	∧ 'A'	∧	
L	Missing Verb	∧ 'v'	∧	
L	Missing Auxiliary Verb	∧ 'v'	∧	
L	Missing Preposition	∧ 'Prep'	∧	
L	Missing 'to' infinitive	∧		
L	Missing Noun / Pronoun	∧		
L	Missing Adverb	∧		
L	Singular vs. Plural Form	'N'	'Underline' the redundant 's''	
L	Adjective-Noun Confusion	'form'	'Adj'	'Noun'
L	Finite vs. Non-finite Verbs	'wrong verb'	'wrong form'	'
L	Subject-Verb Agreement	'Ag'		
L	Wrong Tense	'T'		
L	Active-Passive Voice	'voice'	'form'	
L	Direct & Indirect Quotation	'✗'	'Underline' the problematic chunk'	
L	Wrong Relative Pronoun Reference	'Underline' the relative pronoun with an '✗' inserted above		
L	Awkward Expression	'Awkward'	'Underline' the problematic expression'	
L	Run-on Sentence (Comma Spice)	'Run-on sentence'	'Circle' the comma'/ 'Underline' the problematic chunks of words'	
L	Sentence Fragment	'Incomplete sentence'	'Circle'/ 'Underline' the problematic chunks of words'	
L	Imbalance Structure	'wrong structure'		
L	Mixed Structure	'paraphrase it'	'Noun phrase'	
L	Relative clause to modify the meaning in the main clause	'Relative Clause'	'link the ideas'	
L	Relative clause or Participle phrase to modify the main clause	'Relative Clause' or 'Participle Phrase'	'form'	
L	Using 'that' to begin a subordinating clause	∧		
L	Inappropriate Register (level of formality)	'Underline' / 'Circle' the problematic chunk	'No question form in a formal article'	
L	Misuse of 'Concern' / 'Benefit' / 'Besides' / 'On the other hand'	'wrong word'	'✗'	
L	Good Usage	✓		
O	New Sentence	'New Sentence'		
O	New Paragraph	'New Paragraph'		
O	Lack of focus across the paragraphs (Coherence)	'jump?'		

O	Lack of transitions within the paragraph (cohesion)	'poor linkage'
O	Punctuation	∧ 'Punctuation'

Appendix C: Relative Frequency of Errors by Category

	Experimental Group (Total Errors made in Draft 1 & Draft 2)				
	Arranged in descending order by the frequency of errors in Draft 1				
	Error Types	**Error Categories**	**Draft 1**	**Draft 2**	**% Reduction**
1	Awk / Unclear Expression	Awkwardness	47	5	89%
2	Sngular-Plural Form	Word-level	43	7	84%
3	Choppy Ideas (Cohesion)	Organization	28	6	79%
4	Missing Determiner	Word-level	26	3	88%
5	Sentence Fragment	Clausal-level	25	2	92%
6	Wrong Verb Form	Word-level	23	5	78%
7	Wrong Word Collocation	Collocation	22	0	100%
8	Redundant Word	Word-level	22	1	95%
9	Run-on Sentence	Clausal-level	21	0	100%
10	Imbalanced Structure	Clausal-level	19	1	95%
11	Inappropriate Register	Tone & Style	19	1	95%
12	Spelling	Word-level	18	0	100%
13	Subject-Verb Agreement	Word-level	18	0	100%
14	To General / Unclear	Content	16	2	88%
15	Redundant Determiner	Word-level	15	0	100%
16	Wrong Preposition	Word-level	15	1	93%
17	Punctuation	Word-level	15	3	80%
18	Tense Error	Word-level	13	1	92%
19	Voice	Word-level	13	2	85%
20	Wrong Word / Expression	Word-level	11	3	73%
21	Inadequate Topic Sentence	Organization	10	0	100%
22	Missing Preposition	Word-level	10	1	90%
23	Supporting Details not follow the TS	Organization	9	0	100%
24	Noun-Adjective Confusion	Word-level	9	3	67%
25	Redundant Preposition	Word-level	8	0	100%
26	New Sentence	Organization	7	0	100%
27	Missing Noun / Pronoun	Word-level	7	0	100%
28	Misuse Auxiliary Verb	Word-level	7	0	100%
29	Wrong Relative Pronoun Reference	Clausal-level	6	0	100%
30	Missing Auxiliary Verb	Word-level	6	1	83%
31	Missing Possessive	Word-level	6	1	83%
32	Redundant Conjunction	Word-level	6	3	50%
33	Logic	Content	5	0	100%
34	Choppy Ideas across Paragraphs	Organization	5	0	100%
35	Adjective-Adverb Confusion	Word-level	5	0	100%
36	Finite vs. Non-finite Verbs	Word-level	5	0	100%
37	Misuse of 'Besides'	Word-level	5	0	100%
38	Ineffective Intro (Thesis Statement)	Organization	4	1	75%
39	Comparing' vs. 'Compare'	Word-level	4	0	100%
40	Wrong Adverb / Connective	Word-level	4	1	75%
41	Cause & Effect Problem	Content	3	0	100%

