

To correct or not to correct: error correction in L2 writing instruction

Christopher Alton Baldwin

Aston University
School of Languages and Social Sciences

January 2008

Submitted in partial completion of the MSc degree in Teaching English to Speakers of
Other Languages

To correct or not to correct, that is the question – whether 'tis nobler in the classroom to suffer the tenses and syntax of outrageous grammar or to take red pens against a sea of errors and by correcting end them (to misquote Shakespeare, 1602).

Abstract

This study sets out to examine the question of the efficacy of written error correction in the writing component of a general English course. Two research questions are posed: is grammar correction an effective way to improve grammatical competence in L2 writing? and are some types of error more amenable to correction than others? Participatory action research was carried out with two classes of Italian primary school teachers learning English in order to qualify as English teachers. Each class was divided into a correction group and a non-correction group and seven essays over a two month period were written by the students. Statistical analyses were carried out to answer the research questions. The major findings are: a. written grammar error correction does not lead to any statistically significant improvement over time. b. lexical errors were found to be correctable, as were simple grammatical errors, whereas complex morphosyntactic errors showed a deterioration with correction and an improvement without correction. A questionnaire was administered in order to investigate student attitudes to the research. Findings are that students prefer correction but may feel freer to experiment without correction. Pedagogical implications and weaknesses of the present study are also discussed.

Contents

1. Literature review	1
1.1 The debate for and against correction	1
1.2 The SLA perspective	5
1.3 Re-analysis of the studies on error correction	8
2. Experiment	12
2.1 Participants and method	12
2.2 Correction method	14
2.3 Questionnaire	15
3. Results analysis	16
3.1 Total errors	16
3.2 Grammatical breakdown of errors	18
3.3 Questionnaire results	21
4. Discussion	23
4.1 Statistical results	23
4.2 Questionnaire results	27
4.3 Pedagogical Implications	29
4.4 Strengths and weaknesses of the current study and ways forward for future research	36
5. Conclusions	43
6. References	45
7. Appendices	52
Appendix 1	52
Essay questions	
Appendix 2	53
Error breakdown	
Appendix 3	54
Example of student writing	
Appendix 4a	55
Questionnaire for correction groups	
Appendix 4b	56
Questionnaire for non-correction groups	
Appendix 5	57
Full results by student	
Appendix 6	58
Full results of questionnaires	

1. Literature review

1.1 The debate for and against correction

This study was stimulated by the debate between Truscott (1996, 1999) and Ferris (1999, 2004) on the efficacy of grammar correction in L2 writing. Truscott's thesis is that “grammar correction has no place in writing courses and should be abandoned” (1996:328). His argument against grammar correction is based on four main points:

- “(a) Research evidence shows that grammar correction is ineffective;
 - (b) this lack of effectiveness is exactly what should be expected, given the nature of the correction process and the nature of language learning;
 - (c) grammar correction has significant harmful effects; and
 - (d) the various arguments offered for continuing it all lack merit.”
- (Truscott 1996:328)

On the first point Truscott deems it necessary to consider only studies which are both longitudinal, in order to observe long-term effect on SLA, and with a control group which received no correction, to determine if any observed improvements would have been caused by factors other than correction. There are many studies which look at the effects of various correction methods but which do not have a non-correction group (for example Chandler 2003, Ferris et al. 2000 (cited in Ferris 2004), Lalande 1982, Lizotte 2001, Robb et al. 1986,). These studies show the relative merits of one method over another, and generally show an improvement, but they cannot say if no correction would have been equal or even better in the various measures of accuracy and fluency looked at.

Some studies discussed in the light of these criteria by Truscott (1996) are Kepner (1991), Lalande (1982), Robb et al. (1986), Semke (1984), and Sheppard (1992). These studies all show correction to be ineffective, although Ferris (2004) interprets Kepner's study to show

evidence in favour of correction and Semke's to be inconclusive (see page 10 for a discussion of these studies).

Truscott's second point refers to how correction could relate to the SLA process. He takes the position that SLA is not a simple transfer of knowledge from teacher to student, but is rather a gradual, poorly understood process. The first theoretical point mentioned is the hypothesised natural order of language acquisition. If a student is not cognitively ready to acquire a form, no amount of correction can give it to him or her. Syntactic, morphological and lexical knowledge are possibly acquired in different ways, so applying the same corrective techniques to different problems is unlikely to be effective. Corroborating this, Hu (2002) finds that there are major limits to the use of metalinguistic knowledge in L2 writing and error correction. Truscott then talks about “pseudolearning” (Truscott 1996:345) or the acquiring of metalinguistic knowledge, which has little impact on the underlying interlanguage systems (see also Krashen, 1987). It is important to remember that we are discussing grammar correction only at this stage, not lexical or punctuation errors for example. These are possibly acquired and governed by simpler systems, so are possibly more amenable to correction (Truscott 2001).

Truscott's third point is the harmful effects of correction, in that students simplify their texts in order to avoid known problems, and class and teacher time is spent struggling over errors which takes valuable time from more productive activities such as producing new texts (Truscott, 1996). See Kubota (2001) for an example of this. It is also noted that correction is intrinsically negative, so can produce stress which is counter-productive to learning.

Truscott's (1996) last main point is that the reasons given for correction are not valid. The two principal arguments in favour of correction are a fear of fossilisation and student survey research which shows that students want correction. The response is that there is no evidence that correction actually helps in the case of fossilisation and that students may want correction, but that does not mean that it is good for them.

Ferris (1999) wrote a strong response against Truscott's position, to which Truscott (1999) responded. Both sides of the reasoning are now presented.

