

**The Effects of Comprehensive Direct and Indirect Written
Corrective Feedback on Accuracy in English as a Foreign
Language Students' Writing**

by

Darby McGrath (BA, MA TESOL)

**Department of Linguistics
Macquarie University**

October 10th, 2016

Table of Contents

Abstract.....	3
Chapter 1: Introduction.....	4
1.1 Statement of the problem.....	4
1.2 The purpose of this study.....	4
1.3 The context of this study.....	6
1.4 Organisation of this thesis.....	6
Chapter 2: Literature Review.....	8
2.0 Introduction.....	8
2.1 Theoretical perspectives	8
2.1.1 Explicit and implicit knowledge as a dichotomy.....	8
2.1.2 Processability theory.....	9
2.1.3 Skill acquisition theory	10
2.1.4 Sociocultural theory.....	10
2.1.5 Interactionist perspectives.....	11
2.2 Is written corrective feedback effective?	12
2.2.1 Pre-Truscott (1996).....	12
2.2.2 Post-Truscott (1996)	14
2.3 Which forms of WCF are most likely to be effective?	18
2.4 How focused should feedback be?.....	21
2.5 A synopsis of the studies reviewed.....	23
Chapter 3: Methods.....	25
3.0 Introduction.....	25
3.1 Design of the study	25
3.2 Participants.....	26
3.3 Ethical issues.....	29
3.4 Treatments	29
3.5 Writing tasks	31
3.6 Data collection crocedures.....	32
3.6.1 Session one (Pretest).....	32
3.6.2 Sessions two and three.....	32
3.6.3 Session four (Delayed posttest)	33
3.7 Data processing procedures	33
3.8 Reliability measures.....	33
3.9 Measuring the outcome variables	34

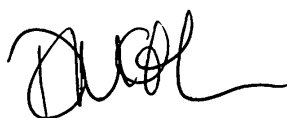
3.10 Data analysis procedures	34
Chapter 4: Findings and discussion.....	37
4.0 Introduction	37
4.1 Findings	37
4.1.1 Inferential statistics	37
4.1.2 Effect sizes	40
4.2 Implications of the findings for the research questions	43
4.2.1 On the efficacy of explicit, comprehensive, indirect WCF	44
4.2.2 On the efficacy of comprehensive, direct WCF	45
4.3 Interpretation	46
4.4 The learners' perspective	50
4.5 Concluding remarks	52
Chapter 5: Conclusions	53
5.0 Introduction	53
5.1 Summary of key findings	53
5.2 Pedagogical implications	53
5.3 Implications for further research	54
5.4 Limitations	55
5.5 Conclusion.....	55
References	57
Appendices	64

Abstract

This study investigated the effect of two forms of written corrective feedback (WCF) on written accuracy: comprehensive indirect feedback, in which students received coded feedback on all errors, and comprehensive direct feedback, in which all errors were corrected by the rater. The study looked at four pieces of writing over two months in a group of 30 students at a Japanese university. The study was quasi-experimental in that it took place outside the classroom and participants were randomly assigned to treatment and control groups, but sampling was non-random. No significant effect was found either for time or feedback type on accuracy. This is contrary to the findings of a majority of recent research, most of which has found some effect for WCF. Possible explanations for these findings are discussed. The first is type II error. The second is that the study supports the argument that WCF is ineffective. The possibility that comprehensive feedback overwhelms students and is therefore ineffective is also discussed. Participant motivation is an additional consideration. A final possibility is that the results were due to the context of the study, which meant that it was largely controlled for language input for its duration. The ramifications of these possibilities for WCF are discussed, and directions for future research are offered.

Candidate Statement

This work has not been submitted for a higher degree to any other university or institution.

A handwritten signature in black ink, appearing to read 'Darby McGrath', with a stylized, cursive script.

Darby McGrath

Chapter 1: Introduction

1.1 Statement of the problem

There is an intuitive appeal to written corrective feedback (WCF) in foreign language writing instruction. The notion that teachers should offer some form of error correction as a means of helping students to improve their written accuracy seems self-evident. In fact, prior to the publication of Truscott's (1996) paper calling for the abandonment of WCF in students' L2 writing, the use of WCF was uncontroversial. In the time since, there has been vigorous debate. At issue is the question of whether WCF does in fact have any positive effect on students' long-term written accuracy or language acquisition. While much of the recent research points to some efficacy for WCF (see, for example, Bitchener, 2008; Bitchener & Knoch, 2008, 2009, 2010a, 2010b; Bitchener, Young, & Cameron, 2005; R. Ellis, Sheen, Murakami, & Takashima, 2008; Mawlawi Diab, 2015; Van Beuningen, De Jong, & Kuiken, 2008, 2012), the question remains disputed. Truscott (1999, 2004, 2007, 2009, 2010; Truscott & Hsu, 2008) continues to hold that rather than being beneficial WCF may in fact have a detrimental effect on L2 writing because it reduces fluency and complexity.

The extent to which WCF facilitates improvement in written accuracy is important because it plays a significant role in many writing classrooms. Many teachers spend a lot of time giving WCF based on the assumption that it has long-term benefits for students' writing. Students are also asked to devote a lot of time to attending to this feedback. Furthermore, there is evidence to suggest that teachers give much more error-driven WCF than they think they do, and even that they tend to give more WCF than they profess to think is useful (Alshahrani & Storch, 2014; Montgomery & Baker, 2007). As Truscott has pointed out (1996, 1999; Truscott & Hsu, 2008), if WCF is ineffective, this represents a tremendous waste of time and resources. It is therefore in the interests of all stakeholders that we move closer to understanding the extent to which WCF facilitates learning.

1.2 The purpose of this study

One aim of this study was to contribute to our general understanding of the value of WCF as a vehicle for language acquisition. This study also attempts to fill three smaller niches in the literature on the efficacy of WCF. First, because it is not classroom-based it is one of the few studies for which language input is not a significantly confounding

variable. While it is impossible to eliminate input entirely over the course of two months, the fact that the participants in this study did not receive language instruction and the study took place in an English as a foreign language (EFL) context is unique in this area of research. One intention of this study was to investigate the efficacy of different forms of WCF without the influence of other input. Second, much of the recent research has indicated an effect for WCF on students' written accuracy. However, most of this research has been highly focused, inasmuch as it has only dealt with feedback on one language point. Although this degree of focus is interesting from a theoretical perspective, it is arguably of limited use in the classroom because it bears little relationship to what teachers actually do, or to what their students expect. While the focused approach has a place, there is also a need for research that moves away from the efficacy of highly focused WCF to comprehensive WCF¹. This shift in focus will inform teachers' decisions about whether the WCF they give their students is worthwhile. Finally, this study aimed to offer a unique perspective on the debate over two different types of WCF: direct WCF, in which the learner is given the correction, and indirect coded WCF, in which the error type is indicated for the learner, but the correction is not provided. Much of the recent literature has argued for the superiority of the direct approach (Bitchener & Knoch, 2010a, for example). However, it is not clear that research has been conducted in conditions in which we might expect indirect feedback to be effective. This study was designed to provide such conditions in order to give a better indication of the relative merits of direct and indirect WCF.

Given the purposes of this study, it sought to answer the following research questions:

1. Do students who are required to correct grammatical errors in their writing by way of comprehensive, explicit, indirect WCF make fewer grammatical errors over time?
2. Do students who are required to correct grammatical errors in their writing by way of comprehensive, direct WCF make fewer grammatical errors over time?
3. Do students who are required to correct grammatical errors in their writing by way of comprehensive, explicit, indirect WCF make fewer grammatical errors in a over time than students who receive no WCF?

¹ Comprehensive WCF is corrective feedback for all or many error types in a piece of writing. It is arguably more representative of the type of feedback that teachers are likely to give in a classroom setting.

4. Do students who are required to correct grammatical errors in their writing by way of comprehensive, direct WCF make fewer grammatical errors over time than students who receive no WCF?
5. Do students who are required to correct grammatical errors in their writing by way of comprehensive, explicit, indirect WCF make fewer grammatical errors over time than students who are required to correct grammatical errors in their writing by way of comprehensive, direct WCF?
6. Is comprehensive, direct WCF more effective in improving student accuracy in orthographic errors than in non-orthographic errors?
7. Is comprehensive, explicit, indirect WCF more effective in improving student accuracy in orthographic errors than in non-orthographic errors?

1.3 The context of this study

This study was conducted at the School of Policy Studies at a private Japanese university. Policy studies is an eclectic field, but its primary focus is policy analysis and policymaking. In this context, all students are required to complete four semesters of academic English classes. The English Language Programme (ELP) is divided into a number of skills-focused courses, including listening, reading, presentation, seminar, and writing. The writing programme is typical of such programmes in the Japanese university EFL context. There is an emphasis on structure and organisation, with classes initially focusing on paragraph construction, moving onto simple four-paragraph essays, and ending with research papers. This structural focus leaves little time for explicit language input, which is the source of some debate within the programme. However, all writing classes employ an indirect coded approach to corrective feedback, using the same symbols as were used in this study to code students' errors (see Appendix A). All participants in this study had recently completed the ELP in the School of Policy Studies, and so were familiar with the coded symbols.

1.4 Organisation of this thesis

This thesis consists of five chapters. Following this introduction is a review of the key literature related to the efficacy of WCF. This review helped to identify gaps in the literature which this study was designed to investigate. The third chapter details the

methodological approach taken in this study, including the setting and participants, the research design, and data collection and analysis. Following this, the results of the quantitative analysis are presented and discussed in Chapter 4, and the key findings are outlined and interpreted. The final chapter summarises the results of the study, discusses its limitations, and outlines its implications for language pedagogy, as well as possible directions for future research.

Chapter 2: Literature Review

2.0 Introduction

This chapter offers an overview of the literature associated with WCF. First it presents a discussion of different theoretical perspectives that underpin the disagreement over the efficacy of WCF. The contention is that at least some of the disagreement can be attributed to competing theoretical perspectives on second language acquisition (SLA). Next it reviews the WCF research base and what it contributes in to our knowledge of the efficacy of WCF. This is approached chronologically to show development of the topic over time, but with a stronger focus on the research carried out over the last twenty years. Finally, it identifies the issues raised in the literature that formed the specific focuses of this study: what we know, what the gaps are in our understanding with respect to what kind of WCF works best, and how much WCF learners should be given.

2.1 Theoretical perspectives

This section will examine five theoretical perspectives on second language acquisition that are most relevant to the debate over the efficacy of WCF. The first perspective is that of theories that discount a role for WCF in SLA, such as those of Schwartz (1986) and Krashen (1982). The second is the processability theory of Pienemann (2003; Pienemann & K  bler, 2011), which suggests that until we know more about the sequences in which learners acquire grammar, we are unable to provide effective written feedback. The third perspective, DeKeyser's (2007) skill acquisition theory, admits a role for WCF in SLA inasmuch as it assumes a direct pathway between explicit and implicit linguistic knowledge. The role of WCF for the fourth theoretical perspective, sociocultural theory, is less clear. In principal sociocultural theory has a place for WCF. However, sociocultural theorists such as Lantolf (2012) argue that to be effective feedback must cater to the needs of individual learners, which presents a substantial challenge in practice. Finally, this section will look at the role of WCF from interactionist perspectives. Schmidt's (2001) concepts of attention and noticing are particularly relevant in this context.

2.1.1 Explicit and implicit knowledge as a dichotomy

When we ask what effect WCF has on language acquisition, we are effectively looking at the relationship between explicit and implicit grammatical knowledge. In simple terms, WCF can be seen as a means by which explicit knowledge is imparted to the learner.

Long-term improvements in accuracy can, in a controlled environment, be understood as evidence that this explicit knowledge has led in some way to implicit knowledge, or acquisition. With this in mind, the debate over the efficacy of WCF can be traced to a degree to differing theoretical perspectives. In other words, opposition to or support for the use of WCF as a teaching and learning tool is to some extent informed by whatever theoretical position one adopts with respect to the interplay between explicit and implicit knowledge. For example, Truscott's (1996, 1999, 2004, 2007; Truscott & Hsu, 2008) opposition to the use of WCF appears to stem in part from a perspective that rejects the notion that explicit and implicit knowledge might lie on a spectrum. In this sense, when Truscott (1996, 1999) argues that WCF not only does not work, but also cannot work, he does so from the point of view of theorists like Schwartz and Krashen, who view implicit and explicit knowledge as effectively dichotomous. Truscott (1996) cites both Krashen and Schwartz in support of his argument, neither of whom accepts a significant role for negative feedback in L2 acquisition (Bitchener, 2012; R. Ellis, 2010; Polio, 2012). Schwartz and Krashen each draw a distinction between their own analogues of explicit and implicit knowledge. To Schwartz (1986) there is no relationship between linguistic knowledge (explicit) and competence (implicit). According to Krashen's (1982) acquisition-learning hypothesis, the same is true of learning (a means of acquiring explicit knowledge) and acquisition (the means by which knowledge becomes implicit). Truscott's (1996, p. 345) insistence that WCF can effect nothing better than "pseudolearning", seems to be a claim that explicit knowledge is the best we can expect from WCF (Polio, 2012). This carries the assumption that declarative explicit knowledge has no effect on implicit knowledge.

2.1.2 Processability theory

Truscott's (1996) second theoretical claim is based on processability theory (see, for example, Pienemann, 2003; Pienemann & K  bler, 2011). The central tenet of processability theory is the concept of developmental sequences in grammar acquisition, and, as Truscott (1996, p. 344) puts it, the notion that "When students are corrected on a point for which they are not yet ready, the correction is unlikely to have much value". From this perspective, the problem with WCF is not so much conceptual as a question of timing. Truscott holds that we do not know enough about developmental sequences to be able to give developmentally appropriate WCF. There are questions, however, over how much processability theory has to say about written language. Polio (2012) points out

that it is a theory of oral production. In writing, learned rules, such as those which WCF can provide, may overcome the processing constraints imposed by working memory. What this means is that although developmental sequences seem uncontroversial for oral language, not enough is known about L2 writing to say that it occurs under similar constraints.

2.1.3 Skill acquisition theory

Other theoretical frameworks are more accommodating of WCF. Anderson's (1982) Adaptive Control of Thought Model suggests that cognitive skills are acquired in two stages: declarative and procedural. DeKeyser (2007) applies this notion directly to L2 acquisition. He views L2 learning as a difficult skill analogous to any other difficult skill. He posits a connection between declarative (explicit) knowledge and procedural (implicit) knowledge, a connection that practice can strengthen. What this means for WCF is that its role in the development of explicit knowledge, and also the sense in which it can inform practice, could lead to procedural knowledge. DeKeyser himself feels that we are some way short of fully understanding the effects of feedback, including “explicit error correction”, but does point out that it “appears to have a substantial positive effect” (2007, p. 9). From within his skill-acquisition framework, we might expect it to do so.

2.1.4 Sociocultural theory

The sociocultural perspective represents an attempt to integrate the psychological and social elements of learning into a unified theory. For sociocultural theorists, language is viewed as a tool which mediates users’ mental processes and action (Lantolf & Thorne, 2007). Learning takes place when learners engage in dialogue which necessitates the use of this tool (Lantolf & Thorne, 2006). With respect to second language acquisition this dialogue will tend to be interaction and collaboration with instructors or other more proficient interlocutors (Lantolf, 2012). The interlocutor can offer support for any language production that takes place within what Vygotsky (1978) termed the learner’s “zone of proximal development” (ZPD). For Vygotsky (1978, p. 86), the ZPD represented the gap between what a learner (for Vygotsky a child) is capable of independently, and what he or she can do “under...guidance, or in collaboration with more capable peers”. In other words, the ZPD sits at the edge of what the learner can do without support the interlocutor’s support, and, according to sociocultural theory, is

where language learning occurs. The goal of both the learner and the interlocutor will be to reach a stage at which the learner can produce language without the interlocutor's support (Lantolf & Thorne, 2006). The ZPD will differ from learner to learner, depending on their level of development, and so feedback will need to be adjusted accordingly. This emphasis on individual differences can limit the practical value of sociocultural theory in some contexts. There is little question that individual differences play a substantial role in the uptake of WCF (Ferris, 2004, 2006; Ferris, Liu, Sinha, & Senna, 2013; Hyland, 1998, 2003; Kormos, 2012; Storch & Wigglesworth, 2010). However, the requirement that the instructor tailor WCF to the ZPD of each learner may restrict its usefulness to relatively small groups of learners.