#	Error Types	Error Categories	Draft 1	Draft 2	% Reduction
42	Ineffective Conclusion	Content	3	0	100%
43	New Paragraph	Organization	3	0	100%
44	Misuse / Missing Determiner	Word-level	3	0	100%
45	Misuse of 'Concern'	Word-level	3	1	67%
46	Missing Adverb / Connective	Word-level	3	3	0%
47	Missing Verb	Word-level	3	0	100%
48	Wrong Form	Word-level	3	0	100%
49	Ineffective Intro (Scene Setting)	Organization	2	0	100%
50	Line of Reasoning difficult to follow	Organization	2	0	100%
51	Missing Word	Word-level	2	0	100%
52	Misuse of 'On the other hand'	Word-level	2	0	100%
53	Redundant Adjective	Word-level	2	0	100%
54	Be-Do Confusion	Word-level	2	0	100%
55	Verb-Noun Confusion	Word-level	2	1	50%
56	Redundant Verb, Auxiliary or Non-Finite	Word-level	2	0	100%
57	Capitalization	Word-level	2	0	100%
58	Word Order in Question	Word-level	2	2	0%
59	Mixed Construction	Clausal-level	1	0	100%
60	Relevance	Content	1	0	100%
61	Claims or Ideas without Adequate Support	Content	1	1	0%
62	Ambiguous	Content	1	0	100%
63	Inappropriate Tone	Tone & Style	1	0	100%
64	Cliché	Tone & Style	1	0	100%
65	Misuse / Missing Subordinating Clause/Conj	Word-level	1	0	100%
66	Adjective-Verb Confusion	Word-level	1	0	100%
67	Be-Have Confusion	Word-level	1	0	100%
68	Missing Infinitive 'to'	Word-level	1	0	100%
69	due to' vs. 'because'	Word-level	1	0	100%
70	Negative Form	Word-level	1	0	100%
71	Wrong Adverb Form	Word-level	1	0	100%
72	Instransitive Verb	Word-level	1	0	100%
73	Unclear Topic Sentence	Organization	0	0	N/A
74	Wrong Relative Pronoun	Word-level	0	0	N/A
75	Misuse of 'Consist'	Word-level	0	0	N/A
76	Misuse of 'on the contrary'	Word-level	0	1	-100%
77	Comparative vs. Superlative Confusion	Word-level	0	1	-100%

Control Group (Total Errors made in Draft 1 & Draft 2)					
Arranged in descending order by the frequency of errors in Draft 1					
	Error Types	**Error Categories**	**Draft 1**	**Draft 2**	**% Reduction**
1	Singular-Plural Error	Word-level	77	32	58%
2	Awk / Unclear Expression	Clausal-level	72	53	26%
3	Too General / Unclear	Content	34	11	68%
4	Wrong Verb Form	Word-level	34	10	71%
5	Missing Determiner	Word-level	30	20	33%
6	Inappropriate Register	Tone & Style	29	10	66%

7	Run-on Sentence	Clausal-level	24	12	50%
8	Sentence Fragment	Clausal-level	23	9	61%
9	Subject-Verb Agreement	Word-level	21	9	57%
10	Wrong Preposition	Word-level	19	9	53%
11	Wrong Word / Expression	Word-level	19	1	95%
12	Tense Error	Word-level	17	5	71%
13	Noun-Adjective Confusion	Word-level	17	3	82%
14	Spelling	Word-level	16	5	69%
15	Wrong Word Collocation	Collocation	14	13	7%
16	Voice	Word-level	13	8	38%
17	Redundant Determiner	Word-level	11	3	73%
18	Mixed Construction	Clausal-level	10	15	-50%
19	Imbalanced Structure	Clausal-level	10	6	40%
20	Wrong Form	Word-level	9	0	100%
21	Wrong Relative Pronoun Reference	Clausal-level	8	3	63%
22	Ineffective Intro (Thesis Statement)	Organization	7	3	57%
23	Punctuation	Word-level	7	2	71%
24	Adjective-Adverb Confusion	Word-level	7	3	57%
25	Missing Auxiliary Verb	Word-level	7	7	0%
26	Choppy Ideas (Cohesion)	Organization	6	4	33%
27	Finite vs. Non-finite Verbs	Word-level	6	5	17%
28	Choppy Ideas across Paragraphs	Organization	5	2	60%
29	Missing Preposition	Word-level	5	4	20%
30	Misuse / Missing Determiner	Word-level	5	1	80%
31	Logic	Content	4	3	25%
32	Redundant Conjunction	Word-level	4	4	0%
33	Verb-Noun Confusion	Word-level	4	3	25%
34	Redundant Verb, Auxiliary or Non-finite	Word-level	4	0	100%
35	Claims / Ideas without Adequate Support	Content	3	0	100%
36	Redundant Word	Word-level	3	2	33%
37	Missing Noun / Pronoun	Word-level	3	0	100%
38	Adjective-Verb Confusion	Word-level	3	3	0%
39	Be-Have Confusion	Word-level	3	1	67%
40	Word Order in Question	Word-level	3	0	100%
41	New Paragraph	Organization	2	0	100%
42	Redundant Preposition	Word-level	2	4	-100%
43	Misuse of 'On the other hand'	Word-level	2	1	50%
44	Be-Do Confusion	Word-level	2	5	-150%
45	Capitalization	Word-level	2	0	100%
46	Cause & Effect Problem	Content	1	0	100%
47	New Sentence	Organization	1	0	100%
48	Ineffective Intro (Scene Setting)	Organization	1	0	100%
49	Missing Verb	Word-level	1	3	-200%
50	Missing Word	Word-level	1	1	0%
51	Missing Infinitive 'to'	Word-level	1	0	100%
52	Wrong Adverb Form	Word-level	1	0	100%