Ferris (1999:1) states that Truscott's claim that correction should be abandoned is “premature and overly strong.” Her argument is based on two points: problems with definition and problems with support. She notes that the term “error correction” is not defined and that “grammar correction” is defined very loosely (Ferris 1999:3). Truscott's (1999:112) response is that he did not use the term “error correction” so it did not need to be defined, and that he objects to all forms of grammar correction, not just that which is considered “poorly done” (Ferris 1999:4), thus the criticisms of definition seem weakly founded.

Ferris' (1999:4-5) second point is more interesting. She notes three problems with the studies that Truscott cites:

“(a) The subjects in the various studies are not comparable; (b) The research paradigms and teaching strategies vary widely across the studies; and (c) Truscott overstates negative evidence while disregarding research results that contradict his thesis.” (Ferris 1999:4)

Truscott (1999:113-115) replies to the first two points that variability actually leads to generalisability, so this is in fact an argument against correction. Ferris (2004:52) responds that these studies which differ in design and scope also lead to different results, so there is no way to generalise, although she admits that even the anonymous reviewer of her article did not accept her point. The problem with her argument is linked to the reply to the third point above, which is that Ferris rejects Truscott's reasoning regarding several studies. She notes that Kepner (1991) was a study of journal writing, not multiple draft essays, so cannot be considered. Truscott's response is that this kind of journal writing plays a large part in many writing courses, so deserved attention. Later on she cites Kepner's study as supporting correction (Ferris 2004:51). Ferris also rejects Truscott's reasons for not taking into consideration studies by Fathman and Whalley (1990) and Lalande (1982), “both of which found positive effects for error correction” (Ferris 1999:5). However, Fathman and Whalley (1990) is a short-term study, so cannot show effects on long-term SLA, and Lalande (1982) did not have a control group, so it is impossible to say if a non-correction

group would have performed better or worse in the same situation than the experimental groups. To link back to the second point above, if the non-control group and short-term studies are rejected, then there is little variation in the results of the remaining studies – Truscott argues that they all show correction to be ineffective, even across varying situations.

Ferris then goes on to present three reasons for continuing error correction: a. as students want correction, withholding it could be demotivating and reduce student confidence; b. university subject teachers are not tolerant of typical ESL errors (see also Johns 2004:83 for a contradictory anecdote); and c. students need to become more self-sufficient as writers, so need to improve their editing skills (Ferris 1999:8). Truscott's (1999:116-117) rejoinder was: a. teachers should train students that correction is unhelpful; b. this argument presupposes that correction actually helps students become better writers, but Truscott argues that this claim is unjustified; and c. this argument confuses grammar correction with strategy training – but these are two separate aspects.

Ferris (2004:54-56) seems to drop her second and third points, but focuses on the danger of fossilisation if correction is not given and reasons that correction can be a helpful “step on the road to long-term accuracy” (Ferris 2004:54). She also claims that short-term studies are useful in predicting long-term effects, although the evidence shows this not to be the case, as the short-term improvements noted do not seem to last over time (see studies cited in table 1, page 9).

Other researchers have added their voices to the debate, for example Chandler (2003 and 2004). She accused Truscott of ignoring evidence against his thesis, for the reasons mentioned above, but she then comes to the same conclusion as Truscott from the same study (Fathman and Whalley, 1990), as noted by Truscott (2004:341). Truscott (2004) responds to Chandler (2003), noting many flaws in Chandler's experiment, for example considering all errors, not just grammatical errors; the possibility that the reduction in errors observed was caused by simplification and avoidance of known problem areas; and the fundamental problem – the lack of any real control group. Chandler (2004) replies to Truscott, but does not make any new points.

Hyland and Hyland (2006) discuss Truscott's position, citing longitudinal studies in favour of correction, but which again have no control groups. They make the point that

“it is unlikely that feedback alone is responsible for long-term language improvement, it is almost certainly a highly significant factor.” Hyland and Hyland (2006:4)

Although without control groups this remains conjecture.

Guénette (2007) presents a stimulating analysis of many of the studies in the area of written error correction, and draws the conclusion that:

“findings can be attributed to the research design and methodology, as well as to the presence of external variables that were beyond the control and vigilance of the researchers.”

This implies that after decades of research on the question we are still at “square one”, and that there is a need for more rigorous research designs in order to satisfactorily answer the question of grammar correction in L2 writing (Ferris 2004:49).

1.2 The SLA perspective

An analysis of the SLA literature can help to see why correction may or may not work. Essentially there are two paradigms to consider, leading to opposing conclusions – behaviourist (Skinner, 1957) and naturalist (Chomsky, 1959). The behaviourist perspective states that errors need to be corrected immediately in order to avoid fossilisation (Myles, 2002), whereas the naturalist approach sees errors as part of the process of SLA, which correction can do little to change.

1.2.1 Why correction should work – The behaviourist paradigm

When students write in L2 they tap into deeply seated interlanguage (Selinker, 1972), which lies under the consciousness. This interlanguage is something approaching correct L2, but with its own unique grammar and vocabulary. The object of correction is to try to make the interlanguage more like standard L2. When students are corrected they are pushed to “notice the gap” between their production and the correct form (Schmidt and Frota, 1986; see also Krashen, 1983). This noticing will then be fed back into interlanguage, thus improving student accuracy. This can then be linked to the output hypothesis (Swain, 1985; see also Qi and Lapkin, 2001) which states that as students produce output they become aware (consciously or not) of problems in their interlanguage which they modify in successive outputs, which leads to acquisition. When students re-write their corrections or alternatively speak about them in teacher conferences (Bitchener et al., 2005) they produce the form correctly which pushes towards acquisition, or destabilisation of fossilised features. This output then becomes input as students read their work, thus strengthening the correction (Pulido, 2007; Sharwood Smith, 1996). The output hypothesis has, though, been attacked on the grounds that high levels of competence can be attained even without any or much output, and on a lack of experimental evidence (Krashen, 1998). Work on fossilisation (in spoken language) finds correction on its own to be of little help (Han 2003; Selinker, 1972; Selinker & Han, 2001). However, correction could be more promising on newly learnt features, although little research has been done on this question yet.