This practical limitation is instantiated in a study by Aljaafreh and Lantolf (1994), which adopted a sociocultural perspective to the provision of corrective feedback. In this study, learners and the researchers held 35-40 minute meetings in which learners were invited to self-correct their errors by way of negotiation with the expert interlocutor. The support the learners were given was initially very general, but gradually became more specific until the learner was able to self-correct. This represented the support that allowed learners to correct themselves with guidance, and which for sociocultural theory is the precursor to independent use. Such a time-consuming approach will be beyond most language instructors in most contexts. However, it does suggest possible compromises that may be appropriate for more conventional classroom use. One possibility is a system whereby WCF is provided over a series of drafts, each of which the student is asked to correct. Feedback on each draft would gradually become more specific, moving implicit WCF, to explicit indirect WCF, to direct WCF, focusing each time only on the error the learner had failed to correct previously. The need for direct WCF would suggest a language point beyond the learner's ZPD, and therefore something for which the learner is not ready.

2.1.5 Interactionist perspectives

The interactionist perspective is, according to Bitchener (2012), the language learning theory that offers the biggest role for WCF in learning. Interactionism looks at the extent to which language acquisition is influenced by the relationships between the learner and those with whom he or she interacts (Richards & Schmidt, 2007). Long (1996), in his Interaction Hypothesis, stresses the value of both positive and negative input. The Interaction Hypothesis shares with Krashen (1982) the view that comprehensible input is

a necessary condition for language learning. Unlike Krashen though, Long accepts that negative feedback can have a role in this process. However, Long (1996, p. 453) characterises *written* feedback as “static”, and therefore less amenable to the negotiation for meaning between learner and interlocutor that he believes facilitates learning. This said, some forms of written feedback are at least analogous to the “semantically contingent” (Long, 1996, p. 452) utterances that comprise this negotiation. Rephrasings, reformulations and recasts all have direct analogues in written feedback, and as Polio (2012) points out, the interactionist perspective has provided the theoretical framework for a number of studies of WCF, including Qi and Lapkin (2001) and Sachs and Polio (2007).

Schmidt shares Long’s belief that input has a role in directing learners’ attention. Schmidt (2001, p. 11) contends that through feedback learners can be encouraged to notice the gaps in their interlanguage, and what occurs “within attentional space largely determines the course of language development”. This noticing focuses the learner on form, either explicitly or implicitly, which should then inform future output. WCF is therefore an invaluable attention-directing device. It may even be that WCF is more beneficial than oral corrective feedback in that it affords learners more time to engage with and respond to it (Bitchener, 2012).

WCF does not have a place in all theories of second language acquisition. At the same time there are clearly theoretical frameworks into which it can fit quite comfortably. The inherent plausibility of WCF as a means of acquiring language will depend in the first instance on which theoretical perspectives we adopt.

2.2 Is written corrective feedback effective?

Research into WCF can be separated into two periods. Prior to the publication of Truscott’s (1996) article in which he argued for the abandonment of WCF, and the time since the publication of Truscott’s article. Before 1996 there were a small number of interesting but flawed studies. Since 1996 there have been a larger number of more robust studies, although many have had problems of their own.

2.2.1 Pre-Truscott (1996)

Truscott (1996, p. 341) discounted the early positive studies of WCF, stating that: “None of the studies that purportedly support the practice of grammar correction actually do so”. To the extent that a number of the early positive studies focused only on accuracy in

revisions, not in new pieces of writing, he had a point. Truscott's argument was that these revision studies did not measure the effect of WCF on learning. Haswell (1983) and Fathman and Whalley (1990), for example, found an effect for WCF on accuracy in text revisions by US university students. However, Truscott accepts that WCF produces increased accuracy in revisions. He simply argues that this is not particularly interesting or useful (1996, 1999; Truscott & Hsu, 2008). That WCF has a positive effect on accuracy in revisions has been consistently borne out in subsequent studies (Ashwell, 2000; Ferris, 1997, 2006; Ferris & Roberts, 2001; Polio, Fleck, & Leder, 1998; Sachs & Polio, 2007).

For Truscott, the only real value in WCF would lie in the extent to which it facilitates learning, manifested in improved accuracy in new pieces of writing. Early studies that looked at WCF from this perspective were generally negative (see A. D. Cohen & Robbins, 1976; Kepner, 1991; Robb, Ross, & Shortreed, 1986; Semke, 1984; Sheppard, 1992). However, these studies have methodological limitations which undermine their conclusions. First, a significant problem for all of these early studies was a lack of control groups. Researchers often attempted to mitigate this issue by including a group that received only content or holistic feedback. However, without an actual non-treatment group it is impossible to know what effect these "control" treatments may have had. In the case of Sheppard (1992) for example, the "holistic" treatment was very similar to what later studies call implicit indirect feedback (in which the error is underlined and left for the student to correct). Furthermore, without control groups it is difficult to be sure of exactly what was being measured. For example, Robb et al. (1986) found a positive effect for four types of WCF on written accuracy. However, because this effect was of a consistent size for all four treatment groups, they concluded that this was an artefact of writing practice. It is, however, difficult to see how this conclusion can be reached without a control group to form a baseline. It is at least possible, if unlikely, that that all of the treatments were effective, but equally so.

Beyond the lack of control groups, there are other factors that compromise the early negative studies. A.D. Cohen and Robbins (1976) looked at only three learners, which makes it very difficult to generalise based on the results. Semke's (1984) two treatments were very inconsistent, including different grading criteria and classroom activities (Guénette, 2007), and a post-test accuracy measure that consisted of ten minutes of free writing, with no topic or direction provided (Bitchener, 2008). Sheppard (1992) also suffered from a lack of consistency between treatment groups. Kepner (1991)

is an interesting case in that both sides of the debate have claimed her study as offering support for their perspective. Truscott (1996) points to the fact that there was no statistically significant effect for error correction over “message-related comments” (Kepner, 1991, p. 305), and that therefore this study is evidence against the efficacy of WCF. Ferris (2004, p. 51), on the other hand, says that Kepner “finds positive evidence for error correction but curiously interprets it as negative”. Her position appears to be that the small effect size would have been bigger had it not been for the study’s methodological flaws. While it’s impossible to say that this would necessarily have been the case, it is clear that Kepner’s (1991) study, along with the other early negative studies (A. D. Cohen & Robbins, 1976; Robb et al., 1986; Semke, 1984; Sheppard, 1992) do not constitute strong evidence against the efficacy of WCF.

One pre-1996 study found an effect for WCF on new writing, not only revisions. Lalande (1982) looked at 60 students of German to compare the effects of indirect and direct feedback on students’ writing, and found an edge for indirect over direct WCF. However the difference was not significant and this study had no control group. There was also an asymmetry between treatment groups, including activities that favoured the indirect WCF group (Van Beuningen et al., 2012).

Truscott (1996) was therefore justified in suggesting that there was, at the time, little evidence to support the use of WCF for language learning. When he published his call to abandon WCF, the only solid evidence we had was in support of WCF for revisions. However, there was also a lack of negative evidence. The fact is that the research had, until then, not been of a sufficiently high standard to inform an approach to WCF either way. It was for this reason that Ferris (1999, p. 4) argued that Truscott’s position was “premature and overly strong”. She suggested that what was needed was more research of a higher standard.

2.2.2 Post-Truscott (1996)

Since 1996, the evidence for the efficacy of WCF has been more positive. However, it remains inconclusive. Although the research conducted over the past twenty years has been more rigorous, a proportion of studies continue to be compromised by shortcomings. Guénette (2007) argued that eleven years after Truscott’s (1996) article, four issues in particular were holding the research back. Firstly, Guénette contended that all studies should account for language proficiency, but had often not done so. Language proficiency is likely to have a bearing on a student’s ability or readiness to uptake WCF,

not least because more proficient learners are more likely to understand the feedback they are given (McGrath, 2015). She also called for control groups in all studies. Control groups are ethically troublesome in classroom research, but Guénette insisted that without them we can have no reliable measure of the extent to which WCF works. She also argued that treatment groups should experience the same classroom input. This is a difficult proposition, and is a problem that this current study was designed to avoid. Finally, she called for more longitudinal research. In her critical review of the WCF research base, Storch (2010) made the point that it is not enough that the research be longitudinal. She also argued that it should be based on more than one treatment, and that participants should be given sufficient opportunity to engage with the feedback they are given. Studies that lack such features “lack theoretical and pedagogical validity” (2010, p. 42). Guénette and Storch's articles will inform this review of post-1996 WCF research.

Chandler (2003) studied the effects of WCF on the written accuracy of 31 ESL students. The study found an effect for WCF on accuracy in both revisions and in new texts. However, the paper has received criticism because, as Guénette (2007) and Truscott (2004) have pointed out, it did not have a genuine control group. Bitchener, Young, and Cameron (2005) also found that WCF improved accuracy in revisions and new writing. However, there were substantial differences in the amount of instruction received by the two treatment groups and the control group in this study. Where the control group received only four hours of instruction per week, the two treatment groups each had ten or twenty hours of classes each week. As a result, the treatment groups received as many as 192 hours more instruction than the control group. This weakens claims the study has to a genuine control group. In addition, the participants chose which of the three classes to attend. We cannot be sure that less committed students did not chose the less intensive control-group class (McGrath, 2015).

Ferris (2006) also found significant effects for WCF on new texts in her study of 92 ESL students. However, the study found that WCF was effective for only certain types of grammatical errors. For others, such as the use of articles, accuracy actually declined. She also concedes that the raters did not follow the established marking scheme, and each diverged in his or her own way. This is a significant limitation which contravenes Guénette's (2007) and Storch's (2010) call for consistency between treatment groups.

Truscott and Hsu (2008) conducted a rare study published in the last twenty years to find WCF ineffective for learning. There are however reasons to suspect this may be

due the design of the study. Truscott and Hsu gave only implicit WCF using underlined errors, from which students were required to infer both the type of error and a suitable correction. This is an approach that asks perhaps too much of students to be effective, and for this reason it is almost entirely unused in recent research. All we can take from this study therefore, is that offers support for the notion that implicit WCF is ineffective, but it has nothing to say about the more explicit modes of feedback (McGrath, 2015).

Bitchener (2008) conducted one of the more rigorous studies in recent years. It satisfied all of Guénette's (2007) requirements to some degree². However the study performed less well with respect to Storch's (2010) criteria. Although it took place over two months, the study only included a single treatment, something which Storch argues limits its ecological validity. The study looked at 75 students at two language schools in New Zealand. It found an effect for WCF over the control group which was still present in the delayed posttest two months later.

Bitchener's (2008) study is typical of a recent preference for focused WCF research that investigates the effects of WCF on a single language point. Sheen (2007), R. Ellis, Sheen, Murakami, and Takashima (2008), Bitchener and Knoch (2009, 2010a, 2010b), Farrokhi and Sattarpour (2012), and Shintani and Ellis (2013) have all published studies with the same focus. All of these studies were limited to the effects of WCF on article use. This does not compromise the findings of the studies, but there is doubt as to the usefulness of such a limited focus, a point which will be discussed further in section 2.4. While these studies kept to a very narrow application of WCF, a number of them (Bitchener, 2008; Bitchener & Knoch, 2010a, 2010b; Shintani & Ellis, 2013) also used a very broad conception of WCF. For Bitchener (2008) and Bitchener and Knoch (2010b), for example, "written corrective feedback" included oral feedback in the form of mini-lessons that dealt specifically with the students' errors. Again, this is not cause to question the findings, but it is something that many teachers will not have time to do, and in the case of mini-lessons is not WCF (McGrath, 2015).

A more recent study had a similar focus on metalinguistic feedback. Mawlawi Diab (2015) compared the effects of direct WCF plus metalinguistic feedback with the effects of metalinguistic feedback only on pronoun agreement errors and lexical errors.

² Xu (2009) argued that the study did not control adequately for proficiency, but Bitchener (2009) responded that any variation in proficiency was not statistically significant.

In her immediate post-test she found an advantage for the direct WCF plus metalinguistic group, but by the delayed post-test this had disappeared.

Two recent studies by Van Beuningen et al. (2008; 2012) have diverged from this trend towards focused research. They looked at the effects of WCF on written accuracy for high school students of Dutch. These studies differed from other recent studies in that they looked at comprehensive feedback rather than the focused approach. The results of these studies, which showed an effect for WCF on accuracy, suggest that the comprehensive approach may have been abandoned prematurely. This finding was a factor in the decision to focus on comprehensive WCF in the current study.

At this point the research base points strongly to some positive effect for WCF on accuracy. The current study was intended to augment our understanding of this apparent effect in a number of ways. First, it appears to be the first WCF study conducted in an input-controlled EFL setting. What this means is that the study was effectively controlled for substantive L2 input outside WCF. The participants in this setting all lived in Japan and none studied in English-medium classes while the data were collected. Although this approach is contrary to Storch's (2010) call for more ecologically valid classroom-based research, there were good theoretical and practical reasons for adopting it. While all of the studies discussed above were confounded to some extent by their ESL setting, their classroom setting, or both, the degree of control over formal language input in this study is unique. At the very least it can help us begin to understand whether any apparent effect from WCF is due to WCF alone. The study was also somewhat longitudinal, taking place over the course of two months, and was controlled for the proficiency of the participants. Because of its duration, the study had time for two feedback and revision cycles in addition to its pretest and posttest. While this is perhaps fewer treatments than is ideal, it meets Storch's (2010) call to avoid WCF studies based on a single treatment.

This study therefore largely satisfies Guénette's (2007) and Storch's (2010) requirements. Where it diverges from Storch's criteria, it is with the aim of occupying a unique space in the literature. In this sense it is well positioned to add to our knowledge of the extent to which WCF affects learning. However, the debate over WCF is not simply a matter of whether it is broadly effective. There are a number of different types of WCF, each of which may facilitate learning where others do not. This study also investigated some of these variations, which are discussed below.

2.3 Which forms of WCF are most likely to be effective?

Two considerations inform this question. The first concerns the degree of explicitness in feedback. One approach is implicit WCF, in which students are directed to their errors, but not to what kinds of errors they are. The other is explicit feedback, in which students are made aware of the nature of their error. This distinction has not been well researched, perhaps because the implicit approach is intuitively less appealing. However, it is probably revealing that of the many notable WCF studies published in the last twenty years, the only study that focused solely on implicit WCF was also the only one that found WCF to be entirely ineffective for written accuracy (Truscott & Hsu, 2008).

While Ferris and Roberts (2001) and Ferris (2006) compared the effects of implicit and explicit WCF on revision-writing, direct comparison of their effect on learning is rare. One study that did make this comparison was Chandler (2003). Her results suggested that while both approaches were effective, implicit feedback had more substantial long-term effects. Chandler aside, there is a lack of data to inform research design with respect to this matter. Truscott and Hsu (2008) suggest that implicit feedback is ineffective, while Chandler's (2003) study suggests that it is preferable to explicit feedback. The decision in the current study to focus on explicit feedback was motivated by two factors. First, it is based on a belief that the design of most research until now has failed to create an environment where explicit, indirect WCF might be expected to be effective. This has potentially created a mistaken impression that such feedback is ineffective, something to which we will return later in this section. Secondly, explicit, indirect WCF is the approach taken by instructors in the context in which this study was carried out. It therefore bolsters the local generalisability of the study³.

The second distinction to be made here is between two forms of explicit WCF: direct feedback and indirect feedback. With the former, the teacher corrects the error for the student, while with the latter the error types are identified by the teacher, but the students are required to correct them. Before 1996, three studies (Lalande, 1982; Robb et al., 1986; Semke, 1984) compared these two forms of feedback, and none found a significant difference. Since then, results have varied. Ferris and Hely (2001, as cited in Bitchener et al., 2005 and Bitchener & Ferris, 2012) found direct WCF was more effective in improving accuracy in revisions, while indirect WCF was more effective for

³ 'Local generalisability' refers to the extent to which the results might be generalisable only to classes within the context in which the study took place, as distinct from external validity.

learning. In contrast, Bitchener & Knoch (2010a) found that direct WCF was more effective for learning. Chandler (2003) found that although both types of WCF were effective, the effects of direct WCF were greater. Van Beuningen et al. (2008) found that both direct and indirect WCF were effective, but the effects of direct WCF lasted longer. This is somewhat at odds with their more recent study (2012). In what R. Ellis and Shintani (2014) called “one of the best studies carried out to date” (p. 268), Van Beuningen et al found that direct WCF was better for improving grammatical accuracy, where indirect WCF was more effective at treating non-grammatical errors.