53	Intransitive Verb	Word-level	1	0	100%
54	Relevance	Content	0	0	N/A
55	Ambiguous	Content	0	0	N/A
56	Ineffective Conclusion	Content	0	0	N/A
57	Inadequate Topic Sentence	Organization	0	0	N/A
58	Unclear Topic Sentence	Organization	0	0	N/A
59	Supporting Details not follow the TS	Organization	0	0	N/A
60	Line of Reasoning difficult to follow	Organization	0	0	N/A
61	Inappropriate Tone	Tone & Style	0	0	N/A
62	Cliché	Tone & Style	0	0	N/A
63	Misuse Auxiliary Verb	Word-level	0	0	N/A
64	Missing Possessive	Word-level	0	0	N/A
65	Misuse of 'Besides'	Word-level	0	0	N/A
66	Comparing' vs. 'Compare'	Word-level	0	0	N/A
67	Wrong Adverb / Connective	Word-level	0	0	N/A
68	Misuse of 'Concern'	Word-level	0	0	N/A
69	Missing Adverb / Connective	Word-level	0	0	N/A
70	Redundant Adjective	Word-level	0	0	N/A
71	Wrong Relative Pronoun	Word-level	0	0	N/A
72	Misuse of 'Consist'	Word-level	0	0	N/A
73	Misuse / Missing Subordinating Clause/Conj	Word-level	0	0	N/A
74	due to' vs. 'because'	Word-level	0	0	N/A
75	Negative Form	Word-level	0	0	N/A
76	Misuse of 'On the contrary'	Word-level	0	0	N/A
77	Comparative vs. Superlative Confusion	Word-level	0	6	-600%

Appendix D: **STUDENT QUESTIONNAIRE**   Section # _____

> The purpose of this study is to investigate students' perceptions of teacher written feedback in improving the overall accuracy and appropriateness of student writing in the areas of the:
> (11) Legibility of feedback
> (12) Student ability to correct errors
> (13) Types of feedback (i.e. grades / error feedback / written comments)
> (14) Focus of error feedback (i.e. content / organization / language)
> (15) Amount of error feedback
> (16) Focus of written comments (i.e. content / organization / language)
> (17) Types of error feedback strategy (i.e. explicitness of error feedback: direct vs. indirect feedback)
> (18) Students' responsiveness of error types (i.e. easier / more difficult to understand & correct)
> (19) Types of post-writing activity
> (20) Usefulness of feedback
> It is hoped that this study can contribute to our understanding of the ways students react to and act on teacher written feedback in LANG 206. Thank you very much for your participation.

1.   Was your language instructor's feedback **legible (understandable)**? (Please circle only **ONE** answer) ('3' = Partially)

5	4	3	2	1
Totally < - ------------------------------------- Partially -----------------------------> Not Legible at all				

Comments: _____
_____

2.   To what extent were you able to **correct the error accurately** according to the language instructor's feedback? (Please circle only **ONE** answer) ('3' = Some)

5	4	3	2	1
Totally < - --------------------------------------- Some --------------------------------> Not Legible at all				

Comments: _____
_____

3.   In future compositions, which of the following **types of feedback** do you prefer to get from the language instructor? (Please tick only **ONE** answer)

	Types of Feedback	**Preference**
a.	Only Grades	
b.	Only Response to Errors (i.e. error feedback)	
c.	Only Written Comments	
d.	Grade + Error Feedback	
e.	Grade + Written Comments	
f.	Error feedback + Written Comments	
g.	Grade + Error Feedback + Written Comments	
h.	None of the Above	

Comments: _____
_____

4. Which of the following **types of feedback** would you like your language instructor to give **more** in future? (Please tick only **ONE** answer)

	Types of Feedback	Preference
a.	Error Feedback	
b.	Written Comments	
c.	None of the above	

Comments: _____
_____

5. Which of the following **types of feedback** would you like your language instructor to give **less** in future? (Please tick only **ONE** answer)

	Types of Feedback	Preference
a.	Error Feedback	
b.	Written Comments	
c.	None of the above	

Comments: _____
_____

6. In future compositions, which of the following **types of feedback** would you be most interested in receiving? (Please tick only **ONE** answer)