It is also hypothesised that attention is necessary to varying degrees for learning to take place (Izumi, 2002:542-543) so correction can raise student consciousness of the problem, thus helping correct acquisition (Fotos, 1993). This consciousness raising can be seen as the first step in noticing. Both consciousness raising of general problem forms and noticing of specific occurrences of correct forms (which were produced erroneously) in context may be helpful in pushing towards acquisition, and this could be brought about by correction.

Skill acquisition theory (DeKeyser, 2007) talks about three stages of knowledge – declarative, procedural and automatic, saying that it is necessary to pass through these stages in order to acquire any new skill. This has been applied to SLA for lower level students with simple structures in formal learning environments. Correction can give students the declarative knowledge if carried out competently. Re-writing corrections

could help to give the procedural knowledge, and repetition in varying contexts (Folse, 2006) could possibly pass this into automatic knowledge in interlanguage.

1.2.2 Why correction should not work – The naturalist paradigm

Krashen's (1981) dualsystem hypothesis draws a distinction between acquisition and learning, saying that what is consciously learnt has little bearing on SLA. Therefore correction only changes conscious metalinguistic knowledge, but has little impact on interlanguage. This theory is in accord with principles of universal grammar (Chomsky, 1965).

This idea is strengthened by work on sequences in SLA. Many studies (for example Ellis 1994; Larsen-Freeman & Long 1991; Bardovi-Harlig 1997, and Bardovi-Harlig & Reynolds 1995. See Lightbrown (2000) for a review) show that grammatical features are learnt in a predetermined order, thus any effort to correct an error when a student is not ready to acquire that particular point is doomed to failure (see Truscott 1996, 2001). Correction may change metalinguistic knowledge, which is why students can improve over multiple drafts of a single essay (see table 1), or perform well in grammar tests (Cardelle and Corno, 1981), but does not influence interlanguage.

1.2.3 Bridging the gap – which errors could be correctable and which not?

Linguistic competence can be divided into semantic, morphological, and syntactic competences – knowing words, knowing how to change them for tense, inflection and so on, and knowing how to put them together in sentences, to convey meaning. It is unlikely that these competences are acquired by the same processes, so an understanding of the mechanisms which govern them can help in understanding which forms may be more susceptible to correction. Ferris (1996:6) draws a distinction between “treatable” and “untreatable” errors, those deemed treatable “occur in a patterned, rule-governed way.” Truscott (2001) expands this idea, introducing the concepts of simplicity and discreteness, where items which are easily explicable and easily repeatable are more easily correctable. Morphology comes from complex systems, so morphological errors are seen as poor candidates for error correction, whereas individual words are discrete items, so the use of a

wrong word should be more easily correctable. Syntactic systems are seen as being highly complex and not discrete, so correction of syntactic errors is unlikely to be effective. In short many grammatical errors are probably not amenable to correction, whereas lexical errors are more so.

1.3 Re-analysis of the studies on error correction

The following section analyses some of the studies which have been considered by both Ferris, Truscott and other authors as well as some more recent research on the question. This section includes the phrase “error correction” even though the first question being studied in this paper is “grammar correction” as many of the studies did not draw a distinction.

One fundamental aspect of the studies is that they must be longitudinal, in order to observe long-term effects on interlanguage not just short-term improvements, and have control groups which received no error correction, to discount other variables leading to changes in accuracy such as grammar instruction or in second language contexts the improvement expected from living in a country where the target language is spoken. Groups receiving feedback on content only are acceptable as control groups since content feedback is not grammar correction. Table 1 presents 17 studies which have looked at this question, analysing them in the light of the longitudinal/control group question. Studies which lack one of these aspects are considered irrelevant, or in the words of an anonymous journal reviewer they are evidence which “tells us nothing” (Ferris 2004:54) in terms of the long term effects of correction, although they may be interesting for other reasons. These studies take place in a wide variety of settings – both FL and SL, with different languages as L2. It will be noted that using these criteria only 6 of the studies are considered relevant, so worthy of further investigation on the question of the long term effect of error correction on SLA (in bold in the table). They will be examined in chronological order.

Table 1 – Summary of previous studies

Study	Longitudinal / short-term	Non correction	Major findings	Relevant
Ashwell 2000	Short	Yes	Correction better than non-correction on 3 draft essay	NO
Bitchener et al. 2005	Long	Yes (content feedback)	Improvement with written and conference feedback for 2 error types. No improvement when all 3 error types looked at together. No difference between written feedback only and non correction groups	YES
Chandler 2003	Long	No	Direct correction & underlining better than code	NO
Fathman & Whalley 1990	Short	Yes	Correction helped	NO
Fazio 2001	Long	Yes	No difference between groups	YES
Ferris & Roberts 2001	Short	Yes	Codes and underlining effective; non-correction not effective on two draft essay	NO
Ferris et al. 2000 and Ferris and Helt 2000 (cited in Ferris 2004, two reports of the same study)	Long	No	Indirect better than direct correction	NO
Greenslade and Félix-Brasdefer 2006	Short	No	Coded error correction better than underlining errors in two drafts of one essay	NO
Kepner 1991	Long	Yes	No difference between groups	YES
Lalande 1982	Long	No	Indirect better than direct correction	NO
Lee 1997 (Experiment on correction, not writing)	Short	Yes	Feedback better than correction. Correction codes need to be handled with care. Not all errors are the same	NO
Lizotte 2001	Long	No	Students improved with self correction	NO
Polio et al. 1998	Long	Yes	Both groups improved; no significant difference between groups	YES
Robb et al. 1986	Long	No	All groups improved	NO
Sachs and Polio 2007	Short	No	Reformulation better than correction	NO
Semke 1984	Long	Yes	No differences between groups on accuracy; non correction leads to better fluency	YES
Sheppard 1992	Long	Yes (content feedback)	Looked at verb forms: no significant differences of writing accuracy between groups. Non-correction group improved in grammatical accuracy	YES

Semke (1984) studied German FL students at a US university doing free-writes, and found that there were no statistically significant differences between the correction and non-correction groups. She went on to conclude that correction can be harmful for fluency and produces a negative effect on student attitudes.