Clearly this is a confused area, which makes it somewhat surprising that Bitchener and Knoch (2010a) stated outright that direct WCF has “a superior longitudinal effect” (p. 215). The research as a whole does not appear to support this position, but it has clearly taken hold. Since 2005 a significant majority of the research has focused on direct WCF only (see Bitchener, 2008; Bitchener & Knoch, 2008, 2009, 2010b; Bitchener et al., 2005; R. Ellis et al., 2008; Farrokhi & Sattarpour, 2012; Sheen, 2007). There are reasons other than the lack of strong evidence to suggest that this may be premature. Since 2001 there has been a second strand of research into WCF which has focused on learners’ interactions with WCF through such instruments as think-aloud protocols (Qi & Lapkin, 2001), stimulated recall (Adams, 2003), and recorded discussions between participants (Storch & Wigglesworth, 2010). These studies have provided some interesting insights into the ways in which learners engage with WCF. What they have highlighted is the importance of noticing in the efficacy of WCF. In one way they mean this in the sense of Schmidt’s noticing hypothesis, whereby the noticing that is a necessary component of learning is “at a very low level of abstraction” that is distinct from metalinguistic awareness (Schmidt, 2001, p. 5). However, they have also found that something beyond this is necessary for WCF to be effective. Qi and Lapkin (2001) found in their case studies of two ESL students that it was not just noticing, but quality of noticing that determined uptake. To them, quality of noticing is determined by the extent to which it is “substantive” rather than “perfunctory” (p. 291), which implies a greater degree of engagement than Schmidt’s noticing hypothesis.

Similarly, Adams (2003), in a study of 56 ESL students, found that the accuracy of students who noticed and acted upon reformulations improved significantly in a written post-test over those who received direct correction. She attributes this to the passivity with which students engage with direct feedback. While Adams looked at reformulations rather than the coded WCF used in this study, implicit in her conclusion is

that we should expect a form of feedback that asks more engagement from students, such as indirect coded feedback, to be more effective. Further support for this comes from Storch and Wigglesworth (2010), who investigated the extent to which direct and indirect WCF were noticed by and engaged with by 40 ESL students. They found that not only do learners engage more with indirect feedback, but that “retention seemed to relate to the level of engagement” (p. 327).

What these qualitative studies have suggested is that we should expect indirect WCF to be more effective for learning than direct WCF. This poses the question of why the results have been so inconclusive in the quantitative classroom-based studies. One possibility is that the studies which have reported a greater effect for direct WCF have not provided a setting which allows indirect WCF to be effective. It seems likely that for WCF to be effective, learners require time to engage with it (Polio et al., 1998; Qi & Lapkin, 2001). To be beneficial, indirect coded WCF will need to be provided to students in a setting in which they are given time to consider their errors, consider and look into how to repair those errors, and make the necessary changes to their writing. This will engage learners in the “reflective learning processes” which is the intuitive appeal of indirect WCF (Ferris, 2006, p. 83).

Returning to the studies that have directly compared indirect and direct WCF, this notion receives some support. Bitchener and Knoch (2010a) found that direct WCF was more beneficial to learning than indirect. However, it is interesting that the participants in this study were not required to revise their writing; they were “given a few minutes to consider the feedback” (Bitchener & Knoch, 2010a, p. 213). If we accept that indirect feedback will require time and revision to effect learning, then we would expect indirect WCF to fare poorly in this study, as it did. Moreover, if this approach is compared to studies in which learners were required to revise their writing based on indirect feedback, we get different results. Ferris and Hely (2001 as cited in Bitchener et al., 2005, and Bitchener & Ferris, 2012) found that indirect WCF was more beneficial to learning, while Van Beuningen et al. (2012) found that indirect WCF was more beneficial for lexical errors. Both of these studies required the participants to make revisions based on the feedback they were given.

What this suggests is that recent research has prematurely discounted the possibility that indirect WCF can be more effective for learning. There are good intuitive and empirical reasons to believe that, given the right conditions, indirect WCF might be preferred. This study aimed to create those conditions to test this possibility.

2.4 How focused should feedback be?

The distinction here is between focused WCF (in which one or very few error types are treated) and comprehensive WCF (in which many or all error types are treated). The tendency in recent WCF research has been to use highly focused treatments. Since 2007, at least six studies have focused on the treatment of articles in student writing (Bitchener & Knoch, 2009, 2010a; R. Ellis et al., 2008; Farrokhi & Sattarpour, 2012; Sheen, 2007; Shintani & Ellis, 2013). One rationale for this is that the more mature field of oral feedback has shown that feedback on a single grammar point is more beneficial (Sheen, 2007). Another reason is a perception that comprehensive WCF is likely to overwhelm students, particularly those of lower proficiency (Bitchener, 2012; R. Ellis et al., 2008).

However, as with the growing focus on direct WCF over indirect, this may be premature. First, it is worth noting that the belief that comprehensive feedback is overwhelming for students is based largely on intuition. While it seems likely that comprehensive feedback would have this effect, for writing the research is inconclusive. At least four studies have directly compared the two approaches. R. Ellis et al. (2008) compared the effects of focused and comprehensive WCF on 49 EFL learners' use of articles, and found that both treatments outperformed controls equally in a posttest. Sheen, Wright and Moldowa (2009), in contrast, found that both the focused treatment group and a control group outperformed the comprehensive WCF group in post-tests. Farrokhi and Sattarpour (2012) looked at the effects of focused and comprehensive WCF on the use of articles by 60 advanced EFL students and found that while both groups outperformed the control group, the focused feedback group did better than the comprehensive. Finally, Frear and Chiu (2015) compared the effects of focused and comprehensive WCF on the use of "weak verbs" and on general accuracy in a group of 42 Taiwanese university students. They found that both the focused and the comprehensive treatment groups outperformed the control group equally. As with direct and indirect feedback, these studies suggest that the research comparing focused and comprehensive WCF cannot inform a firm decision as to which to prefer.

There is also a question of what amounts to efficacy when comparing focused and comprehensive WCF. The issue is whether we should be concerned with effects on a single grammar point, or whether more modest gains in a range of language points might not be preferable. For example, both R. Ellis et al. (2008) and Farrokhi and Sattarpour (2012) compared the relative effects of focused and comprehensive WCF on article use.

What they did not do was look also at the effects of these approaches on general accuracy, which from a learner and teacher perspective is more interesting. It is possible that the comprehensive WCF had a better effect on overall accuracy, but this was not measured.

Sheen et al. (2009), on the other hand, did compare effects beyond their focus. In addition to looking at articles, they compared the effects of focused WCF, comprehensive WCF, and writing practice on a number of grammatical categories. Surprisingly, the focused WCF group outperformed the comprehensive group in all categories, even those for which they received no feedback at all. It is also surprising that the focused group outperformed the writing practice group, who also received no feedback on those categories. In other words, the focused WCF group outperformed a group which had identical treatment for these language categories. While focused WCF might be helping learners across a range of grammar categories by some mechanism that has yet to be identified, this result appears so unlikely that it is tempting to wonder if there may be some other cause of this in the study's design. Frear and Chiu (2015) had similarly odd results inasmuch as both the comprehensive and the focused WCF groups outperformed controls in all error categories. As with Sheen et al. (2009), this means that the focused WCF group outperformed the control group even for error categories for which they received no WCF. One possible reason for this is that the control group was not asked to take part in the revision sessions, and that the improvement in the accuracy of the two treatment groups was due more to practice than the feedback they received. Whatever the reason, the results of both Sheen et al. (2009) and Frear and Chiu (2015) suggest that the apparent superiority of focused WCF may be due more to accidents of research design than to effectiveness.

Another reason to question the concentration on focused WCF is similar to an issue already raised with regard to indirect feedback. As with indirect feedback, we would expect comprehensive WCF to be more time and labour intensive for learners to attend to than focused WCF, given that there will be more error types and tokens in any given piece of writing. However, we find that this is often not accounted for in studies that found the focused approach to be more effective. Sheen et al. (2009, p. 562), for example, had students “look over their errors...for a few minutes”. It is questionable whether this will be sufficient for learners to assimilate comprehensive feedback. Van Beuningen et al. (2012), on the other hand, had students revise their papers based on the

feedback they received, and gave them the time to do so. It is no surprise that the authors found comprehensive WCF effective for learning where Sheen et al. (2009) did not.

Finally, focusing on comprehensive feedback increases the ecological validity of the research. Although the recent studies concentrating on focused WCF seem to show an effect for WCF on language learning, what they represent is so far removed from the classroom reality that they are of little practical use for teachers and students. For many teachers the idea that WCF is only effective if it focuses on one language point would be functionally equivalent to it being ineffective. They do not have the luxury of being so selective. The suggestion of R. Ellis et al. (2008) that teachers treat a different language point in each paper they set is no real solution in academic contexts in which students only write one or two papers a semester. Storch's (2010) suggestion that a semi-focused approach be taken in WCF feedback is more useful, and certainly has an intuitive appeal, but the research is not yet so conclusive that the comprehensive approach should be entirely abandoned. Ultimately the decision was taken to use comprehensive WCF for the current study because it is the approach taken to WCF in the English programme the participants had recently completed.

Truscott perhaps exaggerates when he argues that focused studies are an admission that “making sacrifices in validity and focusing on the simplest of targets are the keys to obtaining positive results” (p. 333), but the ecological validity of such an approach is a legitimate concern. Van Beuningen et al. (2012, p. 6) are more measured, saying that “most important to us is the fact that the comprehensive approach most closely resembles the correction method used in actual teaching practice”. This is the primary sentiment behind the decision to focus on comprehensive WCF in this study.

2.5 A synopsis of the studies reviewed

This chapter presented a review of the four elements of the WCF literature that have informed the approach taken in the current study. First, it discussed a number of theoretical perspectives on SLA and the extent to which they can accommodate WCF. There is room for effective WCF in these perspectives, but the degree to which we find WCF for language learning plausible will depend on what we believe about the processes by which second languages are acquired. Next, it examined what the research literature suggests about whether or not WCF can facilitate learning. While the degree of effectiveness is unclear, there is considerable evidence to suggest that some WCF is more effective than none. A unique niche was also identified for this study: an input-controlled

EFL context. Next, the different approaches to WCF were examined. It was argued that the growing consensus that direct feedback is more effective than indirect coded feedback is premature. For this reason the decision was taken to focus on the relative effectiveness of indirect and direct WCF in this study. Finally, this chapter looked at the research into focused and comprehensive WCF. It was argued that the tendency in recent research to concentrate almost exclusively on focused WCF is limiting, and that if it is to be of use to teachers and language learners, more investigation of the comprehensive approach is warranted. For this reason, this study looked at the effects of comprehensive WCF.

WCF is a heavily researched area, but due to its variety and complexity it remains poorly understood. The literature reviewed in this chapter, and more specifically the gaps identified in the current literature, informed the choice of research research questions presented in section 1.2.

Chapter 3: Methods

3.0 Introduction

This chapter explains the approach taken to measuring the efficacy of WCF in this study. It begins by detailing and justifying the approach before describing the study's participants. Next, the design is discussed, including the treatments, the procedures employed in the study, and the approach to data collection. The chapter concludes with a description of the data processing and analysis procedures.

3.1 Design of the study

It is difficult to categorise this study in terms of the experimental or quasi-experimental research dichotomy. Experimental design is typically defined by two criteria: a controlled environment and random assignment of participants to treatment and control groups (Jupp, 2006; Zedeck, 2014). This study satisfied both of these criteria and to this extent was true experimental research. However, if we accept that random assignment is a necessary criterion for experimental research, there is also an implication that a random selection of participants is necessary. If individual groups can be unrepresentative of a sample without random assignment, it follows that the sample itself can become unrepresentative of the population from which it comes without random selection. In the current study the sample was self-selected in that it consisted of paid volunteers. It was not therefore randomly selected. To the extent that this study did not satisfy all of the criteria for true experimental design it is categorised as quasi-experimental design in this thesis, although in truth it lies somewhere between the two approaches.

The quantitative approach was considered most appropriate for answering the research questions of the study, as written errors are a readily quantifiable element. This is in keeping with a majority of the prior research investigating the efficacy of WCF, something which allows the results of this study to be viewed within the context of the research base. The small number of qualitative studies have tended to focus more on teacher and student attitudes to and beliefs about WCF than on its effect on writing (see, for example, Hyland, 1998, 2003; Lee, 2005, 2009). Some nominally qualitative data were collected in the form of an entry and an exit questionnaire (see Appendices B and C), but these were only used to gain insight into the nature of the sample.

Table 3.1 shows how the study was organised. It was semi-longitudinal, taking place over two months. The aim was to measure changes in three categories of error in written English over time: total errors, orthographic errors, and non-orthographic errors.

In sessions two and three, the treatment was based on asking participants to revise their previous essay based on the feedback they had been given.

Table 3.1: *Task Type by Session*

Session	Treatment	Essay writing task
1 (pre-test)		✓
2 (practice essay)	✓	✓
3 (post test)	✓	✓
4 (delayed post test)		✓

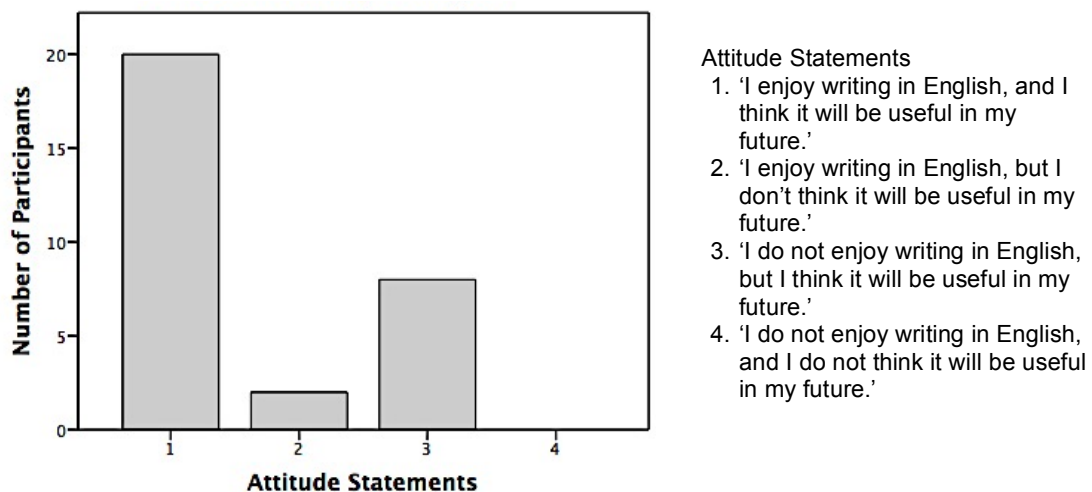
3.2 Participants

The participants were 30 undergraduate students at a private Japanese university. They came from a population of approximately 400 third-year students in the School of Policy Studies. All of the participants were native Japanese speakers and all were 20 to 21 years old. Twenty-one of the participants were female, and nine were male. They represented a range of English learning backgrounds. All had studied English for between six and fifteen years, although most had studied for eight to ten years. All had recently completed the School of Policy Studies' English Language Programme (ELP), but none studied English in any formal sense for the duration of the study. Ten of the participants had lived for six months or more in an English-speaking country, with one student having spent half his life in the United States. This variation meant a broad range of English language proficiency among the participants.

There was an initial concern that the participants might be less-motivated EFL students as they had chosen to take only Japanese-medium classes in their third year. This proved to be misplaced. According to an entry questionnaire, which asked students to choose one of four statements that best aligned with their attitude to writing in English (see Appendix B), the majority were favourably disposed to writing in English, and all expressed some degree of intrinsic or extrinsic motivation for doing so (see Figure 3.1).