	Types of Feedback	Preference
a.	The grade	
b.	Teacher's comments on my writing	
c.	The errors I have made	
d.	Others (Please specify)	

Comments: _____
_____

7. Which of the following **focus of error feedback** would you like your language instructor to emphasize **more** in future? (Please tick only **ONE** answer)

	Focus of Error Feedback	Preference
a.	Content	
b.	Organization (e.g. flow of ideas, transitions, paragraphing)	
c.	Language (e.g. grammar, vocabulary, sentence pattern)	
d.	None of the above	
e.	Others (Please specify)	

Comments: _____
_____

8.   Which of the following **focus of error feedback** would you like your language instructor to emphasize **less** in future? (Please tick only **ONE** answer)

	Focus of Error Feedback	Preference
a.	Content	
b.	Organization (e.g. flow of ideas, transitions, paragraphing)	
c.	Language (e.g. grammar, vocabulary, sentence pattern)	
d.	None of the above	
e.	Others (Please specify)	

Comments: _____
_____

9.   Tick **ONE** box below to indicate the **amount of error** you would like your language instructor to respond to.

	Amount of Error	Preference
a.	All	
b.	Some only	
c.	None	

Comments: _____
_____

10. Which of the following **focus of written comments** would you like your language instructor to emphasize **more** in future? (Please tick only **ONE** answer)

	Focus of Error Feedback	Preference
a.	Content as a whole	
b.	Organization as a whole (e.g. flow of ideas, transitions, paragraphing)	
c.	Language as a whole (e.g. grammar, vocabulary, sentence pattern)	
d.	None of the above	
e.	Others (Please specify)	

Comments: _____
_____

11. Which of the following **focus of written comments** feedback would you like your language instructor to emphasize **less** in future? (Please tick only **ONE** answer)

	Focus of Error Feedback	Preference
a.	Content as a whole	
b.	Organization as a whole (e.g. flow of ideas, transitions, paragraphing)	
c.	Language as a whole (e.g. grammar, vocabulary, sentence pattern)	
d.	None of the above	
e.	Others (Please specify)	

Comments: _____
_____

12. Which of the following **error feedback strategies** would you like your language instructor to use **more** in future when responding to errors? (Please tick **ONE** answer)

	Error Feedback Strategies	Preference
a.	**Underline / Circle my errors** *e.g. ... stop using pesticides due to it caused pollution.*	
b.	**Underline / Circle my errors + Categorize my errors** *e.g. ... stop using pesticides due to it caused pollution.* [noun phrase]	
c.	**Underline / Circle my errors + Provide Corrections** *e.g. ... stop using pesticides due to it caused pollution.* (its environmental unfriendliness)	
d.	**Underline / Circle my errors + Provide example sentences in another context** *e.g. ... stop using pesticides due to it caused pollution.* (People reduce the use of fossil fuels due to its environmental unfriendliness.]	
e.	**Underline / Circle my errors + Categorize my errors + Explain my errors** *e.g. ... stop using pesticides due to it caused pollution.* [noun phrase] ('due to' has to be followed by a noun phrase]	
f.	**Underline / Circle my errors + Categorize my errors + Provide Corrections** *e.g. ... stop using pesticides due to it caused pollution.* [noun phrase] (its environmental unfriendliness)	
g.	**Underline /Circle my errors + Categorize my errors + Provide example sentences in another context** *e.g. ... stop using pesticides due to it caused pollution.* [noun phrase] [People reduce the use of fossil fuels due to its environmental unfriendliness.]	
h.	**Underline / Circle my errors + Categorize my errors + Explain my errors + Provide Corrections** e.g. ... *stop using pesticides due to it caused pollution.* [noun phrase] ('due to' has to be followed by a noun phrase) (environmental unfriendliness)	
i.	**Underline / Circle my errors + Categorize my errors + Explain my errors + Provide example sentences in another context** e.g. ... *stop using pesticides due to it caused pollution.* [noun phrase] ('due to' has to be followed by a noun phrase) (People reduce the use of fossil fuels due to its environmental unfriendliness.)	
j.	**Underline / Circle my errors + Categorize my errors + Explain my errors + Provide example sentences + Provide Corrections** *e.g. ... stop using pesticides due to it caused pollution.* [noun phrase] ('due to' has to be followed by a noun phrase) (People reduce the use of fossil fuels due to its environmental unfriendliness.) (environmental friendliness)	
k.	None of the above	
l.	Others (Please specify)	

Comments: _____
_____

13. Based on your answer in Q. 12, do you prefer making your own decision on the **explicitness of error feedback** (i.e. the types of error feedback strategy provided in Q. 12) you would like to receive on various error types from your language instructor? (Please circle only **ONE** answer) ('3' = Neutral)

5	4	3	2	1
Most preferred < ------------------------------- Neutral -------------------------------→ Not preferred				

Comments: _____
_____

14. Which of the following **types of error** indicated in the feedback is/are **more difficult to understand**? You can tick a **maximum of 2** boxes.