Kepner (1991) is a study worth reading carefully as Truscott (1996:6) interprets these results as showing the ineffectiveness of correction, whilst Ferris (2004:3) says that Kepner “finds positive evidence for error correction but curiously interprets it as negative.” Kepner looked at four classes of college students of Spanish as a foreign language in the USA, giving half the students only grammar correction and the other half only content feedback. She also used high and low level L1 verbal ability as an independent variable. The study found that both the high and low verbal ability groups with content feedback produced significantly better texts in ideational terms on a count of high level propositions than the grammar correction groups, thus showing harmful effects on quality for grammar feedback. This was not offset by an improvement in surface level errors for either verbal ability group, as there was no statistically significant difference between any of the groups. Ferris' comments above may come from Kepner's (1991:310) observation that “the error-corrections model is only helpful in that it permits low-verbal-ability students to perform at the same level as high-verbal-ability students on measures of accuracy.” The same lack of any statistically significant difference between high and low level verbal ability groups was also found in the content feedback group, so again this improvement of the low ability groups cannot be accredited to grammar correction. The conclusions which can be drawn from this is that L1 verbal ability level has no effect on surface level errors, and again that grammar correction is ineffective.

Sheppard (1992) looked at 26 US university ESL students, giving one group “discrete-item attention to form” and another “holistic feedback on meaning” (Sheppard 1992:103). He found that the non-grammar-correction group improved in grammar more than the form focus group, thus negative effects for correction on grammatical ability were seen.

Polio et al. (1998) studied 65 undergraduate ESL students, split into one correction and one non-correction group. The students wrote journal entries over a 15 week period, and no

statistically significant differences were found between the groups, with both making gains in accuracy.

Fazio (2001) considered 112 primary school students in French speaking Canada, looking at the learning of French both as L1 and L2. Of the students, 66 were native French speakers and 46 were from diverse linguistic backgrounds. The students were assigned to three groups: form feedback, content feedback (i.e. no grammar correction) and a mixture of the two. There were found to be no significant differences between any of the treatment groups for either L1 or L2 students.

Bitchener et al. (2005) studied 53 post-intermediate ESL migrant learners in New Zealand. The students were split into three groups – written error correction with a 5 minute teacher conference, written correction only, and feedback on content, but not form (it was felt unethical to give no feedback at all). Errors of prepositions, the past simple tense, and the definite article were analysed, and it was found that when all three error groups were analysed together there were no significant differences between the groups. However, when the error categories were analysed individually it was found that the correction with conference group improved on the past simple and definite article but not with prepositions. This fits in with Ferris' (1999:6) notion of 'treatable' and 'untreatable' errors, and Truscott's (2001) idea that simple, easily explicable errors are more amenable to correction. The fact that the correction only group did not show any improvement whereas with a conference there was an amelioration could be explained by the extra cognitive effort put in talking to the teacher, while the correction only group were not required to do anything with the corrections, thus no cognitive processing was required so there was no observed change in interlanguage.

This study raises important points as to what to look for in future studies – not just overall changes in grammatical accuracy, but specific grammatical forms.

In conclusion, of all the studies which address the grammar correction question from a longitudinal and control group perspective the evidence is clear – grammar correction has little effect, except possibly with certain simple, rule bound errors. It should be noted that some of these studies (Fazio, 2001; Kepner, 1991; Polio et al., 1998; and Semke, 1984)

have been criticised for looking at journal writing, which would not usually be corrected (Ferris 1999:5). Journal writing is, though, an important part of many courses so deserves to be studied, and the SLA processes are largely the same, even if affective factors may differ, whether one is writing a journal entry or an essay. Another criticism is that in some of these studies (Polio et al., 1998; and Semke, 1984) the non-corrected students wrote more than the corrected students (Chandler 2003:268-9). This is precisely the point – that time is better spent writing new texts than correcting errors, as both groups had the same amount of instruction, but used in different ways.

2. Experiment

2.1 Participants and method

The first research question addressed in this paper is: is grammar correction an effective way to improve grammatical competence in L2 writing?

Many researchers (Bitchener et al., 2005; Chandler, 2003; Ferris 2004; Guénette, 2007) on this question have raised the ethical issue of the potential harm caused to students by not correcting. It could equally be argued that as there is evidence for the harm of correction in terms of content and grammatical ability (for example Kepner, 1991; Semke, 1984; and Sheppard, 1992) that it is unethical to correct errors, even though it is standard practice in most L2 writing classrooms. This study was carried out with two classes of experienced Italian primary school teachers learning English in order to be able to teach it, thus these pedagogically aware students with experience as L1 instructors (two students were practising English teachers following the course as a refresher and one was an L2 teacher of German) provide an excellent student sample for conducting participatory action research (Taylor, 1994), which helps to address the ethical issue as the students were aware of the nature of the research throughout and were given the choice of which group they wanted to be in (correction or non-correction).