The participants were paid volunteers recruited through noticeboard advertisements posted in the university, and by way of a flyer which several Japanese-medium lecturers agreed to distribute to their students. Although this sampling procedure came at the expense of some of the local generalisability afforded by intact classes, it was considered appropriate for three reasons.

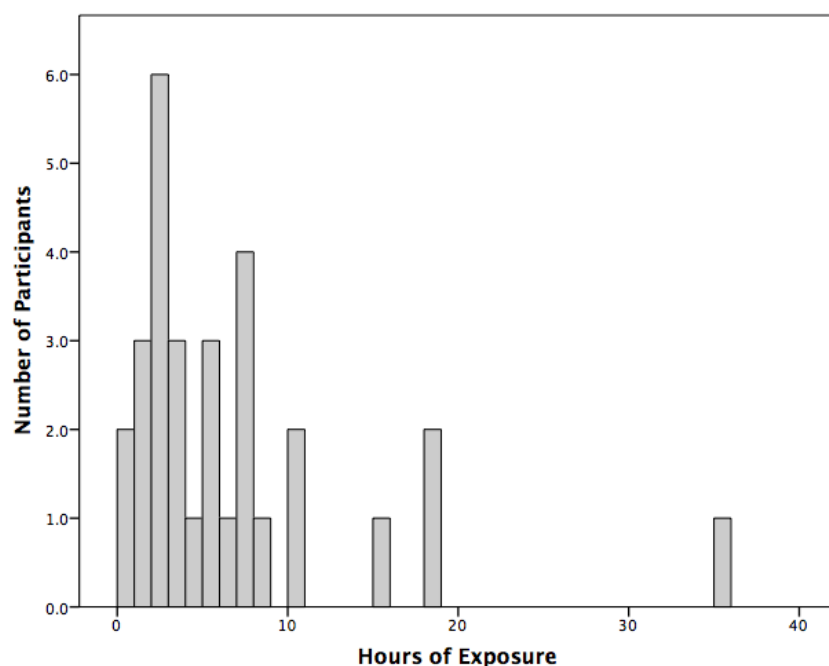
Figure 3.1: *Participant Attitudes to Writing in English.*



The first reason was a matter of logistics. The author's ELP writing class was too small to provide an adequate sample size. Given that the ELP is a tightly coordinated programme, in which all classes teach from the same materials at the same pace, it would have been a significant imposition to ask a second instructor to participate in the research. Secondly, it allowed for the use of a control group without the ethical concerns this would engender in a classroom study. Finally, because the final sample of participants had already completed the ELP, by avoiding the use of intact classes it was possible to ensure that none of the participants received any language instruction for the duration of the study. Much of the existing literature in this field may be confounded by the fact that the participants are English language students, and as such receive substantial language input during the study. This can make it unclear as to what is being measured in these studies. This creates a unique niche for the current study in the research base. This said, it is very difficult to control entirely for input. In the pre-study questionnaire students were asked how many hours of exposure to English they averaged per week, including such activities as reading for pleasure, listening to music in English and watching English language media. As Figure 3.2 shows, the results were positively skewed, with five participants reporting no exposure to English in an average week, and one reporting an average of 35 hours.

While this shows that the study was not controlled for language input entirely, we can say for certain that none of the participants received any formal instructional input or WCF outside the study. Moreover, because the participants were randomly assigned to the treatment and control groups, informal exposure to English was fairly evenly distributed between the three groups.

Figure 3.2: *Participants' Average Exposure to English (Hours Per Week)*



Although the study was not classroom-based, care was taken to replicate the conditions of the writing classes the students had encountered while in the ELP to the extent that this was possible. Most specifically, the meetings took place in the department's computer rooms, and all writing was done using computers. Error correction settings were disabled. This is not to suggest that the study was somehow conducted in both controlled and naturalistic settings, only that the actual act of writing was performed in conditions that resembled those to which the participants had become accustomed in their ELP writing classes.

It is worth noting that the setting was EFL in nature, not ESL. Only a small proportion of recent studies (R. Ellis et al., 2008; Farrokhi & Sattarpour, 2012) have been carried out in an EFL context, and the current study is possibly unique in comparing comprehensive direct WCF with comprehensive indirect WCF in an EFL setting. The EFL setting limited the amount of incidental exposure the participants had to English for the duration of the study.

Initially there were 32 participants, who were divided into two experimental groups (indirect feedback ($n=11$) and direct feedback ($n=10$)) and one control group ($n=11$). Two participants withdrew after the pre-test, leaving the indirect feedback group as the largest ($n=11$), followed by the direct feedback group ($n=10$) and the control group ($n=9$). All data analysis used this final sample of 30 participants. Based on the results of a recent Test of English as a Foreign Language (TOEFL) the participants had

taken, five different proficiency groups of roughly equal size were created. The participants within these groups were then distributed randomly among the treatment groups to prevent any English language proficiency disparity between them. Given the non-random sampling and the small sample size, the assumptions for parametric testing were not met. A Kruskal-Wallis rank sum test was therefore run on the TOEFL scores to determine that there were no significant initial differences between the proficiency levels of the three groups, $H(2)=0.03$, $p=.99$. The same test was applied to the pre-test error ratios, and this also suggested no significant language proficiency disparity between the three groups, $H(2)=0.69$, $p=.71$.

3.3 Ethical issues

One of the reasons the study was set outside the classroom was to avoid the ethical issues presented by the use of a non-treatment control group in classroom studies. This is also one of the reasons the participants were paid. The assumption was that paid voluntary participants would not be concerned by which group they were placed in. However, some of the control-group participants expressed concern that they had hoped the study would give them the opportunity to learn, something they did not think would be possible without WCF.

This presented an unexpected complication. Not only was it ethically problematic, but it also introduced the possibility that the motivation of the participants in the control group might be negatively affected, which in turn could have affected their writing performance. The solution was to offer the control-group participants one-to-one feedback on their four essays after the study was concluded. This appeared to satisfy the participants. In the questionnaire completed at the end of the final session, none said that they found their WCF demotivating, and four of the nine members of the control group took advantage of the offer.

3.4 Treatments

It has been suggested that certain error types should not be included in WCF research. Truscott (2007) argued that orthographic errors, particularly spelling, are very amenable to correction, and so will tend to give an overly favourable impression of WCF in general in studies where distinctions are not made between error types. However, while this may be true of revision studies, it seems much less likely to be the case for studies into the effects of written feedback on new pieces of writing. In fact, it seems more likely that

corrections of spelling will be ineffective, given the small likelihood that the words learners misspell will occur again in later texts. Nevertheless, given Truscott's prominence as a critic of WCF, the decision was taken for the current study to compare the effects of WCF on orthographic and non-orthographic errors. This affected data analysis only, and had no bearing on the WCF the participants received.

Much has also been made of the distinction between WCF for 'treatable' errors and for more idiosyncratic error types. Ferris (1999), Ferris and Roberts (2001), and Bitchener et al. (2005) suggest that rule-governed errors, such as the use of articles, are more likely to be amenable to WCF than idiosyncratic, context-dependent errors, such as the use of prepositions. Qualified support for this perspective comes from Van Beuningen et al. (2012), who found that only direct WCF resulted in grammatical accuracy gains in new writing, while indirect WCF benefitted non-grammatical (lexical and orthographic) errors more.

In view of these competing perspectives, a genuinely comprehensive approach in keeping with the local context was considered most appropriate. As discussed in Chapter 1, one of the aims of this study was to examine whether WCF is effective in the ways in which it is used in language classes, not under ideal conditions designed to be favourable to the treatment. For the indirect WCF group this meant a thirteen-item list of codes that designate the type of error being made (see Appendix A). This had two advantages. First, it is a system in which the participants were well versed, having used it for two years in the ELP. (One of Truscott's (1996) reservations about coded WCF was that students may not understand the feedback they are given.) Secondly, it adds to the local generalisability of the study. The approach to WCF in this study was therefore comprehensive, offering feedback on all grammatical, orthographic, and lexical errors. Table 3.2 provides a summary of the treatments.

Table 3.2: Summary of Treatments by Group

Group	WCF received	Revision task
1. Control ($n=9$)	None	Yes (self-correction)
2. Indirect WCF ($n=11$)	Error and error-type identified by code ^a	Yes
3. Direct WCF ($n=10$)	Error identified and corrected	Yes

Note. Appendices D and E show examples of indirect WCF and direct WCF, respectively.

^a See Appendix A.

3.5 Writing tasks

The tendency in recent research into WCF has been to use narratives or picture descriptions as task types (see, for example, Bitchener, 2008; Bitchener & Knoch, 2008, 2009, 2010a; Farrokhi & Sattarpour, 2012; Sachs & Polio, 2007; Sheen, 2007; Sheen et al., 2009; Shintani & Ellis, 2013; Truscott & Hsu, 2008). Another approach has had students write essays or essay-like texts (see, for example, Ferris et al., 2013; Mawlawi Diab, 2015; Van Beuningen et al., 2012). The latter method was used in this study for two reasons. The first was that the narrative tasks appear to lead, at least sometimes, to comparatively short pieces of writing. The sample text provided by Sheen et al. (2009), for example, is approximately 100 words. The hope in this study was that by using tasks that encouraged participants to write more, the greater range and number of errors would give them more to work with and learn from in the revision stage. Secondly, the choice was made to use a type of writing task that all the participants were familiar with. The descriptive essay genre had been a significant focus of the ELP writing programme, so the tasks themselves were unlikely to prove too demanding for the participants. This proved to be the case when the tasks were piloted.

The four tasks were all descriptive essays with the same thematic focus (see Table 3.3). The participants had encountered an identical theme, albeit from a more academic perspective, in the ELP. Task assignment was counterbalanced among the groups to ensure that results were not influenced by variation in task difficulty, although a pilot study suggested no significant variation. The pilot essays were of roughly equal length and with comparable error ratios. Piloting also suggested that these topics would produce errors consistent with what Japanese learners might produce in their academic writing, such as verb agreement, misuse of conjunctions, errors in relative clauses, part of speech errors, article errors, and lexical errors. They tended, with some exceptions, not to produce the tense errors characteristic of narrative writing.

Table 3.3: *Writing Tasks*

Session	Task
Task one (pre-test)	What are the features of a good parent?
Task two (practice essay)	What are the features of a good student?
Task three (post test)	What are the features of a good friend?
Task four (delayed post test)	What are the features of a good teacher?

3.6 Data collection procedures

All three groups completed four pieces of writing: a pretest, a practice essay, a posttest, and a delayed posttest. They were asked to revise the first two pieces of writing, but not the posttest nor the delayed posttest. The study was run across four one-hour sessions over a period of two months. The first three sessions took place over the course of one month and the fourth session, the delayed post-test, was one month later. Details of each session are provided below.

3.6.1 Session one (Pretest)

Session one consisted of a preamble, after which the participants signed consent-forms, and completed an entry questionnaire. Little research has been done into potential confounding variables in research into WCF, but it seems likely that L2 proficiency is a factor (Guénette, 2007). For this reason participants were asked in the questionnaire to report the score they had gained in a recent department-wide TOEFL. These scores were used after the first meeting to distribute participants into treatment groups. The remainder of the first session was spent on writing task one (the pretest). The participants were presented with the essay question and given seven minutes to generate content ideas in small groups. Finally, the participants were given thirty minutes to type their essay draft. The first written task served as a baseline measure of written accuracy and also as the first treatment. This approach has become common in research of this type (see, for example, Bitchener, 2008). Refer to section 3.5 for the writing tasks and prompts.

3.6.2 Sessions two and three

An identical approach was taken in the second and the third sessions. Session two took place two weeks after session one, and session three two weeks after session two. Essays from the previous session were returned to participants and they were given fifteen minutes to revise them based on the feedback they had been given. The control group was required to self-correct, as they received no WCF.

The importance of this revision stage was discussed in depth in sections 2.3 and 2.4. There is reason to believe that WCF can only be effective if learners are required to attend to it by way of making revisions. This may be particularly true of comprehensive and indirect feedback, but little recent research has required participants to make revisions. During the revision stage, care was taken to replicate as closely as possible the approach taken in ELP writing classes. Participants were allowed access to online tools

and dictionaries, just as they were in their writing classes (use of translation websites was not permitted). The researcher circulated and assisted with any problems participants had with the WCF, just as in ELP writing classes. It is important to point out here that this does not undermine the validity of the study. The intention is not to measure the effect of WCF on revisions, but the effect of engagement with WCF on new writing. Accessing online tools and dictionaries is a means by which students in the study's context respond to WCF.

The remainder of the second and third sessions was spent on writing tasks two and three. Participants were presented with the essay question then given approximately seven minutes to generate content ideas in small groups. Finally, the participants were given thirty minutes to type their essay draft. Writing task three served as the post-test. See section 3.5 for the writing tasks and prompts.

3.6.3 Session four (Delayed posttest)

Session four took place one month after session three, and served as the delayed posttest. No revision of previous drafts was required. The participants were presented with the essay question then given approximately seven minutes to generate content ideas in small groups. The participants were then given thirty minutes to type their final essay. See section 3.5 for the writing tasks and prompts.

3.7 Data processing procedures

All rating of the essays was done by the researcher. The rating procedure was blinded. In order to do this, the texts were randomised and initially checked using indirect coded feedback. The error ratios used in analysis were calculated from these texts. Once the rating was completed, fresh copies were made, and direct feedback was given to the corresponding group, while the control group was left without feedback.

3.8 Reliability measures

Intra-rater reliability was established by re-rating 10% of the total number of essays three weeks after the initial rating. A strong degree of correlation was found by means of an intraclass correlation coefficient [$ICC = .99$, 95% CI (.96, .99), $p < .001$]. The same procedure was used to establish inter-rater reliability. An experienced colleague was asked to rate 10% of the essays. The error ratios were correlated with those of this study's rater. This also found a high level of agreement, [$ICC = .99$, 95% CI (.91, .99), $p < .001$].

3.9 Measuring the outcome variables

A number of approaches have been taken to measuring accuracy in WCF research, but two have become most common. Recent studies looking at focused feedback for article use have tended to apply obligatory use analysis, whereby accuracy is measured by the use of articles when the context requires it (Bitchener, 2008; Bitchener & Knoch, 2009, 2010a; Farrokhi & Sattarpour, 2012; Shintani & Ellis, 2013). Given the comprehensive approach to WCF in this study, this method was not appropriate, as most errors are not a question of obligatory use. Form and inappropriate use must also be considered. This study therefore measured written accuracy by means of error ratios, an approach which has been employed by Truscott and Hsu (2008), Chandler (2003), and Van Beuningen et al. (2012). More specifically, given the possibility that some participants would not reach the target of 250 words per task, Van Beuningen et al.'s technique was used because it is appropriate for texts of any length: $[\text{number of errors}/\text{total number of words}] \times 10$.

Error ratios were also generated for two distinct types of error: orthographic and non-orthographic errors. Orthographic errors included spelling, punctuation, and capitalisation, while non-orthographic errors included the other ten categories presented in Appendix A. This was calculated in the same way as for the overall error ratio: $[\text{number of orthographic errors}/\text{total number of words}] \times 10$ for the former, and $[\text{number of non-orthographic errors}/\text{total number of words}] \times 10$ for the latter. This was to investigate Truscott's (2007) claim that orthographic errors are amenable to improvement through WCF, but non-orthographic errors are not. If Truscott is correct, then overall error ratios will not necessarily be a good indicator of general gains in accuracy. As it transpired, the data were not amenable to investigating this question, and so the focus on orthographic and non-orthographic error types was dropped during the data processing stage (see section 3.10 for further discussion of this decision).

3.10 Data analysis procedures

First, the three error ratios (total errors, orthographic errors, and non-orthographic errors) were calculated for the first draft of each of the four writing tasks, but not for the revised tasks, as this study was not intended to measure the effects of WCF on revision.