	Types of Error	**Preference**
a.	Wrong Word / Wrong Word Choice	
b.	Missing / Redundant Word	
c.	Word-level error (e.g. Tense; Preposition; Singular-Plural; Wrong Verb Form)	
d.	Clausal-level error (e.g. Wrong Sentence Pattern; Run-on Sentence; Sentence Fragment)	
e.	Awkward Expression (e.g. 'Chinglish'; Unclear / Confusing Meaning due to the Wrong Grammatical Structure used)	
f.	Tone & Style error (e.g. informal word; informal phrase)	
g.	Organization-related error (e.g. Paragraphing; Lack of Transitions; Formatting)	

Comments: _____
_____

15. Which of the following **types of error** indicated in the feedback is/are **easier to understand**? You can tick a **maximum of 2** boxes.

	Types of Error	**Preference**
a.	Wrong Word / Wrong Word Choice	
b.	Missing / Redundant Word	
c.	Word-level error (e.g. Tense; Preposition; Singular-Plural; Wrong Verb Form)	
d.	Clausal-level error (e.g. Wrong Sentence Pattern; Run-on Sentence; Sentence Fragment)	
e.	Awkward Expression (e.g. 'Chinglish'; Unclear / Confusing Meaning due to the Wrong Grammatical Structure used)	
f.	Tone & Style error (e.g. informal word; informal phrase)	
e.	Organization-related error (e.g. Paragraphing; Lack of Transitions; Formatting)	

Comments: _____
_____

333

16. Which of the following **types of error** indicated in the feedback is/are **more difficult to correct**? You can tick a **maximum of 2** boxes.

	Types of Error	Preference
a.	Wrong Word / Wrong Word Choice	
b.	Missing / Redundant Word	
c.	Word-level error (e.g. Tense; Preposition; Singular-Plural; Wrong Verb Form)	
d.	Clausal-level error (e.g. Wrong Sentence Pattern; Run-on Sentence; Sentence Fragment)	
e.	Awkward Expression (e.g. 'Chinglish'; Unclear / Confusing Meaning due to the Wrong Grammatical Structure used)	
f.	Tone & Style error (e.g. informal word; informal phrase)	
e.	Organization-related error (e.g. Paragraphing; Lack of Transitions; Formatting)	

Comments: _____
_____

17. Which of the following **types of error** indicated in the feedback is/are **easier to correct**? You can tick a **maximum of 2** boxes.

	Types of Error	Preference
a.	Wrong Word / Wrong Word Choice	
b.	Missing / Redundant Word	
c.	Word-level error (e.g. Tense; Preposition; Singular-Plural; Wrong Verb Form)	
d.	Clausal-level error (e.g. Wrong Sentence Pattern; Run-on Sentence; Sentence Fragment)	
e.	Awkward Expression (e.g. 'Chinglish'; Unclear / Confusing Meaning due to the Wrong Grammatical Structure used)	
f.	Tone & Style error (e.g. informal word; informal phrase)	
e.	Organization-related error (e.g. Paragraphing; Lack of Transitions; Formatting)	

Comments: _____
_____

18. Which of the following **post-writing activities** do you think your language teacher should ask you to do **more** often after he/she has returned your composition? You can tick a **maximum of 3** boxes.

	Error Feedback Strategies	Preference
a.	Read the grade	
b.	Read the written comments	
c.	Correct all the errors	
d.	Correct some of the errors	
e.	Rewrite the whole composition	
f.	Ask the teacher for clarifications, explanations or help in class	
g.	Consult dictionaries, grammar books or writing textbooks	
h.	Refer back to previous compositions	
i.	Work with a partner to help each other improve the composition	
j.	Work on a proofreading exercise	
k.	Read aloud some good sentences in class	
l.	Show and explain the most common errors/mistakes made by students	

m.	Hold an individual consultation with the instructor to get his/her advice	
n.	None of the above items	
o.	Others	

19. To what extent do you think the teacher written feedback (as a whole) helps you with your long-term writing? (Please circle only **ONE** answer) ('3' = Partially Helpful)

5	4	3	2	1
Very Helpful < ----------------------------- Partly Helpful --------------------------------> Not Helpful at all				

Comments: _____
_____

Appendix E: Statistical Results

Table 6.1 Summary Statistics of the Experiment group & Control Group on Pre-test & Post-test

		N	Mean (Score)	SD	SE
Experiment Group	Pre-test	32	31.12	7.22	1.44
	Post-test	32	40.05	9.33	1.70
Control Group	Pre-test	30	31.7	10.50	1.92
	Post-test	30	41.39	13.32	2.35

Table 6.2 Two Sample t-tests for Pre-test & Post-test

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Pre-test *Experimental vs. Control*	-0.234	53	0.816	-0.58	-5.556	4.396	N/A
Post-test *Experimental vs. Control*	-0.303	60	0.763	-.89	-6.769	4.988	N/A

Table 6.3 Summary Statistics of the Experiment group & Control Group on 'Total Error'

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	20.46	10.852	1.981
	Draft 2	32	2.13	2.315	0.423
Control Group	Draft 1	30	19.16	10.349	1.829
	Draft 2	30	9.5	7.509	1.327