The classes were elementary (n=11) and pre-intermediate (n=17) levels (A1-A2 and A2-B1 on the common European framework), following a general English course, of which

writing played a part, both to implement recently taught grammar (in accordance with the output hypothesis (see Swain, 1995)), and as practice for the final exam. The lessons were 3 hours long, once a week, using New English File Elementary and Pre-intermediate course books (Oxenden et al., 2004, 2005).

Each class was divided into two groups – correction and non-correction. The division was made by first asking students who had a preference to choose groups. One student requested non-correction and four asked to receive correction. The remaining students were divided in order to have approximately equal mean error rates per 100 words in the first essay of the study. There are therefore four conditions in this study:

Table 2 – Groups

	Elementary	Pre-intermediate
Non-correction	NE	NI
Correction	CE	CI

The study took place over a period of two months, with seven essays in total of a required length of approximately 100 words (average on first essay 100.79 words). Both groups were given the same initial writing task during class time, with different tasks for the two levels. The essays were submitted to the teacher/researcher who then marked all essays and returned those of the correction group. Phase two of the cycle was for students in the correction group to read and try to understand the corrections and re-write their essays. The teacher/researcher was available to explain any difficulties. At the same time the non-correction group received a new writing task designed to elicit similar language as the first task, but without repeating the same ideas (see appendix 1 for the questions). The cycle was repeated three times and then one final essay was set as a measure of final improvement. Thus the non-correction groups wrote a total of seven different essays, whilst the correction groups wrote four different compositions and three second drafts. One essay per week was written, each essay taking approximately half an hour to complete.

This design directly addresses the question of corrected students writing less than uncorrected students (Chandler 2003:268-9) as exactly the same amount of class time was

used for both groups, the difference being what is done with that time, either understanding and re-writing corrections, or writing new texts. This experimental design is possible using participatory action research.

2.2 Correction method

There are two basic types of error correction: direct, that is re-writing the problem word or sentence, and indirect methods, such as underlining the error, and using correction codes. The short-term studies presented in table 1 (page 9) offer conflicting results as to the value of these methods. Chandler (2003) found both direct correction and underlining to be better than a correction code, Ferris & Roberts (2001) concluded that codes and underlining are equally effective, whereas Greenslade and Félix-Brasdefer (2006) noted that coded correction is more effective than underlining. Chandler's results could be explained by the fact that her correction code was very complex, so possibly not understood by her students.

It was thus decided to use a simple correction code, following Sugita's (2006:35) advice "clarity is the first thing to bear in mind in writing a comment." The following code was adopted:

Table 3 – Error codes used

Capitalisation	Cap
General grammar errors	Gr
Omission	Om
Spelling	Sp
Word errors	W
Word order	Ord

This code is not intended as an in-depth analysis of all possible error types, but rather as a simple way to guide the student to the error. Omission and order errors are usually grammatical in nature, order errors being classified by Ferris (1999) as untreatable (see page 18). The 'general grammar errors' category encompasses a very wide range of errors

from tense problems to the plural and possessive 'S', generally coming under Ferris' treatable category (points taken up in the analysis, below). It should be noted at this point that whilst all errors were corrected only grammatical errors were taken into account in the first analysis (Truscott 2004:6). Errors in student texts were underlined and the code was written next to the error. This code was explained to the correction groups during the lesson of the first revision. Photocopies of all essays were made before correction to allow for blind second marking, to be carried out at the end of the experiment, in random order.

The word 'error' is difficult to define and many of the studies in this area simply do not define it. However, Lennon's (1991) definition “a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts,” is used in this paper. For a full list of errors see appendix 2, which shows which errors were counted as grammatical errors. Appendix 3 gives an example of student writing with error codes used.

2.3 Questionnaire

In order to gauge student reactions to the methods used and to try to understand if students felt these methods to be effective two questionnaires were designed. It has been argued (Truscott, 1996, 1999) that students want grammar correction because they are used to getting it, so their opinions are moulded by current practice. Participatory action research with this particular type of student (experienced teachers) gives extra validity to student opinions as they had the opportunity to reflect on their experience both as students and as teachers throughout the experiment.

Two questionnaires were written for the correction and non-correction groups to probe the following issues: correction groups were asked whether they like being corrected or not, if they think that correction helped their grammar, and if they understood the correction codes. The issue of avoidance of problem areas both in the short-term and long-term was also considered as it is one of Truscott's (1996:333) criticisms of correction, but which is hard to gauge using textual analysis. Non-correction groups were asked whether or not they liked not being corrected, if they felt that their written grammatical accuracy had improved, and if they felt that not being corrected encouraged them to write more freely

(one of Truscott's 1996 arguments in favour of non-correction). All students were also asked if they had any additional comments to add on the subject of correction. See appendix 4 for the questionnaires.

The questionnaires were administered at the end of the experiment in English, with the teacher/researcher available to clarify and explain any misunderstandings. In both of the classes a lively discussion followed the questionnaire answering session, and student opinions from this were noted.

3. Results analysis

3.1 Total errors

There are many ways to analyse students' texts for improvement over the course of an experiment. Many of the studies presented in table 1 (page 9) use ratios of number of errors per total number of words (for example Ashwell, 2000; Chandler, 2003; Fazio, 2001; Greenslade 2006; Kepner 1991; and Polio et al. 1998) or errors per occurrence of the form under analysis (Bitchener et al. 2005). In the present study the number of errors per 100 words of text was calculated and the first and last essays compared. All four groups (NE, NI, CE and CI) were found to be statistically uniform at the start of the experiment, using both the t test and the F test, so the elementary and pre-intermediate groups were analysed together. Six students were excluded from the study due to absence, leaving twenty two. The following results were obtained:

Table 4

		Group		
		Correction n=10	Non-correction n=12	t test(p <.05)
1 st essay	Mean errors/100 words	5.72	4.93	0.68
	Standard deviation	2.35	2.76	
Last essay	Mean errors/100 words	6.12	3.84	1.64
	Standard deviation	3.45	2.77	
				t(tabulated)=2.086

From these results it can be seen that the correction groups actually got slightly worse, whereas the non-correction groups improved slightly. However, the t test shows these results to be statistically insignificant. This is confirmed by an ANOVA test: $F(\text{correction group})=0.08$; $F(\text{non-correction group})=0.87$; $F(\text{tabulated})=2.69$. These analyses show that neither group made fewer grammatical errors to a statistically significant degree over the course of the experiment. Non-significant results are also obtained when the groups are analysed separately.