The results here lead to a reduction in the scope of the study. As the descriptive statistics show (see Appendix F), the mean error ratios for orthographic errors were so low that comparison between orthographic and non-orthographic error rates became effectively meaningless. Not only did the data produce a floor effect, whereby the mean

orthographic error rates were close enough to zero that there was very little variation, but it also meant that the mean non-orthographic error rates were sufficiently similar to the overall error rates as to be functionally the same. In other words, the low orthographic error rates were such that the entire study essentially became a measure of the effect of WCF on non-orthographic error rates. For this reason, the decision was taken to dispense with research questions 6 and 7 (see section 1.2), which focused on the relative efficacy of WCF for orthographic and non-orthographic errors. These questions were intended to investigate Truscott's (2007) claim that only orthographic errors can be affected by WCF, but the data were not suitable. However, it does not follow that the study could therefore have nothing to say about these questions. Given that the overall error rates were comprised almost entirely of non-orthographic errors, which Truscott contends are not amenable to WCF, any significant reduction in error rates would be an indication that Truscott was not correct.

The remaining data called for a two-way 3×4 mixed ANOVA. The two factors were time (i.e. essays 1 to 4) and condition (i.e. group). Time was a repeated-measures within-group independent variable, and condition a between-group independent variable. The dependent variable was error ratios.

Due the sample's non-random nature, its small size ($N=30$), the non-normal distribution of data, and its lack of sphericity, the assumptions for a parametric ANOVA were not met. Although Levene's tests showed that homoscedasticity was satisfied for the between-group factor⁴, the data violated the assumption of sphericity for the repeated-measures factor. Mauchley's test indicated that for Group 1 ($W=0.11$, $p=.005$) and Group 2 ($W=0.22$, $p=.024$) the condition of sphericity was violated, which is telling given that the test lacks power for small sample sizes (Field, Miles, & Field, 2012). Furthermore, a series of Shapiro-Wilk tests of normality by group for each essay showed that nearly half of the data subsets differed significantly from a normal distribution (see Appendix G). Even allowing for the potential for familywise error a series of tests represents, it is clear that the data violate the assumption of normal distribution. A related problem was that of outliers (see Appendix H). A small number of participants had unusually high error rates for each of their writing tasks.

⁴ Pretest: $F(2,27)=0.26$, $p=.77$; Essay 2: $F(2,27)=0.48$, $p=.62$; Posttest: $F(2,27)=0.16$, $p=.84$; Delayed posttest: $F(2,27)=0.14$, $p=.86$

One possible response to data of this nature is to adopt non-parametric testing. However, no non-parametric equivalent of a two-way mixed ANOVA exists (Field et al., 2012), so the researcher opted for a robust mixed ANOVA with trimmed means and bootstrapping. Robust tests are resistant not only to non-normal distribution and outliers, but also to the loss of information inherent to traditional nonparametric testing (Larson-Hall, 2010; Larson-Hall & Herrington, 2009). There is also evidence that bootstrapping in repeated measures designs controls for Type I errors from data that lack sphericity better than more traditional adjustments, such as Geisser-Greenhouse (Berkovits, Hancock, & Nevitt, 2000). In fact, there is an argument that in most cases robust statistics are preferable to both standard parametric and nonparametric testing, due to the former's susceptibility to even small deviations from normality, and the latter's vulnerability to type II error (Field, 2013; Field et al., 2012; Larson-Hall, 2010; Larson-Hall & Herrington, 2009; Wilcox, 1997, 1998, 2010). A drawback of this approach is that there is no single robust test with trimmed means and bootstrapping which can support both a between-subjects variable and a within-subjects variable. However, the feeling was that the benefits provided by robust testing with bootstrapping and trimmed means outweighed the problems introduced by multiple tests. Three tests were therefore carried out using the `sppba`, `sppbb` and `sppbi` functions in the `WRS2` package (Mair, Schoenbrodt, & Wilcox, 2015) within the R statistical software environment (R Core Team, 2016), with a Bonferroni correction to reduce the likelihood of this resulting in familywise error.

A final consideration was effect sizes. The choice here was whether to focus on omnibus effect sizes or individual contrasts within the study. In other words, is it preferable to report generalised overall effects, or to look more closely at how each level of each independent variable compares with each of the other levels? Kline (as cited in Larson-Hall, 2010) suggests that using both is redundant. Although it adds to the complexity of the data presentation, the decision was taken to focus on individual contrasts for three reasons. The first is that eta-squared, which is the omnibus effect size most commonly associated with ANOVA, tends to have a strong positive bias when small sample sizes are involved (Larson-Hall, 2010). Secondly, the approach taken for the robust ANOVA outlined above is not amenable to the omnibus effect sizes because it is based on three tests. Finally, and probably most importantly, focussing on individual contrasts is arguably more useful for a mixed design in that it gives a clearer picture of exactly where the largest effects are.

Chapter 4: Findings and discussion

4.0 Introduction

This chapter presents the results of the data analysis in this study. The chapter begins with a brief discussion of the descriptive statistics before moving on to the inferential statistics, and then an examination of effect sizes. Finally, the significance of the data for the research questions in this study is discussed, contextualised, and interpreted.

4.1 Findings

Descriptive statistics for error ratio by session, split by group, are shown in Table 4.1. These data were analysed using the robust mixed ANOVA described in section 3.10. Trimmed means were automatically generated and applied by the sppba, sppbb, and sppbi tests using a modified M-estimator based on Huber's Psi (Mair, Wilcox, & Schoenbrodt, 2015). This method identifies and removes problematic outliers (Wilcox, 2012). 2,000 bootstrap samples were taken. Initially α was set at .05, but then adjusted using a Bonferroni correction to account for the fact that three tests were used. Final $\alpha=.016$.

Table 4.1: *Descriptive Statistics for Total Error Ratio by Group*

Group	Session	<i>n</i>	Minimum error ratio	Maximum error ratio	Mean error ratio	Std. Deviation
Control	Pretest	10	0.51	2.25	1.25	0.53
	Essay 2	10	0.65	2.98	1.45	0.76
	Posttest	10	0.57	2.87	1.30	0.79
	Delayed posttest	10	0.70	2.29	1.28	0.49
Indirect WCF	Pretest	11	0.30	1.86	1.07	0.49
	Essay 2	11	0.43	2.52	1.02	0.58
	Posttest	11	0.33	2.98	1.11	0.72
	Delayed posttest	11	0.24	2.86	1.03	0.69
Direct WCF	Pretest	9	0.39	2.23	1.15	0.64
	Essay 2	9	0.42	2.33	1.16	0.68
	Posttest	9	0.29	2.14	1.08	0.61
	Delayed posttest	9	0.70	2.36	1.39	0.56

4.1.1 Inferential statistics

Table 4.2 shows the main effect for group on error ratio. It shows whether error ratios for the three treatment groups differed in a statistically significant sense. At $p=.46$, the test was well short of significance, even without the Bonferroni correction. Based on this, the

null hypothesis, that mean error ratios were not affected by treatment type, cannot be rejected.

Table 4.2: *Main Effect for Group*

DFn	DFd	Comparison (Group × Group)	$\bar{\psi}$	<i>p</i>
2	27	Control × Indirect	0.22	.46
		Control × Direct	0.11	
		Indirect × Direct	-0.12	

Table 4.3 shows the main effect for essay. This shows the level of significance in error ratios across time, irrespective of treatment group. Again, the results are not statistically significant ($p=.49$). In other words, there was no statistically significant variation between any combination of pretest, essay two, posttest, or delayed posttest.

Table 4.3: *Main Effect for Time (Essay)*

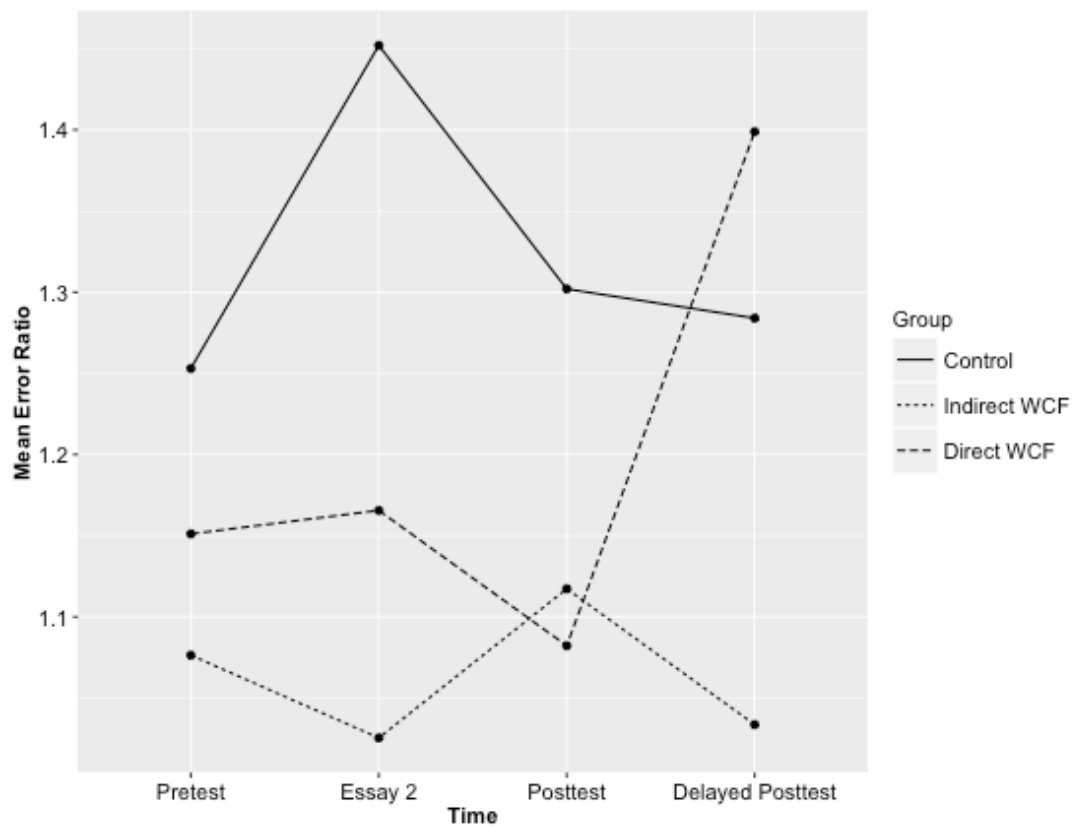
DFn	DFd	Comparison (Essay × Essay)	$\bar{\psi}$	<i>p</i>
3	81	1 × 2	-0.08	.49
		1 × 3	0.07	
		1 × 4	<0.01	
		2 × 3	0.09	
		2 × 4	-0.03	
		3 × 4	-0.07	

Perhaps most important in a mixed design of this type is the interaction effect for the two independent variables in the study. Table 4.4 shows the combined effect for group and time, and as such reflects both within-groups and between-groups measures. If the results for this test were significant, it would suggest one of two things. First, that for at least one of the three groups, the mean error ratio changed significantly more over the course of the study than for the others. Secondly, it could be an indication that written accuracy in one (or more) of the three groups improved significantly over time relative to itself. However, the results for this test were even further from significance than from the previous two ($p=.69$). The null hypothesis cannot therefore be rejected: there is no difference in change in accuracy over time between any of the three groups, and none of the groups improved significantly over time. The variation in accuracy over time for each of the three groups is represented visually in Figure 4.1. Because of the lack of significance for all three tests, post hoc analysis was not necessary.

Table 4.4: *Interaction Effect for Group × Time*

DFn	DFd	Group	Interaction (Group/Essay × Group/Essay)	$\hat{\Psi}$	<i>p</i>
6	81	Control	1/1 × 1/2	-0.15	.69
			1/1 × 1/3	0.12	
			1/1 × 1/4	0.27	
			1/2 × 1/3	-0.12	
			1/2 × 1/4	-0.02	
			1/3 × 1/4	0.10	
		Indirect WCF	2/1 × 2/2	-0.24	
			2/1 × 2/3	0.22	
			2/1 × 2/4	0.45	
			2/2 × 2/3	0.19	
			2/2 × 2/4	-0.08	
			2/3 × 2/4	-0.27	
		Direct WCF	3/1 × 3/2	0.11	
			3/1 × 3/3	0.27	
			3/1 × 3/4	0.16	
			3/2 × 3/3	-0.20	
			3/2 × 3/4	0.11	
			3/3 × 3/4	0.31	

Figure 4.1: *Error Ratios Over Time by Group*



Note. Lower values denote greater accuracy.

4.1.2 Effect sizes

Null hypothesis significance testing only speaks to probability; it has nothing to say about the magnitude of the effect of any given treatment. In other words, it is possible for a result to be statistically significant but insubstantial, or non-significant but important (P. D. Ellis, 2011). For a study such as this, therefore, which has returned non-significant results and has a small sample, effect sizes are of substantial value. They can point us to effects that inferential statistics may miss.

As discussed in section 3.10, this study reports effect sizes for individual focuses. Firstly, the between-groups focuses are presented in Table 4.5. A visual representation of these contrasts is given in Figure 4.2. These data show effect sizes for the between-group comparisons for each essay. Based on J. Cohen's effect size benchmarks (1988), there is only one pattern which might point to a notable between-groups effect; there is a consistent difference in means in favour of the indirect WCF group over the control group. This is manifested in a small effect for the pretest ($d=0.35$), a medium effect for essay 2 ($d=0.63$), a small effect for the posttest ($d=0.25$), and a small effect for the delayed posttest ($d=0.42$). This could be construed as indirect WCF having a positive effect on accuracy over no WCF at all. However, the fact that this effect was present in the pretest, and therefore preceded any treatments, suggests that it is more likely to be indicative of a small disparity in English proficiency between the two groups that was not detected by the earlier inferential tests intended to control for proficiency.

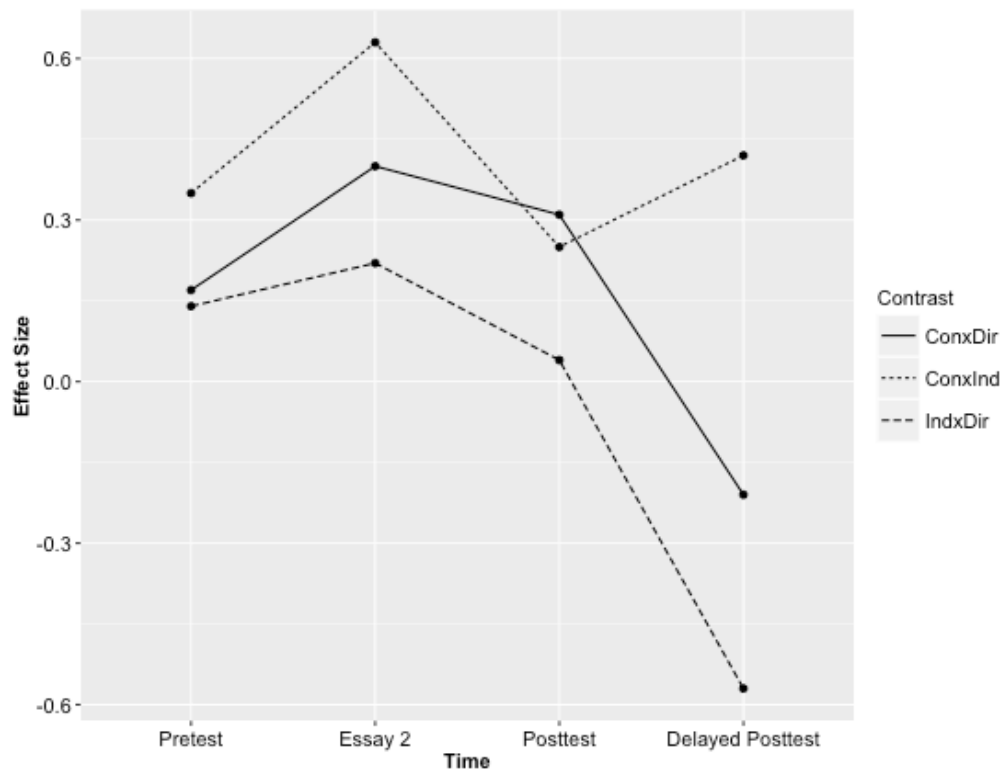
Table 4.5: *Effects Sizes for Between-Group Comparisons by Time*

Time (Essay)	Comparison (Groups)	d^a	95% CI
1	Control×Indirect WCF	0.35	-0.57, 1.27
	Control×Direct WCF	0.17	-0.77, 1.12
	Indirect WCF×Direct WCF	0.14	-1.09, 0.80
2	Control×Indirect WCF	0.63	-0.31, 1.57
	Control×Direct WCF	0.40	-0.55, 1.35
	Indirect WCF×Direct WCF	0.22	-1.19, 0.75
3	Control×Indirect WCF	0.25	-0.67, 1.17
	Control×Direct WCF	0.31	-0.64, 1.26
	Indirect WCF×Direct WCF	0.04	-0.92, 1.01
4	Control×Indirect WCF	0.42	-0.50, 1.35
	Control×Direct WCF	-0.21	-1.16, 0.74
	Indirect WCF×Direct WCF	-0.57	-1.56, 0.42

^a Negative values denote overall higher accuracy for the fixed variable (i.e. the group listed first in the contrast).