Table 6.4 Two Sample t-tests for Total Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	0.486	59.244	0.629	1.3111	-4.085	6.706	N/A
Draft 2 *Experimental vs. Control*	-5.2879	37.197	5.697e-06	-7.367	-10.189	-4.544	-1.326

Table 6.5 Two Sample t-tests for Total Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	9.049	31.635	2.736e-10	18.333	14.205	22.462	2.34
Control Group *Draft 1 vs. Draft 2*	4.272	56.557	7.496e-05	9.65	5.129	14.183	1.07

Table 6.6 Summary Statistics of the Experiment group & Control Group on 'Awkwardness' Error

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	1.57	1.43	0.261
	Draft 2	32	0.167	0.5301	0.097
Control Group	Draft 1	30	2.25	1.984	0.351
	Draft 2	30	1.66	1.911	0.338

Table 6.7 Two Sample t-tests for 'Awkwardness' Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	-1.563	56.388	0.124	-0.683	-1.559	0.192	N/A
Draft 2 *Experimental vs. Control*	-4.238	36.048	0.000149	-1.489	-2.202	-0.777	-1.047

Table 6.8 Two Sample t-tests for 'Awkwardness' Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	5.025	36.83	1.32e-05	1.4	0.835	1.964	1.297

Control Group	1.219	61.914	0.227	0.594	-0.380	1.567	N/A
Draft 1 vs. Draft 2							

Table 6.9 Summary Statistics of the Experiment group & Control Group on 'Clausal' Error

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	2.4	2.127	0.388
	Draft 2	32	0.1	0.305	0.056
Control Group	Draft 1	30	2.34	1.877	0.332
	Draft 2	30	1.4	1.316	0.233

Table 6.10 Two Sample t-tests for 'Clausal-level' Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	0.110	57.921	0.913	0.056	-0.966	1.079	N/A
Draft 2 *Experimental vs. Control*	-5.459	34.534	4.153e-06	-1.306	-1.792	-0.820	-1.347

Table 6.11 Two Sample t-tests for 'Clausal-level' Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	5.863	30.193	2e-06	2.3	1.50	3.1	1.514
Control Group *Draft 1 vs. Draft 2*	2.31	55.55	0.0244	0.943	0.125	1.75	0.578

Table 6.12 Summary Statistics of the Experiment group & Control Group on 'Word' Error

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	11.73	8.081	1.475
	Draft 2	32	1.5	1.871	0.342
Control Group	Draft 1	30	11.22	7.490	1.324
	Draft 2	30	5	4.932	0.872

Table 6.13 Two Sample t-tests for 'Word'-level Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	0.260	58.83	0.796	0.52	-3.453	4.482	N/A
Draft 2 *Experimental vs. Control*	-3.7379	40.234	0.000577	-3.5	-5.392	-1.608	-0.927

Table 6.14 Two Sample t-tests for 'Word-level' Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	6.7572	32.1	1.217e-07	10.23	7.149	13.318	1.745
Control Group *Draft 1 vs. Draft 2*	3.922	53.624	0.000251	6.218	3.040	9.398	0.981

Table 6.15 Summary Statistics of the Experiment group & Control Group on 'Collocation' Error

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	0.73	0.868	0.716
	Draft 2	32	0	0	0
Control Group	Draft 1	30	0.44	0.158	0.127
	Draft 2	30	0.41	0.560	0.099

Table 6.16 Two Sample t-tests for 'Collocation' Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	1.459	56.325	0.150	0.296	-0.110	0.702	N/A
Draft 2 *Experimental vs. Control*	-4.1043	31	0.000271	-0.406	-0.608	-0.204	-1.00940

Table 6.17 Two Sample t-tests for 'Collocation' Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	4.6256	29	7.172e-05	0.733	0.409	1.058	1.194
Control Group *Draft 1 vs. Draft 2*	0.1946	58.609	0.8464	0.0313	-0.290	0.353	N/A

Table 6.18 Summary Statistics of the Experiment group & Control Group on 'Tone & Style' Error

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	0.7	0.915	0.167
	Draft 2	32	0.033	0.182	0.0333
Control Group	Draft 1	30	0.91	1.552	0.274
	Draft 2	30	0.31	0.780	0138

Table 6.19 Two Sample t-tests for 'Tone & Style' Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	-0.642	50.79	0.524	-0.206	-0.851	0.439	N/A
Draft 2 *Experimental vs. Control*	-1.967	34.6	0.0572	-0.279	-0.567	0.00904	-0.485

Table 6.20 Two Sample t-tests for 'Tone & Style' Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	3.912	31.304	0.000460	0.667	0.319	1.014	1.010
Control Group *Draft 1 vs. Draft 2*	1.933	45.723	0.0594	0.594	-0.0246	1.212	0.483

Table 6.21 Summary Statistics of the Experiment group & Control Group on 'Content' Error

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	0.9	1.0619	0.194
	Draft 2	32	0.1	0.403	0.0735
Control Group	Draft 1	30	1.31	1.615	0.286
	Draft 2	30	0.44	0.716	0.127