These results need to be seen in the light of the tasks set. The essay questions followed the grammatical content of the course, so some questions could have been more difficult than others and so produced more errors, although an attempt was made to set comparable questions for the first and last essays to elicit similar grammatical points (see Bitchener et al., 2005:202).

A random sample of 11% of the essays was taken, with examples from all of the groups, to be blind second marked by a teacher who does not know the students. The essays were second marked in random order, with the second marker not knowing which essay came from which group, nor if they were first or last essays. An interrater reliability of 87% was found (compare 76% in Chandler, 2003; “over 95%” and “almost 99%” in Ferris and Roberts, 2001:170; and 86% in Panova, 2002). Importantly, 89% reliability was found in the first essay and 86% on the last essay, thus indicating an intrarater reliability of 98%. This is important as the change between first and last essays is under consideration here so intrarater reliability is more important than interrater reliability (See Chandler 2003:276). These figures were obtained by calculating the percentage difference between the errors found by the two raters.

See appendix 5 for a breakdown of errors per student.

3.2 Grammatical breakdown of errors

At this point, this analysis moves away from grammar errors only and analyses other error-types, such as lexical errors. Several of the studies in table 1 (page 9) analyse errors in more depth, breaking them into grammatical categories of various types (Bitchener et al.,

2005; Ferris and Roberts, 2001; Lalande, 1982; and Lee, 1997). This is more revealing than analysing all errors together, as it allows an investigation into Ferris' (1999) treatable/untreatable errors and Truscott's (2001) more in-depth analysis.

Ferris defines treatable errors as errors which “occur in a patterned, rule-governed way” such as “subject-verb agreement, run-ons and comma splices, missing articles, verb form errors.” Untreatable errors “included a wide variety of lexical errors and problems with sentence structure, including missing words, unnecessary words, and word order problems”(Ferris 1999:6).

Truscott (2001) considers SLA related reasons why certain errors could be more or less amenable to correction. His thesis is that

“the most correctable errors are those that involve simple problems in relatively discrete items. Least correctable are those stemming from problems in a complex system, particularly the syntactic system. Grammar errors in general are not good targets, though certain types can be identified that are more promising than others”
(Truscott 2001:93)

A more detailed consideration of his conclusions is presented in table 5.

Table 5

Uncorrectable	Syntactic errors
	Morphological errors
	Non-simple preposition errors
	Verb tenses
Moderately correctable	Misclassification of words (e.g. countable/uncountable nouns)
	Incorrect derivational affixes
	Use of a given form with inappropriate words (e.g. who/whom)
Correctable	Spelling
	Simple errors in word meaning
	Idioms and collocations
	Certain preposition errors, when associated with particular words
	Utilisation of “the” before words which do not permit it
	Mistaking of words with similar spelling or pronunciation
	Register

(Adapted from Truscott 2001:104-105)

The main differences between Ferris' and Truscott's analysis is that Truscott is more negative with regard to grammatical errors in general, particularly morphological errors, although they agree that syntactic errors are generally not amenable to correction. Truscott, on the other hand is more positive than Ferris with regard to lexical errors, considering them to be simple discrete items which are easy to generalise, and thus acquire by correction (Truscott 2001:95). The third person singular 's' is governed by a simple rule, so should be treatable according to Ferris, but as it is a morphological feature Truscott would consider it uncorrectable. Prepositions would be classified as untreatable, as the rules which govern them are “idiosyncratic” (Bitchener et al, 2005:201); however, from Truscott's point of view, VERB+PREPOSITION collocations are often discrete items and it could be 'simple' to learn (by correction) that one preposition often goes with one verb. Other uses of prepositions, on the other hand, do fit in with Bitchener's view above, thus some of Ferris' treatable errors would be considered correctable under Truscott's model and others not. Ferris and Roberts (2001:173) moves towards Truscott's position, recognising that “some 'untreatable' errors may be more so than others — specifically, complex sentence structure problems versus single-word errors.”

In order to test these hypotheses, a second research question was formulated: are some types of error more amenable to correction than others?

The data from the present study were analysed by splitting errors into the categories in table 5. The improvement in errors per 100 words was calculated, combining the two levels (elementary and pre-intermediate). The difference in improvement between the correction and non-correction groups was calculated for each error category. This calculation shows the relative merits of correction/non-correction for each error type. Error types with few occurrences were omitted. The results are presented in table 6.