There are also less obvious patterns that are of interest. The relationship of the direct WCF group to both of the other groups follows a very similar trajectory (see Figure 4.2). For essay two and the posttest the direct feedback group outperformed the control group ($d=0.40$ and $d=0.31$, respectively), but the reverse was true for the delayed posttest ($d=-0.21$). Comparing the indirect WCF and direct WCF groups shows a small effect for direct WCF over indirect for essay two ($d=0.22$), which became trivial by the posttest ($d=0.04$), and then became a medium effect size in favour of the indirect WCF group in the delayed posttest ($d=-0.57$). What these data suggest is that over time both the indirect WCF group and the control group improved in accuracy, where the direct WCF did not.

Figure 4.2: *Effect Sizes for Between-Group Comparisons by Time*



Note: Negative values denote overall higher accuracy for the the group listed first in the contrast.

However, the effect sizes for within-groups comparisons show that this cannot be the case (see Table 4.6). These data show that in the case of the indirect WCF group at least, results were very consistent. All of the effect sizes within the indirect group are trivial, which suggests no change in accuracy over the course of the study. What actually occurred is that none of the groups improved, but the direct feedback group declined in accuracy substantially between the posttest and the delayed posttest (Figure 4.1 gives a clear visual representation of this). It is difficult to see how this can be attributed to the

treatment. The decline does not suggest a gain in accuracy that was subsequently lost in the delayed post-test for two reasons. Firstly, the gains made by the direct WCF group from the pretest to essay two and the posttest were trivial ($d=0.02$, $d=0.11$, respectively), so there was effectively no improvement. Secondly, the accuracy in the delayed posttest was substantially below even that of the pretest ($d=-0.40$). This was an actual decline in accuracy, not a decline relative to earlier gains. The notion that direct WCF could lead to a sudden, delayed decline in accuracy is counter-intuitive. It seems more likely that this was either a chance result, or that it reflects waning interest on the part of the direct feedback group at the end of the study.

Table 4.6: *Effect Sizes for Within-Group Comparisons by Time*

Group	Comparison (Time)	d^a	95% CI
Control	1×2	-0.31	-1.25, 0.64
	1×3	-0.07	-1.01, 0.87
	1×4	-0.06	-1.00, 0.88
	2×3	0.19	-0.75, 1.14
	2×4	0.27	-0.68, 1.21
	3×4	0.03	-0.91, 0.97
Indirect WCF	1×2	0.09	-0.80, 0.98
	1×3	-0.06	-0.95, 0.82
	1×4	0.07	-0.82, 0.96
	2×3	0.14	-1.03, 0.75
	2×4	0.02	-0.91, 0.87
	3×4	0.11	-0.78, 1.00
Direct WCF	1×2	0.02	-1.01, 0.98
	1×3	0.11	-0.89, 1.11
	1×4	-0.40	-1.38, 0.58
	2×3	0.12	-0.88, 1.12
	2×4	-0.37	-1.38, 0.61
	3×4	-0.53	-1.55, 0.49

^a Negative values denote an overall decline in accuracy over time.

The effect sizes for the control group were similarly influenced by a single unusual result, not by a pattern of improvement or decline. Accuracy declined in essay two for the control group. This led to small positive effect sizes from essay two to the posttest ($d=0.19$) and the delayed posttest ($d=0.27$). Again, however, this is due to markedly lower accuracy in one essay, not to a pattern of improvement. Effect sizes show a trivial decrease in accuracy for the control group if the pretest is compared to the posttest

($d=-0.07$) and the delayed posttest ($d=-0.06$). There is no pattern of change in accuracy in the control group.

Effect sizes for this study, therefore, are in keeping with the inferential statistics. Firstly, there is no discernable pattern of change between any of the three groups. The most likely explanation for any between-group differences is a disparity in English proficiency (between the control and indirect feedback groups), and chance or waning interest (in the case of the direct feedback group's poor performance in the delayed posttest). This is even more likely given the lack of consistent positive effect sizes for the within-groups comparisons. With no change in accuracy within groups, we would expect to find no difference in accuracy between them beyond what they started with. Effect sizes over time in the indirect WCF group were very consistent in that none of them amounted to even a small effect based on J. Cohen's (1988) framework. The direct feedback group and the control group produced a little more variety in this sense, but this was due entirely to a single unusual result for each. There is no suggestion of a pattern of change for either group. In other words, both the inferential statistics and the effect sizes strongly support a conclusion of no effect for either treatment in this study.

4.2 Implications of the findings for the research questions

If we take the results discussed in section 4.1 at face value, the answer to each of the research questions (presented again in Table 4.7) is no. In other words, the null hypothesis cannot be rejected for any of the research questions investigated in this study.

Table 4.7: *Research Questions Investigated in this Study*

1.	Do students who are required to correct grammatical errors in their writing by way of comprehensive, explicit, indirect WCF make fewer grammatical errors over time?
2.	Do students who are required to correct grammatical errors in their writing by way of comprehensive, direct WCF make fewer grammatical errors over time?
3.	Do students who are required to correct grammatical errors in their writing by way of comprehensive, explicit, indirect WCF make fewer grammatical errors over time than students who receive no WCF?
4.	Do students who are required to correct grammatical errors in their writing by way of comprehensive, direct WCF make fewer grammatical errors over time than students who receive no WCF?
5.	Do students who are required to correct grammatical errors in their writing by way of comprehensive, explicit, indirect WCF make fewer grammatical errors over time than students who are required to correct grammatical errors in their writing by way of comprehensive, direct WCF?

4.2.1 On the efficacy of explicit, comprehensive, indirect WCF

Research questions 1, 3, and 5 deal explicitly with indirect WCF. Question 1 relates to the efficacy of indirect WCF measured within groups. Question 3 and 5 relate to how indirect WCF fares relative to no feedback and direct feedback, respectively. Tables 4.4 and 4.6 show that the null hypothesis cannot be rejected for research question 1. There was no significant improvement in accuracy over time for the indirect WCF group, nor is there any suggestion in the effect sizes of any pattern of improvement that inferential testing was unable to detect. The results are the same when we consider the value of indirect WCF in the context of research questions 3 and 5. Relative to both the direct WCF group and the control group, there was no statistically significant change in accuracy (see Table 4.2). Nor did the effect sizes point to any substantive pattern of change in accuracy in comparison to the control and direct WCF groups (see Table 4.5). In simple terms, explicit, comprehensive, indirect WCF had no effect on written accuracy over the course of the study.

Although the effect of indirect coded WCF is not a well-researched area, the results of this study contradict what much of the literature has previously suggested about indirect WCF. Since 1982 at least six studies have looked directly at the effects of indirect WCF of the type used in this study, and only one is in broad agreement with the current study. Robb, Ross, and Shortreed (1986) compared direct WCF, implicit indirect WCF, and explicit WCF and found no effect for any of them. Robb, Ross, and Shortreed's study shared a similar context to the current study (i.e. an EFL setting in a Japanese university), which is interesting given the similarity of the results. It is possible that there is something in the setting of these two studies that led to their non-significant results, a point that will be discussed further in section 4.3.

Robb, Ross, and Shortreed (1986) aside, the research has tended to support at least some efficacy for indirect WCF. Lalande (1982) found that indirect WCF outperformed direct WCF, but that the difference was not statistically significant. Sheppard (1992) found that although "holistic" WCF was more effective than indirect, indirect WCF nevertheless effected gains in accuracy. Similarly, Chandler (2003) found that indirect WCF had a positive effect on accuracy in student writing, although the effects of both implicit WCF and direct WCF were greater.

It should be noted that the four studies discussed above all had significant flaws that undermine their conclusions, not least of which was the lack of a legitimate control group in all cases. Two more recent studies, however, were more rigorous and still had

different findings with respect to indirect WCF than the current study. Van Beuningen, et al. (2008) found an effect for both indirect WCF and direct WCF, although the effect for direct feedback had greater longevity. The same authors (Van Beuningen et al., 2012) again found an effect for both indirect and direct WCF, although their effectiveness for different categories of error varied. The null results of the current study with respect to the efficacy of indirect WCF were therefore unexpected. Possible reasons for these results will be discussed in section 4.3.

4.2.2 On the efficacy of comprehensive, direct WCF

Research questions 2, 4, and 5 relate directly to the efficacy of comprehensive, direct WCF. Functionally, the results are identical to those for indirect WCF. Table 4.4 shows that the direct feedback group made no statistically significant improvements in accuracy over the course of the study. Although the within-group effect sizes were more varied than those for indirect WCF (see table 4.6), there was no pattern of improvement in accuracy for this group. If anything the direct WCF declined in accuracy over the course of the study. There was a small negative effect size between the pretest and the delayed posttest, and a medium negative effect size between the posttest and the delayed posttest. However, as was suggested in section 4.1, this makes more sense as a single aberrant result than as a genuine decline. Overall the data for the direct WCF group tell the same story as for the indirect WCF group; the treatment had no effect on written accuracy, either in isolation or relative to the other groups.

In the context of the existing research, the results of this study with respect to direct WCF are even more unexpected than for indirect WCF. While it was argued in Chapter 2 that indirect WCF had been abandoned prematurely, the fact is most of the recent research focuses on direct WCF, and most of it shows a significant effect for this form of feedback on accuracy. In the last ten years alone at least eleven studies have found some effect for direct WCF (Bitchener, 2008; Bitchener & Knoch, 2008, 2010a, R. Ellis et al., 2008, 2008; Farrokhi & Sattarpour, 2012; Mawlawi Diab, 2015; Sheen, 2007; Sheen et al., 2009; Van Beuningen et al., 2008, 2012). Clearly the current study is at odds with a preponderance of the research. Possible reasons for this will be discussed in section 4.3.

4.3 Interpretation

When the results of a study stand in such stark contrast to much of the published research it is natural to speculate as to the reasons. There are a number of possible explanations for the results of the current study, some of which are general reflections of the idiosyncrasies of quantitative research, and others that are specific to the field of written corrective feedback.

One important qualification is that the results of the current study may not be as unusual as they appear in the context of the published research. The term “file drawer problem” describes the tendency of journals and researchers to favour studies with positive results over those with null results. Rosenthal, who coined the term (1979, p. 638), characterises the “extreme view of this problem” as the possibility that “the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show nonsignificant...results”. While Rosenthal is exaggerating, his point is that the published research in many fields may be misleading because of this preference for positive results. There is no reason to believe that the research into WCF is any exception. This point is made not to preclude the need for an explanation of the results of the current study, only to suggest that it may not be as much of an outlier as it appears.

One possible cause of the null results relates to a limitation this study: the nature of the sample. While the initial sample size of 32 is not particularly unusual in applied linguistics research, this was compounded by the loss of two participants early in the study, and the fact that the sample was split into three smaller groups. A further factor in this sense was the non-normal distribution of the data collected from this sample, due not only to the sample size, but also to its nonrandom selection process. A sample of this nature compromises the power of statistical tests to detect an effect, and so it is possible that the results of this study are due to type II error. There are however two factors that reduce this likelihood. The first is the use of robust statistical tests, which have been shown to increase power with small, nonrandom samples (Wilcox, 1998, 2010). The fact the current study failed to find significant results where other studies with similar sample types using less appropriate parametric testing have had positive results (see, for example, Chandler, 2003 ($N=31$), R. Ellis et al., 2008 ($N=49$), and Shintani and Ellis, 2013 ($N=49$)) suggests that this is not simply a case of type II error. When parametric tests are used in situations where their assumptions have been violated, type II error becomes more likely. Studies like those of Chandler (2003) would therefore be *more*

likely to generate type II errors than the current study. This does not discount the possibility of type II error in the current study, but makes it less likely.

The second factor that suggests this study has probably not missed an effect due to lack of power is the effect sizes reported in tables 4.5 and 4.6. Unlike inferential statistics, effect sizes are not particularly sensitive to sample size (Clark-Carter, 2003). As discussed in section 4.1.2 there is nothing in the effect sizes for this study to suggest any improvement in written accuracy for any of the three groups. At this point, the question of whether this study generated type II errors becomes academic. Even if inferential statistics had uncovered a statistically significant effect, it would have been rendered inconsequential by the trivial effect sizes.

It seems unlikely therefore that the non-significant results of this study were due to error. The alternative explanations are more interesting in that they all have something to tell us about the relationship between WCF and written accuracy. The first of these is the possibility that Truscott was right. There are two ways in which this statement can be interpreted in the context of the current study. The first is Truscott's claim, first made in 1996 but often repeated since then, that written corrective feedback does not and cannot effect greater accuracy in writing. Ostensibly, the results of this study support the first part of Truscott's claim. For this study, WCF clearly did not work. However, to accept this explanation would be to ignore the weight of evidence other studies have produced. In the last fifteen years, possibly the only study to have shown no effect for WCF was conducted by Truscott himself (Truscott & Hsu, 2008), and it used implicit feedback, an approach to WCF that even Truscott (2007) had previously argued was so unlikely to be effective that it effectively amounted to a control treatment.

A second way in which the current study could be viewed as supporting Truscott's point of view is with respect to the distinction between orthographic and non-orthographic errors. Truscott (2007) has made the claim that only the former may be amenable to WCF, a claim that this study was in part intended to investigate. However, this research question was dropped after it became clear that the orthographic error rate was too low to be statistically meaningful (see section 3.10). A possible implication of this is that the current study may have uncovered an effect for orthographic errors had they been frequent enough to be examined separately, but that this effect was drowned out by the noise of the non-orthographic error rate which, for Truscott, cannot be affected by WCF. Again, however, this is contrary to many other studies that have measured only non-orthographic errors and have found significant effects for WCF. Recently, for

example, a number of studies looking at highly focused WCF have all concentrated entirely on non-orthographic errors, primarily in the use of articles (Bitchener, 2008; Bitchener & Knoch, 2008, 2009, 2010a, 2010b; R. Ellis et al., 2008; Farrokhi & Sattarpour, 2012; Sheen, 2007; Shintani & Ellis, 2013), and all have found at least some effect for WCF. Although the results of the current study coincide with Truscott's claims, given the research base it is difficult to argue that this speaks to Truscott being correct.

Another possible reason that the current study failed to reproduce the results of recent studies is that it did not employ focused feedback, but instead used comprehensive feedback. There is an intuitive appeal to the argument that comprehensive feedback, be it direct or indirect, is best avoided because it is likely to overwhelm students. In fact this argument is often cited as the reason why so many recent studies have dealt entirely with highly focused feedback (see, for example, Bitchener, 2012; R. Ellis et al., 2008). If this argument is correct, then we should expect to find no effect for comprehensive WCF. However, the empirical evidence for this claim in written feedback for written errors is not strong (see section 2.4). Two of the better studies conducted in the last ten years (Van Beuningen et al., 2008, 2012) have looked at direct and indirect WCF, as the current study did, and have found an effect for both. As long as students are given sufficient time to process and act on comprehensive feedback, there seems to be no reason to think it is inherently ineffective. The comprehensive approach to WCF taken in this study does not therefore appear to be a sufficient explanation for its results.

While it is conceivable, despite the reasons discussed above, that the results of this study are due to type II error, the comprehensive approach to WCF, or the fact that WCF is actually ineffective, there is a final possibility which is arguably more compelling, and certainly more promising in terms of positioning this study in the research base. This is the possibility that what separates this study from the majority of published studies is not methods, but context. If this is the case, then this study can contribute to our understanding of the extent to which the efficacy of WCF is not a universal, and that it depends on who receives the feedback and in what setting. There are three salient points here: the study took place in a Japanese academic EFL context, it was not classroom-based, and the participants were volunteers who were not currently studying English. The final two factors here point to the possibility that a lack of participant motivation played a role in the results of the study. It should first be pointed out that some effort was made to measure motivation in both pre and post-study questionnaires, and all the participants expressed a degree of motivation, be it intrinsic,

extrinsic, or both (see sections 3.2 and 3.3). However, what this study missed was any form of short-term extrinsic motivators. Bruton (2009, 2010) has argued that a key component in the uptake of WCF is “such factors as instructions, tasks and grades [which] will make a difference and cannot be ignored” (2010, p. 496). It is these kinds of short-term-goal focused motivators that this study lacked, and which may, at least partially, explain its results.