Table 6.22 Two Sample t-tests for 'Content' Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	-1.1952	53.921	0.237	-0.4125	-1.104	0.279	N/A
Draft 2 *Experimental vs. Control*	-2.307	49.44	0.0253	-0.338	-0.631	-0.0436	-0.576

Table 6.23 Two Sample t-tests for 'Content' Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	3.859	37.168	0.000439	0.8	0.380	1.22	0.996
Control Group *Draft 1 vs. Draft 2*	2.802	42.719	0.00761	0.875	0.245	1.505	0.700

Table 6.24 Summary Statistics of the Experiment group & Control Group on 'Organization' Error

		N	Mean (Error)	SD	SE
Experiment Group	Draft 1	32	2.43	2.269	0.820
	Draft 2	32	0.23	0.504	0.0920
Control Group	Draft 1	30	0.69	0.414	0.145
	Draft 2	30	0.28	0.581	0.103

Table 6.25 Two Sample t-tests for 'Organization' Errors by Drafts (Between Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Draft 1 *Experimental vs. Control*	3.977	36.038	0.000322	1.746	0.856	2.636	1.0364
Draft 2 *Experimental vs. Control*	-0.347	59.652	0.730	-0.0480	-0.324	0.228	N/A

Table 6.26 Two Sample t-tests for 'Organization' Errors by Groups (Within Groups)

	t	df	p-value	Mean difference	95% confidence interval		Effect size
					Lower	Upper	
Experimental Group *Draft 1 vs. Draft 2*	5.1832	31.854	1.181e-05	2.2	1.335	3.065	1.338
Control Group *Draft 1 vs. Draft 2*	2.286	55.844	0.0261	0.4063	0.0501	0.762	0.571

Appendix F: Participants' Prior Perceptions of Teacher Feedback on Writing

1. Was your language instructor's feedback understandable?

> Slightly below 70% of the respondents (43 participants out of 62 in total) respectively from the experimental group (67%) and the control group (68%) indicated that they were able to understand the language instructor's feedback, while 6% and 9% of the respondents respectively from the groups indicated that they did not quite understand the feedback.

2. To what extent were you able to correct the errors?

> A bit over 60% of the respondents (38 participants out of 62 in total) respectively from the experimental group (60%) and the control group (64%) indicated that they were able to correct the errors indicated by their instructor, while 7% of the respondents from each group indicated that they were not quite able to act on the feedback.

3. What types of the following feedback would you like to receive more in the future?

> More than half of the respondents respectively from the experimental group (65%) and the control group (52%) expressed that they demanded more teachers' comments on their writing, while 35% of the respondents from each group expressed that they preferred the teachers put in effort on highlighting their errors. It is noted that 13% of the respondents from the control group expressed that they would like to receive the grade instead of any error feedback or comments, while no respondents from the experimental group chose this preference.

4. Which of the following focus of error would you like our language instructor to emphasize more?

> About 60% of the respondents respectively from the experimental group (58%) and the control group (63%) reported that they preferred the teacher feedback focusing on language while about 30% of the respondents respectively from the experimental

group (34%) and the control group (30%) preferred organization. Less than 10% of the respondents respectively from the experimental group (8%) and the control group (7%) highlighted content as their preference.

5. Please indicate the amount of error you would like your language instructor to respond to:

About 70% of the respondents respectively from the experimental group and the control group reported that they would like their language instructor to indicate all errors they made.

6. To what extent do you think the teacher written feedback helps you with your long-term writing:

The majority of the respondents respectively from the experimental group (74%) and the control group (74%) found the teacher written feedback very helpful and quite helpful with their long term writing, with 15% and 12% respectively from the experimental group and control group claimed that the feedback was very helpful.

Appendix G: Summary of the Statistical Results

<table>
<tr><td colspan="4" align="center">Between-Groups Comparison</td></tr>
<tr><td colspan="4">*Whether there is a difference between the Electronic feedback and Paper-based feedback in their effectiveness in reducing errors in draft 1 (D1).

**Whether there is a difference between the Electronic feedback and Paper-based feedback in their effectiveness in reducing errors in draft 2 (D2).</td></tr>
<tr>
<td>Between Groups</td>
<td>Electronic (D1) vs. Paper (D1)

Electronic (D2) vs. Paper (D2)</td>
<td>Draft 1*

Electronic vs. Paper</td>
<td>Draft 2**

Electronic vs. Paper</td>
</tr>
<tr>
<td>Total Errors

(Ch. 6.2)</td>
<td>Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?</td>
<td>No significant diff. for both groups in draft 1

(P-value = 0.6288)</td>
<td>Significant reduction in electronic feedback in draft 2

(P-value = 5.697e-06)

(Effect size = -1.308)

large effect</td>
</tr>
<tr>
<td>'Awkwardness' Error

(Ch. 6.3)</td>
<td>Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?</td>
<td>No significant diff. for both groups in draft 1

(P-value = 0.1237)</td>
<td>Significant reduction in electronic feedback in draft 2