Table 6. All figures are in errors per 100 words

ERROR TYPE	Non-correction improvement	Correction improvement	Difference	Correctable / Uncorrectable
Simple errors in a word's meaning	-0.3	0.61	0.91	C
Misspelling	0.61	1	0.39	C
Preposition – mistaken association of a particular preposition with a particular other word, or failure to make such an association when necessary	-0.03	0.07	0.1	C
Non-simple preposition errors	0.03	0.1	0.07	U
Verb tenses	0.23	0.27	0.04	U
Syntactic errors	-0.2	-0.54	-0.34	U
Use of “the” before words that do not allow it	0.1	-0.47	-0.57	C
Morphology	0.66	-0.33	-1	U

A positive figure for 'difference' indicates correction to have been more effective, whereas a negative number shows that the error type actually improved more without correction than with correction. In the improvement columns for each group a positive number indicates an improvement, whereas a negative figure signals a deterioration. The correctable/uncorrectable category follows table 5, from Truscott (2001). The table is ordered according to the difference column, hence the higher an error type is on the list, the more amenable to correction it has been, and the lower, the less correctable. It should be noted that the actual number of errors in each category is small (on average 40 errors per

category), and no test for statistical significance has been carried out, but the trends observed here agree with Truscott's hypothesis in that errors in meaning and spelling were found to be correctable, whereas morphological errors improved with no correction and got worse with correction. Syntactic errors deteriorated with both correction and non-correction, but the deterioration was smaller without correction. The “the” error was considered by Truscott to be correctable, but here it has improved without correction and actually deteriorated in the correction groups.

3.3 Questionnaire results

3.3.1 Correction groups

Considering initially the correction groups in both levels, the first question asked, “Did you like being corrected?” 93% of students reported liking correction, with only one student (7%) being neutral.

The second question asked, “Do you think that it helped or damaged your grammar?” Again 93% of students thought that correction helped their grammar improve and one student was neutral.

These two sets of responses are in line with research (Enginarlar, 1993; Ferris, 1995; Hedgcock & Lefkowitz, 1994; Leki, 1991) which found that students want their errors to be corrected.

Question three inquired, “Did you find it clear or confusing?” Yet again 93% of respondents found it clear with only one being unsure. This indicates that the correction method chosen was appropriate for the students and that they understood the corrections. This is an important point as unclear correction could have led to different results, which would have confounded the results of the study.

Question four probed the issue of avoidance of difficult structures. “Did it encourage you to avoid using difficult forms or to try to get it right... (a) in re-writes? (b) in later texts?”

21% of students admitted to avoidance in re-writes of the same essay, whilst only one student admitted that this had had a long-term effect in later texts. One student in conversation said that she avoided an error, but replaced it with another difficult structure, so she did not feel that this was a harmful technique. However, these results add weight to Truscott's (1996:333) argument that students do avoid problem areas, which may have a long-term effect even if students do not recognise this.

3.3.2 Non-correction groups

The first question to the non-correction groups was: “Did you like not being corrected?” 75% of students reported disliking not being corrected, which is in harmony with the results of the correction group. However, two students did say that they liked not being corrected. These two students were both in the pre-intermediate group. Although this is a very small sample, it is possible that the higher the level, the more the students will accept the idea of not being corrected. See below for a discussion of this.

Question two asked, “Do you think that the grammar in your writing got better or worse during the course?” Only one student thought that her written grammar had deteriorated, whereas 50% said that it had improved and 42% were neutral. This shows that even though the students generally did not like non-correction many still felt that they had improved grammatically, which agrees with the statistical results.

The third question, “Did not being corrected encourage you to experiment with difficult forms or make you worried about mistakes?” again lends weight to the argument that non-correction encourages students to write more freely. 58% of students said that they were encouraged to experiment more, thus to try out new forms, which could help improve both content and grammar. Only 25% said that they were worried about mistakes as a consequence of not being corrected. See below for a discussion of the implications of these results.

These trends were generally the same across the levels, except for the point in question one for the non-correction group, as noted above.

The last question for both groups was “Do you have any comments?” Student responses were in harmony with the previous answers, in that corrected students said that they appreciated correction, and uncorrected students were generally sceptical about not being corrected. One pre-intermediate, non-corrected student who reported disliking non-correction said that she felt that knowing her errors would help her to self correct, but that she did feel encouraged to write spontaneously without correction. She asked if this is enough to improve. Another student in the same group wrote that she liked very much non-correction and felt that her writing had improved throughout the experiment.

The discussions which followed the questionnaire sessions reflected these opinions, with students being curious to know the statistical results of the study. Most of the students reported finding the experience interesting, and some said that it helped them to reflect on their own teaching.

See appendix 6 for full results.

4. Discussion

4.1 Statistical results

It can be seen from table 4 that there was no significant difference between the correction groups and the non-correction groups. This finding is in harmony with the longitudinal control group studies presented in table 1, in that correction does not seem to effect the number of errors produced. Thus a preliminary answer to the first research question “is grammar correction an effective way to improve grammatical competence in L2 writing?” is “no”. However, this needs to be balanced with the second research question “are some types of error more amenable to correction than others?” Table 6 indicates that word meaning and spelling errors are correctable in line with Truscott (2001).

Looking specifically at grammar errors, there were very modest gains for preposition errors with correction compared to non-correction, with only a small difference between “simple” and “non-simple” preposition errors. The simple errors were found to be more

correctable than non-simple, again lending weight to Truscott's (2001) simplicity hypothesis. Verb tenses made nearly the same improvement with and without correction, which could be a reflection of grammar instruction during the courses as a whole, which featured a focus on tense forms in both levels. "Verb tense" as an error category is, though, very wide ranging. It includes errors from using present simple instead of past simple, which is a simple error, to more complex present perfect/past simple confusion. Thus some verb tense errors may be more correctable than others. Truscott's pessimism with regard to the correctability of morphological errors has been borne out by this study. These findings can be compared with Bitchener et al. (2005), who found improvements with correction and teacher conference for the definite article and the past simple but no difference with preposition errors. This again bears out the idea that simple errors are more correctable.