It is difficult to say exactly what kind of effect a lack of motivation is likely to have on the uptake of WCF, as it is not a well-researched area. However one concern is that a lack of extrinsic motivators will negatively affect a learner’s “noticing” of the gaps in their interlanguage. Schmidt (2001) suggests that increased attention and noticing is one of the mechanisms by which motivation leads to success in language learning. Without motivation, a learner’s capacity for noticing is therefore reduced. For Schmidt (2001), without noticing, which derives at least to some extent from motivation, there is no learning. In terms of the results of this study specifically, it may be the case that a lack of motivation led participants to attend to WCF in what Qi and Lapkin (2001) describe as a “perfunctory” manner, which in turn meant they did not learn from WCF as they might in an authentic classroom context. The idea that motivation may play a role in the uptake of WCF is hardly novel, but further research into how short-term-goal focused motivators could be used to effect greater noticing and possible uptake of WCF seems merited based on the results of the current study.

It was remarked earlier (see section 4.2.1) that the setting of the current study had similarities to that of Robb et al. (1986), possibly the only other study to show no effect for both indirect and direct WCF. More specifically, both studies took place in a Japanese university EFL context. A survey of the WCF research base shows one more study that shares this context (R. Ellis et al., 2008). This study compared the effects of focused and comprehensive feedback on first mention and anaphoric reference article use, and found both effective, although only the effects for focused feedback extended to the delayed posttest. Unfortunately, as is common in WCF research, effect sizes were not reported, but based on the descriptive statistics published with the study, they appear to have ranged from small to medium. What we have in the Japanese EFL context then are three studies. Two found no effect for WCF, and one shows a small to medium effect for WCF on one use of one language point. EFL studies in other settings are also inconclusive. Truscott and Hsu (2008) found no effect for implicit feedback on accuracy in Taiwanese EFL students, while Farrokhi and Sattarpour (2012) found an effect for

both focused and semi-focused WCF for their “high-proficient” Iranian EFL students. The ratio of negative to positive studies in EFL settings is clearly very different to that for ESL settings. The current study has contributed to growing evidence that WCF is ineffective, or at least less effective, in EFL settings than in ESL settings. This also poses the question of what is happening in the ESL studies. Are we seeing the effects of feedback, or is it an accident of the incidental language input participants are receiving in their day-to-day lives?

Incidental input is not the only possible confounding variable in WCF studies. There is also the question of what effect classroom input is having on apparent WCF uptake. This is an issue for both ESL and EFL studies, and is something upon which the current study is uniquely positioned to offer a different perspective. Because there was no option to conduct this study in a classroom, the decision was taken instead to control for language input as much as possible. In order to do this, only participants who were not currently studying English in any formal sense were enlisted. This, combined with the EFL setting, meant that most of the students experienced little exposure to English over the course of the study. In other words, the current study is as close as we have come to a genuinely experimental study of the effects of WCF on written accuracy. From this perspective, the fact that the results of the study were non-significant and non-substantial is revealing. It hints at two possibilities. The first is that WCF alone is ineffective, and that what we see in positive studies is the result of other input, not correction. Given the fact that in many of these studies the treatment groups have outperformed control groups, this seems unlikely. More likely is the possibility that WCF and general input are interacting in ways that are facilitating greater accuracy in some cases, but not in others. Determining exactly how this interaction might work would be extremely useful to teachers, and by extension, their students.

4.4 The learners’ perspective

Although it does not speak directly to the research questions for this study, it is worth touching briefly on results of the exit questionnaire at this point (see Appendix C). These results offer insight into the participants’ perspective on the WCF given in the study.

The first point to note is that students appear to want WCF in some form. None of the indirect or direct WCF group participants felt strongly that they had received too much or too little WCF. In the non-feedback control group, on the other hand, all but one of the participants agreed outright that they had received too little feedback. This is in

keeping with other research into student expectations and WCF (see, for example, Lee, 2005), and presents a tension between the findings of the inferential statistics in the current study and student expectations. Truscott (1999) suggests that student expectations do not justify the time spent given WCF if it does not effect improved accuracy in the long term, but many teachers will feel uncomfortable at the prospect of ignoring such expectations.

The results of the questionnaire also suggested that students do not feel overwhelmed by comprehensive feedback. None of the participants in either the direct WCF group or the indirect WCF agreed outright that they had received too much feedback or that the feedback they received was demotivating. The fact that students feel this way does not preclude the possibility that comprehensive WCF asks too much of them, but without strong evidence against the efficacy of WCF it ought at least to be a consideration for instructors who are unsure whether focused or comprehensive WCF is preferable. One participant from the direct WCF group touched on this point when he wrote: “Too much or too little feedback could demotivate students. The amount of feedback may be difficult to adjust, but it is important to give the right amount for them.” We might expect a participant expressing this sentiment to feel that the comprehensive feedback he received was excessive, but he disagreed that he received too much feedback and also disagreed that the feedback he received was demotivating.

The open-ended responses favoured indirect WCF over direct WCF, with ten of the thirty participants across the three groups describing something like indirect WCF in what they thought was the best approach, for example: “...pointing out their grammatical errors by using a [coded] paper which you used. That was very easy to see and convenient to correct errors.” None of the participants described direct WCF as preferable, but without a study that exposes all participants to both approaches it is difficult to attribute real significance to this.

Finally, many of the participants felt that WCF combined with an actual meeting between the teacher and the student would be the best approach. Eight of the participants explicitly mentioned that they would like to be able to discuss their errors with their teachers. While this is probably not practicable in many contexts, taking this approach in smaller classes may be beneficial and would allow the provision of feedback to better accommodate more socially mediated conceptions of L2 acquisition, such as sociocultural theory (see section 2.1.4).

4.5 Concluding remarks

This chapter has detailed the results of the current study and has discussed the possible implications of those results. No significant variation was found either between groups or over time. Effect sizes also pointed to no substantive effect for either comprehensive, indirect WCF or comprehensive, direct WCF on accuracy in new pieces of writing.

It was suggested that there are number of factors that may account for this, including type II error, the possibility that WCF is ineffective, and the comprehensive approach taken to WCF. However, it was argued that the most likely explanation for the null results is that they are a product of the setting and design of the study. There is a suspicion that the study may not have offered the participants appropriate motivation to improve. There is also a possibility that EFL settings are not amenable to the uptake of WCF. Finally, it was suggested that the EFL setting may have combined with the controls the study imposed on language input to render the WCF ineffective. All three possibilities point to potentially useful areas for future research.

This chapter also looked briefly at the results of the questionnaire that participants completed at the conclusion of the study. The survey suggested that students feel strongly that they should receive WCF in some form, and perhaps that they have some preference for the indirect approach. It also suggested that students feel able to cope with comprehensive feedback, and that they feel they would benefit from having the opportunity to discuss their WCF directly with their teacher.

Chapter 5: Conclusions

5.0 Introduction

This chapter presents a brief summary of the key findings of this study. The implications of this study, both for teachers and researchers are discussed before going on to look at some of its limitations. Finally, concluding remarks are given on the study as a whole.

5.1 Summary of key findings

This study investigated the effects of direct and indirect WCF on student accuracy in new pieces of writing. It was unique in its efforts to control for language input as a confounding factor in WCF research. At the same time, it focused on the comprehensive approach to WCF, which was essentially a response to the large numbers of recent studies which have focused on the effects of WCF on a single language point. The author considered the comprehensive approach to be more relevant to language teachers and students.

The results of the study were ostensibly straightforward. No effect was found for either form of feedback, either through inferential statistical testing or effect sizes. Viewed within the context of the research base however, it has been argued in this thesis that there are a number of possible explanations for this beyond the simple conclusion that the study was flawed or that WCF is ineffective, including a possible lack of motivation among the participants, the EFL setting in which the study took place, and the extent to which language input was controlled throughout the study.

5.2 Pedagogical implications

There is a certain tension here between the results of this study, and the sense that these results in isolation do not tell the whole story. It would certainly be inappropriate to suggest that teachers abandon WCF on the basis of one study that is contradicted by the research base. This study arguably has more to say to teachers in a Japanese EFL context than to language teachers in general. Such teachers should be open to the possibility that comprehensive direct and indirect WCF may be ineffective. The decision of whether to engage students in the time-consuming cycle of error correction and revision should therefore be based on a number of considerations. First, how many times will I have the opportunity to give feedback to my students? This study suggests that once or twice will be insufficient for any effect. Second, what motivation do my students have to complete this task? More specifically, does the task provide the kinds of short-term goals and

rewards that this study lacked? For example, is there a grade associated with the task? Finally, what are the expectations of my students? The exit questionnaire suggested that students want feedback, but given the possibility that WCF will be ineffective for my students, is it a worthwhile use of my time and my students' time to provide WCF simply to satisfy their expectations?

5.3 Implications for further research

The implications of this study for further research are based primarily on coming to understand why it found no effect for WCF when most other studies have had positive results. In this sense further research into the elements that separate this study from most of the research base is recommended.

First, there is a need for further investigation into the effects of WCF in EFL settings. Of the five EFL-focused studies discussed in Chapter 2 and again in section 4.3, including the current study, three found no significant effect and another showed limited effects. This suggests that different populations respond to WCF differently, and the research should reflect this if it is to inform teaching outside a limited context.

Beyond population differences, this study also points to the need for further research into the interaction between individual differences and uptake of WCF, particularly motivation. While the current study was able to establish a fairly consistent level of motivation among its participants, it failed to address the kind of short-term-goal motivators that can be found in the classroom. This accident of omission, combined with the null results, suggests that these kinds of motivators may play a role in uptake of WCF that we do not yet understand.

The current study was unique in the control it exercised over language input as a confounding factor. While this imposed limitations on the study, it also helped it to identify an area about which we understand very little, but which may be very important in informing classroom practices. That this study found no effect for WCF suggests that there is some interaction between what occurs in the classroom and how effective WCF will be. This may be something straightforward, such as the extent of teacher involvement in the engagement with WCF and revision process, or it might be more subtle, such as the amount and type of form-focused instruction in the classroom. Research which attempts to isolate these factors and establish which, if any, help to facilitate learning from WCF, would unquestionably be useful.

Based on the lack of a positive result from this study, it is difficult to escape the thought that for us to genuinely understand the effects of written corrective feedback, more longitudinal research will also be necessary, both in the sense that it take place over a longer period of time, and that it include significantly more than two treatments. Much of the research into WCF, including the current study, in a sense misrepresents what motivates teachers to give WCF. It is unlikely that many teachers consider it to be a quick fix, but rather it is something that they provide to students based on the belief that through time and repetition, it will have some effect.

5.4 Limitations

One limitation of this study was the size of the sample. A post hoc analysis of the study's power suggests that for a conventional ANOVA a very large effect size ($d=1.72$) would have been required to detect significance with the group sizes this study had. The fact that this study employed robust statistics notwithstanding, none of the effect sizes reached anywhere near that point. Although the effect sizes showed that the non-significant findings were probably not due to the small sample, a small, self-selected sample limits the external validity of the study because it is unlikely to be representative of the population in general (Burns, 2000; Turner, 2014).

Ecological validity was compromised by the fact that this study was an unusual example of quasi-experimental research that took place outside the classroom. Storch (2010, p. 43) makes the point that highly controlled studies are “unlikely to be relevant to language teachers because they do not reflect real classroom conditions.” It should be noted however that it was exactly the degree of control over input exercised in this study that made it unique and was responsible for its more interesting possible implications.

A final limitation of the study is that it could at best be characterised as semi-longitudinal. This is not a limitation of this study alone, but of most research into the efficacy of WCF. In fact, where this study had two treatments and took place over a period of two months, many recent studies have included only a single treatment over a considerably shorter period of time.

5.5 Conclusion

This study was designed to contribute to the debate over the efficacy of WCF by focusing on less researched, but pedagogically relevant approaches to feedback. It investigated the effects of two forms of comprehensive WCF, direct and indirect, on

student accuracy in new pieces of writing. The thirty volunteer participants were divided into three groups: two treatment and one control, and asked to write four short essays. The first two of these essays were treatments in which the participants were asked to review and attend to what feedback they had been given. The third and fourth essays were a posttest and a delayed posttest, respectively.

Statistically, the results of this study were non-significant and non-substantial. No effect on accuracy was found for either form of feedback. It has been argued in this thesis that these results were less straightforward than they appear to be. Possible factors include student motivation, the EFL setting, and the study's control over language input.

The null results were unexpected, but despite being at odds with much of the recent research this study has a contribution to make to both pedagogy and possible directions for future research. It was suggested that teachers who share the study's Japanese EFL setting should consider whether WCF constitutes a worthwhile use of time. Should they choose to adopt it, they should avoid the factors that may have contributed to the lack of efficacy of WCF in this study.

Possible directions for future research were also proposed based on the findings of this study. The role of motivation in the uptake of WCF would be a complex but valuable avenue, as would the interaction between classroom practices and learning from WCF. Finally, it was argued that future research into the effect of WCF on written accuracy should be both longitudinal and include more treatments than is currently typical. It is the author's opinion that to the extent that WCF is genuinely effective, it will prove to be as a gradual, incremental process, based on many exposures and engagements with WCF, and which is heavily moderated by individual learner differences.

References

- Adams, R. (2003). L2 output, reformulation and noticing: implications for IL development. *Language Teaching Research*, 7(3), 347–376.
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78(4), 465–483.
- Alshahrani, A., & Storch, N. (2014). Investigating teachers' written corrective feedback practices in a Saudi EFL context: How do they align with their beliefs, institutional guidelines, and students' preferences? *Australian Review of Applied Linguistics*, 37(2), 101–122.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369–406.
- Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9(3), 227–257.
- Berkovits, I., Hancock, G. R., & Nevitt, J. (2000). Bootstrap resampling approaches for repeated measures designs: Relative robustness to sphericity and normality violations. *Educational and Psychological Measurement*, 60(6), 877–892.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118.
- Bitchener, J. (2009). Measuring the effectiveness of written corrective feedback: A response to “Overgeneralization from a narrow focus: A response to Bitchener (2008).” *Journal of Second Language Writing*, 18(4), 276–279.
- Bitchener, J. (2012). A reflection on “the language learning potential” of written CF. *Journal of Second Language Writing*, 21(4), 348–363.
- Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research*, 12(3), 409–431.
- Bitchener, J., & Knoch, U. (2009). The value of a focused approach to written corrective feedback. *ELT Journal*, 63(3), 204–211.
- Bitchener, J., & Knoch, U. (2010a). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), 207–217.
- Bitchener, J., & Knoch, U. (2010b). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31(2), 193–214.

- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191–205.
- Bruton, A. (2009). Designing research into the effects of grammar correction in L2 writing: Not so straightforward. *Journal of Second Language Writing*, 18(2), 136–140.
- Bruton, A. (2010). Another reply to Truscott on error correction: Improved situated designs over statistics. *System*, 38(3), 491–498.
- Burns, R. B. (2000). *Introduction to research methods*. Frenchs Forest: Pearson Education.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267–296.
- Clark-Carter, D. (2003). Effect size: The missing piece in the jigsaw. *The Psychologist*, 16(12), 636–638.
- Cohen, A. D., & Robbins, M. (1976). Toward Assessing Interlanguage Performance: The Relationship Between Selected Errors, Learners' Characteristics, and Learners' Explanations. *Language Learning*, 26(1), 45–66.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, N.J: L. Erlbaum Associates.
- DeKeyser, R. M. (2007). Introduction: Situating the concept of practice. In R. M. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 1–18). Cambridge: Cambridge University Press.
- Ellis, P. D. (2011). *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results* (reprint). Cambridge: Cambridge Univ. Press.
- Ellis, R. (2010). Cognitive, social, and psychological dimensions of corrective feedback. In R. Batstone (Ed.), *Sociocognitive Perspectives on Language Use and Language Learning* (pp. 151–165). Oxford: Oxford University Press.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36(3), 353–371.
- Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. London ; New York: Routledge/Taylor & Francis Group.