(P-value = 0.0001492)

(Effect size = -1.0472)

large effect</td>
</tr>
<tr>
<td>'Clausal-level' Error

(Ch. 6.4)</td>
<td>Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?</td>
<td>No significant diff. for both groups in draft 1

(P-value = 0.9127)</td>
<td>Significant reduction in electronic feedback in draft 2

(P-value = 4.153e-06)

(Effect size = -1.3471)

large effect</td>
</tr>
<tr>
<td>'Word-level' Error

(Ch. 6.5)</td>
<td>Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?</td>
<td>No significant diff. for both groups in draft 1

(P-value = 0.7961)</td>
<td>Significant reduction in electronic feedback in draft 2

(P-value = 0.000577)

(Effect size = -0.9269)

large effect</td>
</tr>
<tr>
<td>'Collocation' Error</td>
<td>Is there a statistical difference between the average number of errors respectively made by the</td>
<td>No significant diff. for both groups in draft 1</td>
<td>Significant reduction in electronic feedback in</td>
</tr>
</table>

(Ch.6.6)	electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?	(P-value = 0.1502)	draft 2 (P-value = 0.0002731) (Effect size = -1.0261) *large effect*
'Tone & Style' Error **(Ch. 6.7)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?	No significant diff. for both groups in draft 1 (P-value = 0.5238)	Significant reduction in electronic feedback in draft 2 (P-value = 0.05722) (Effect size = -0.4854) *marginal effect*
'Content' Error **(Ch. 6.8)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?	No significant diff. for both groups in draft 1 (P-value = 0.2372)	Significant reduction in electronic feedback in draft 2 (P-value = 0.02529) (Effect size = -0.5763) *medium effect*
'Organization' Error **(Ch. 6.9)**	Is there a statistical difference between the average number of errors respectively made by the electronic feedback group and the paper-based group, respectively in draft 1 and draft 2?	Significant diff. where control group was initially better in draft 1 (P-value = 0.0003222) (Effect size = 1.0364) *large effect*	No significant diff. for both groups in draft 2, which means electronic feedback made significant reduction (P-value = 0.7295)

	Within-Groups Comparison		
#Whether the Paper-based feedback makes a significant difference in reducing errors in draft 2 (D2).			
##Whether the Electronic feedback makes a significant difference in reducing errors in draft 2 (D2).			
Within Groups	**Paper (D1) vs. Paper (D2) Electronic (D2) vs. Electronic (D2)**	**Paper (D1) vs. Paper (D2)#**	**Electronic (D1) vs. Electronic (D2)##**
Total Errors (Ch. 6.2)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	Significant diff. in paper feedback in draft 1 (P-value = 7.496e-05) (Effect size = 1.07) *large effect*	More significant reduction in electronic feedback in draft 2 (P-value = 2.736e-10) (Effect size = 2.34) *very large effect*
'Awkwardness' Error (Ch. 6.3)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	Not significant diff. in paper feedback in draft 1 (P-value = 0.2273)	Significant reduction in electronic feedback in draft 2 (P-value = 1.32e-05) (Effect size = 1.297427) *large effect*
'Clausal-level' Error (Ch. 6.4)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	Significant diff. in paper feedback in draft 1 (P-value = 0.02444) (Effect size = 0.5783045) *medium effect*	More significant reduction in electronic feedback in draft 2 (P-value = 2e-06) (Effect size = 1.513741) *large effect*
'Word-level' Error (Ch. 6.5)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	Significant diff. in paper feedback in draft 1 (P-value = 0.0002511) (Effect size = 0.9806115) *large effect*	More Significant reduction in electronic feedback in draft 2 (P-value = 1.217e-07) (Effect size = 1.744696) *larger effect*
'Collocation' Error (Ch. 6.6)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1	No significant diff. in paper feedback in draft 1 (P-value = 0.8464)	Significant reduction in electronic feedback in draft 2 (P-value = 7.172e-05) (Effect size = 1.194329) *large effect*

	and draft 2?		
'Tone & Style' Error (Ch. 6.7)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	Marginally significant reduction in paper feedback in draft 1 (P-value = 0.05944) (Effect size = 0.4832547) *medium effect*	More significant reduction in electronic feedback in draft 2 (P-value = 0.0004606) (Effect size = 1.010064) *large effect*
'Content' Error (Ch. 6.8)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	Significant reduction in paper feedback in draft 1 (P-value = 0.007606) (Effect size = 0.7004521) *medium effect*	More significant reduction in electronic feedback in draft 2 (P-value = 0.0004391) (Effect size = 0.9962497) *large effect*
'Organization' Error (Ch.6.9)	Is there a statistical difference between the average number of errors respectively made by the paper-based feedback group and the electronic feedback group within their own treatment groups, respectively in draft 1 and draft 2?	Significant reduction in paper feedback in draft 1 (P-value = 0.0261) (Effect size = 0.5713651) *medium effect*	More significant reduction in electronic feedback in draft 2 (P-value = 1.181e-05) (Effect size = 1.338308) *large effect*