- Farrokhi, F., & Sattarpour, S. (2012). The Effects of Direct Written Corrective Feedback on Improvement of Grammatical Accuracy of High- proficient L2 Learners. *World Journal of Education*, 2(2), p49.
- Fathman, A., & Whalley, E. (1990). Teacher response to students writing: Focus on form versus content. In B. Kroll (Ed.), *Second Language Writing: Research Insights for the Classroom* (pp. 178–190). Cambridge: Cambridge University Press.
- Ferris, D. (1997). The Influence of Teacher Commentary on Student Revision. *TESOL Quarterly*, 31(2), 315–339.
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to truscott (1996). *Journal of Second Language Writing*, 8(1), 1–11.
- Ferris, D. (2004). The “Grammar Correction” Debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime ...?). *Journal of Second Language Writing*, 13(1), 49–62.
- Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing* (pp. 81–104). Cambridge: Cambridge University Press.
- Ferris, D., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22(3), 307–329.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10(3), 161–184.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics* (4th edition). Los Angeles: Sage.
- Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London ; Thousand Oaks, Calif: Sage.
- Frear, D., & Chiu, Y. (2015). The effect of focused and unfocused indirect written corrective feedback on EFL learners’ accuracy in new pieces of writing. *System*, 53, 24–34.
- Guénette, D. (2007). Is feedback pedagogically correct?: Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16(1), 40–53.
- Haswell, R. H. (1983). Minimal Marking. *College English*, 5, 600–604.
- Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing*, 7(3), 255–286.
- Hyland, F. (2003). Focusing on form: student engagement with teacher feedback. *System*, 31(2), 217–230.

- Jupp, V. (Ed.). (2006). *The Sage dictionary of social research methods*. London ; Thousand Oaks, Calif: SAGE Publications.
- Kepner, C. G. (1991). An Experiment in the Relationship of Types of Written Feedback to the Development of Second-Language Writing Skills. *The Modern Language Journal*, 75(3), 305–313.
- Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, 21(4), 390–403.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Lalande, J. F., II. (1982). Reducing Composition Errors: An Experiment. *The Modern Language Journal*, 66(2), 140–149.
- Lantolf, J. P., (2012). Sociocultural theory: A dialectical approach to L2 research. In *The Routledge handbook of second language acquisition* (pp. 57–72). New York: Routledge.
- Lantolf, J. P., & Thorne, S. L. (2006). Sociocultural theory and the genesis of second language development. Oxford ; New York: Oxford University Press.
- Lantolf, J. P., & Thorne, S. L. (2007). Sociocultural theory and second language learning. In B. Van Patten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (First Edition, pp. 197–221).
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York ; London: Routledge.
- Larson-Hall, J., & Herrington, R. (2009). Improving data analysis in second language acquisition by using modern developments in applied statistics. *Applied Linguistics*, 31(3), 368–390.
- Lee, I. (2005). Error correction in the L2 writing classroom: What do students think? *TESL Canada Journal*, 22(2), 1–14.
- Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal*, 63(1), 13–22.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413–468). San Diego.
- Mair, P., Schoenbrodt, F., & Wilcox, R. R. (2015). WRS2: Wilcox robust estimation and testing (Version 0.4-0).
- Mair, P., Wilcox, R. R., & Schoenbrodt, F. (2015). Package “WRS2.” Retrieved from <https://cran.r-project.org/web/packages/WRS2/WRS2.pdf>

- Mawlawi Diab, N. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing*, 24, 16–34.
- McGrath, D. (2015). Research into corrective written feedback: An EFL teacher's perspective. *OnCUE Journal*, 8(2), 106–116.
- Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, 16(2), 82–99.
- Pienemann, M. (2003). Language processing capacity. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 679–714). Oxford: Blackwell Publishing Ltd.
- Pienemann, M., & Keßler, J.-U. (Eds.). (2011). *Studying Processability Theory: An Introductory Textbook*. Amsterdam ; Philadelphia: John Benjamins Publishing Company.
- Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, 21(4), 375–389.
- Polio, C., Fleck, C., & Leder, N. (1998). “If I only had more time:” ESL learners’ changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7(1), 43–68.
- Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, 10(4), 277–303.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>
- Richards, J. C., & Schmidt, R. W. (Eds.). (2007). *Longman dictionary of language teaching and applied linguistics* (3. ed., [Nachdr.]). London: Longman.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20(1), 83–95.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Sachs, R., & Polio, C. (2007). Learners’ uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29(1), 67–100.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3–32). Cambridge: Cambridge University Press.
- Schwartz, B. D. (1986). The epistemological status of second language acquisition. *Second Language Research*, 2(2), 120–159.

- Semke, H. D. (1984). Effects of the Red Pen. *Foreign Language Annals*, 17(3), 195–202.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255–283.
- Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37(4), 556–569.
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal*, 23(1), 103–110.
- Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22(3), 286–306.
- Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, 10(2), 29–46.
- Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, 32(Special Issue 02), 303–334.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369.
- Truscott, J. (1999). The case for “The Case Against Grammar Correction in L2 Writing Classes”: A response to Ferris. *Journal of Second Language Writing*, 8(2), 111–122.
- Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing*, 13(4), 337–343.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272.
- Truscott, J. (2009). Arguments and appearances: A response to Chandler. *Journal of Second Language Writing*, 18(1), 59–60.
- Truscott, J. (2010). Some thoughts on Anthony Bruton's critique of the correction debate. *System*, 38(2), 329–335.
- Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17(4), 292–305.
- Turner, J. L. (2014). *Using statistics in small-scale language education research: focus on non-parametric data*. New York: Routledge.

- Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. *ITL - International Journal of Applied Linguistics*, (156), 279–296.
- Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62(1), 1–41.
- Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. (M. Cole, V. John-Steiner, E. Souberman, & S. Scribner, Eds.). Cambridge, Massachusetts: Harvard University Press.
- Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53(3), 300–314.
- Wilcox, R. R. (2010). *Fundamentals of modern statistical methods: substantially improving power and accuracy* (2nd ed). New York, NY: Springer.
- Wilcox, R. R. (2012). *Modern statistics for the social and behavioral sciences: a practical introduction*. Boca Raton: Taylor & Francis.
- Xu, C. (2009). Overgeneralization from a narrow focus: A response to Ellis et al. (2008) and Bitchener (2008). *Journal of Second Language Writing*, 18(4), 270–275.
- Zedeck, S. (Ed.). (2014). *APA dictionary of statistics and research methods* (First edition). Washington, DC: American Psychological Association.

Appendices

Appendix A

Correction Codes for the Indirect Feedback Group

Symbols	Examples	Corrections
o = omit	^o I work <u>for</u> to make money.	<i>I work to make money.</i>
sp = spelling	^{sp} for two <u>mounths</u>	<i>months</i>
ww = wrong word	^{ww} I spent a lot of money <u>for</u> my car ^{ww} Please <u>teach</u> me your email address.	<i>on</i> <i>tell</i>
sv = subject verb agreement	^{sv} She <u>watch</u> television every night.	<i>She watches ...</i>
vt = verb tense	^{vt} I <u>work</u> after school yesterday.	<i>I worked ...</i>
wf = word form	^{wf} I am <u>interesting</u> in English. ^{wf} My hometown is a very <u>safety</u> place.	<i>I am interested in ...</i> <i>... a very safe place.</i>
Λ = missing Word or words	I love listening ^Λ music. I have liked it since ^Λ twelve years old.	<i>... listening to music</i> <i>... since I was twelve years old</i>
# = change singular or plural	I like reading [#] <u>book</u> . I'd like to be a [#] <u>teachers</u> .	<i>... reading books.</i> <i>... a teacher.</i>
△ = article mistake	I work in [△] <u>a</u> store in Osaka. I visited [△] <u>an</u> Eiffel Tower.	<i>... in a store ...</i> <i>... the Eiffel Tower.</i>
c = Capital Letter Mistake	I didn't go to school on ^c <u>monday</u> . My favourite film is ^c <u>HARRY POTTER</u> .	<i>Monday</i> <i>Harry Potter</i>
p = punctuation	^p <u>Also baseball</u> is an interesting sport.	<i>Also, baseball...</i>
wo = word order	^{wo} I caught a <u>fish very big</u> .	<i>... a very big fish.</i>
frg=fragment (not a complete sentence)	^{frg} I love playing sport. <u>For example, football and tennis.</u> Having brothers and sisters is ^{frg} helpful. <u>Because they can give me advice.</u>	<i>I love playing sport, for example football and tennis.</i> <i>Having brothers and sisters is helpful because they can give me advice.</i>

Appendix B

Participant information form

Please answer the questions below. Your information will be kept strictly confidential, and this paper will be disposed of carefully upon completion of the research.

1. Student number: _____
2. Age: _____
3. Most recent TOEFL score: _____
4. Date of most recent TOEFL (month/year): _____
5. How many years have you studied English? _____
6. Have you ever lived in a foreign country or spent a long period of time in another country? If so, which country and for how long?

7. How much exposure do you get to English in an average week (in hours)? _____
What kinds of exposure? (listening / reading / speaking / writing) _____

8. Current zemi instructor: _____
9. Which of these sentences best describes how you feel about writing? Please tick (✓) one.

<input type="checkbox"/>	I enjoy writing in English, and I think it will be useful in my future.
<input type="checkbox"/>	I enjoy writing in English, but I don't think it will be useful in my future.
<input type="checkbox"/>	I do not enjoy writing in English, but I think it will be useful in my future.
<input type="checkbox"/>	I do not enjoy writing in English, and I do not think it will be useful in my future.

Appendix C

Research Completion Questionnaire

Student Number _____

Please rate the statements according to how strongly you agree with them. Your information will be kept strictly confidential, and this paper will be disposed of carefully upon completion of the research.

1. The feedback I received in the study was useful for improving my English errors.

1	2	3	4
I disagree	I disagree a little	I agree a little	I agree

2. I received too little feedback.

1	2	3	4
I disagree	I disagree a little	I agree a little	I agree

3. I received too much feedback.

1	2	3	4
I disagree	I disagree a little	I agree a little	I agree

4. The feedback I received was demotivating (やる気を失う).

1	2	3	4
I disagree	I disagree a little	I agree a little	I agree

Please explain your answer to #4:

5. I think the best way writing teachers could help students reduce errors is...

Appendix D

Example of Indirect WCF

"Essay 3, Draft 1"

There are many features of a good friend. The first feature is being funny and interesting. It is easy to understand why this is a feature of a good friend^P because you would not want to be with a person that has no humor and has no interest in you. Also, being with a person that is funny would also make ^{ww}yourself happy and comfortable. The second feature is being ^{wf}respective. There is a saying in Japan, "Be ^{wf}respective, even to the closest friend". This shows that even close friends need to respect each other in order to maintain a good relationship.

The third feature is being competitive and ^{wf}stimulate each other. In order to improve on your studies or whatever activity you do, it is important to have a friend to compete with and^P stimulate each other. This feature might lead to conflict at times^P but experiencing the process of solving a conflict will be a great experience for later in life.

The fourth feature is being helpful and serious at times. A good friend should be a person that helps and supports you when you are in a tough situation.^P So I think a friend that doesn't care when you are in trouble is not a true friend.

In conclusion^P there are definitely more features of a good friend^P but the ones that I listed above are ^{ww}part of the features of a good friend.

Appendix E

Example of Direct WCF

"Essay 3, Draft 1"

There are many features of a good friend. The first feature is being funny and interesting. It is easy to understand why this is a feature of a good friend, because you would not want to be with a person that has no humor and has no interest in you. Also, being with a person that is funny would also make ^{you} ~~yourself~~ happy and comfortable. The second feature is being ^{respectful} ~~respective~~. There is a saying in Japan, "Be ^{respectful} ~~respective~~, even to the closest friend". This shows that even close friends need to respect each other in order to maintain a good relationship.

The third feature is being competitive and ^{stimulating} ~~stimulate~~ each other. In order to improve on your studies or whatever activity you do, it is important to have a friend to compete with and ^{to} ~~/~~ stimulate each other. This feature might lead to conflict at times, but experiencing the process of solving a conflict will be a great experience for later in life.

The fourth feature is being helpful and serious at times. A good friend should be a person that helps and supports you when you are in a tough situation, ^{so} ~~So~~ I think a friend that doesn't care when you are in trouble is not a true friend.

In conclusion, there are definitely more features of a good friend, but the ones that I listed above are ^{some} ~~part~~ of the features of a good friend.

Appendix F

Descriptive Statistics by Essay and Group

Error type by group and essay		<i>n</i>	Minimum	Maximum	Mean	Std. Deviation
Control	Pretest all errors	10	0.51	2.25	1.25	0.53
	Essay 2 all errors	10	0.65	2.98	1.45	0.76
	Posttest all errors	10	0.57	2.87	1.30	0.79
	Delayed posttest all errors	10	0.70	2.29	1.28	0.49
	Pretest orthographic errors	10	0.00	0.26	0.07	0.08
	Pretest non-orthographic errors	10	0.51	1.98	1.17	0.47
	Essay 2 orthographic errors	10	0.00	0.34	0.11	0.12
	Essay 2 non-orthographic errors	10	0.65	2.64	1.33	0.64
	Posttest orthographic errors	10	0.00	0.17	0.03	0.05
	Posttest non-orthographic errors	10	0.57	2.82	1.28	0.74
	Delayed posttest orthographic	10	0.00	0.30	0.07	0.08
	Del. posttest non-orthographic	10	0.61	2.24	1.20	0.48
Indirect	Pretest all errors	11	0.30	1.86	1.07	0.49
	Essay 2 all errors	11	0.43	2.52	1.02	0.58
	Posttest all errors	11	0.33	2.98	1.11	0.72
	Delayed posttest all errors	11	0.24	2.86	1.03	0.69
	Pretest orthographic errors	11	0.00	0.09	0.03	0.03
	Pretest non-orthographic errors	11	0.25	1.86	1.04	0.50
	Essay 2 orthographic errors	11	0.00	0.16	0.03	0.04
	Essay 2 non-orthographic errors	11	0.43	2.47	0.98	0.57
	Posttest orthographic errors	11	0.00	0.08	0.02	0.03
	Posttest non-orthographic errors	11	0.33	2.98	1.09	0.73
	Delayed posttest orthographic	11	0.00	0.20	0.05	0.05
	Del. posttest non-orthographic	11	0.24	2.66	0.97	0.65
Direct	Pretest all errors	9	0.39	2.23	1.15	0.64
	Essay 2 all errors	9	0.42	2.33	1.16	0.68
	Posttest all errors	9	0.29	2.14	1.08	0.61
	Delayed posttest all errors	9	0.70	2.36	1.39	0.56
	Pretest orthographic errors	9	0.00	0.09	0.04	0.03
	Pretest non-orthographic errors	9	0.32	2.18	1.10	0.64
	Essay 2 orthographic errors	9	0.00	0.23	0.06	0.07
	Essay 2 non-orthographic errors	9	0.38	2.23	1.10	0.68
	Posttest orthographic errors	9	0.00	0.09	0.05	0.04
	Posttest non-orthographic errors	9	0.21	2.05	1.02	0.59
	Delayed posttest orthographic	9	0.04	0.31	0.12	0.10
	Del. posttest non-orthographic	9	0.61	2.08	1.26	0.51

Appendix G

Shapiro-Wilk Tests of Normality for Mean Error Ratio by Group and Essay

	Group	Statistic	df	<i>p</i>
Pretest	Control	0.97	10	.90
	Indirect	0.97	11	.93
	Direct	0.90	9	.24
Essay2	Control	0.80	10	.02
	Indirect	0.84	11	.03
	Direct	0.84	9	.06
Posttest	Control	0.72	10	.002
	Indirect	0.82	11	.02
	Direct	0.90	9	.27
Delayed posttest	Control	0.93	10	.42
	Indirect	0.82	11	.02
	Direct	0.95	9	.70

Appendix H

Boxplots of Error Ratios for Each Essay Split by Group, Including Outliers