Morphological errors improved greatly without correction, but deteriorated with correction, in line with Truscott's thesis. This could be explained by the output hypothesis (Swain 1985,1995), previously cited as evidence in favour of correction (see page 6), as the extra writing practice had by the non-correction groups led to more output. Any possible subconscious noticing occurring during the initial writing, added to the extra opportunity to exercise these noticed points in follow-up writing a week later could have led to acquisition of these morphological forms. This could have happened because the extra essays written by the non-correction groups were all using similar grammatical and lexical points to the previous essay in the same cycle, so the students had time between essays to think about the points and then apply them in different settings. This contrasts with the deterioration of morphological forms with the correction groups which could be explained if different aspects of language are acquired in different ways (see Ellis and Schmidt, 1997; Tyler et al., 2005) The conscious effort required to understand the errors may have blocked any subconscious progress made on these difficult forms. Thus the depth and method of processing (see page 25) required for one language feature are different for another.

The output hypothesis seems to apply in different ways for different error types, as morphological and syntactic errors responded best with more original output, whereas lexical errors responded better to outputting corrected errors in re-writes. Whether or not it is the different forms of output, or other cognitive processes which lead to the observed

differences is a question for further research. The type of noticing carried out in the two situations may be crucial. The subconscious noticing fostered by non-correction and extra writing could be more effective for complex errors, whilst the conscious noticing encouraged by correction seems to work better for simpler errors such as word meaning and spelling. This is in harmony with Swain and Lapkin (1995:371) who note that “sometimes, under some conditions, output facilitates second language learning, ” but not under all circumstances. Maybe 'different types of output' could be added to this caveat.

'The' as an error category was considered by Truscott to be simple, thus amenable to correction; many errors related to the definite/indefinite article are highly complex features of the syntactic system, thus less correctable. This result could also be explained by different “the+NOUN” collocations occurring in the first and last essays, thus the forms may have been learnt by correction but simply not used in the last essay.

This could be explained by the finding that vocabulary retention is fostered not by depth of processing, but by number of retrievals (Folse, 2006). The students were encouraged to write down in vocabulary note books any newly learnt words, which was done during the correction sessions. Students' reviewing of their notes along with the essay re-writes would mean repeated retrievals for the previously wrongly used word, thus explaining why correction seems to work for lexical errors. This is in accord with Skehan (1998), who argues that students process meaning over form, when attention capacities are stretched, so when students are faced with a number of errors, lexical points are more easily processed to the exclusion of morphological items, leading to acquisition of lexis with correction. Laboratory research by Robinson (2005) also implies that grammar is better learnt by deep processing, not by frequent retrievals. It would be interesting to explore this area by conducting research on correcting only morphological errors, although this could raise ethical issues, in that this could harm students' grammar acquisition (Ferris, 2004).

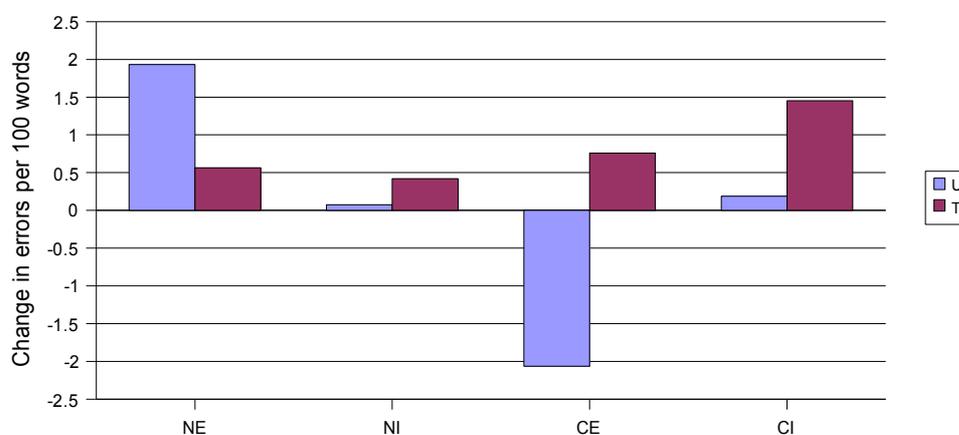
Contrarily to lexis, morphological features may be learnt more by depth of psychological processing (Izumi, 2002). This idea is supported by the dual mechanism model (Pinker 1991) which states that there are different mechanisms for learning regular and irregular inflectional forms – irregular forms are memorised as discrete items, whereas regular forms are produced as the result of the application of morphological rules, thus requiring

more than simple memorisation to learn. The extra cognitive effort required to write more in the non-correction groups, considering the fact that every pair of essays covered similar grammatical ground, could have led to deeper processing, leading to the improvements seen in this study. It has also been suggested that different aspects of morphology (derivational and inflectional) could be learnt in different ways, although this is beyond the scope of this paper (Lardiere 2006).

The deterioration in both groups of syntactic errors is possibly a reflection of the progression of the course as a whole, with students in both groups trying out more complex syntactic structures. This is in contrast to the improvement of tense errors in both groups, there being an explicit focus on tense, but not on syntax in lessons at both levels. It is noteworthy that this deterioration was smallest in the non-correction group, again lending weight to Truscott's simplicity hypothesis (2001). This is also in line with Sheppard (1992), who noted a deterioration of grammatical errors with correction.

It is also of note that the biggest effects were observed in the elementary groups. The only error category to decline in all of the groups was untreatable errors in the elementary correction group, whereas untreatable errors in the elementary non-correction group made the biggest improvement. All other groups made improvements. See chart 1.

Chart 1. U = untreatable errors
 T = treatable errors
 N = non-correction group
 C = correction group
 I = intermediate group
 E = elementary group



This could be explained by different SLA processes and affective factors at the different levels. This is mirrored in the questionnaire results (see page 27). Ayoun (2004), Hasbún (2000) and Salaberry (2000) in studies into the acquisition of morphological features in Spanish (a morphologically richer language than English) all found differences between different levels of competence. Elementary students are learning to learn and so are probably greatly effected by changes in teaching methodology possibly caused by affective factors, whereas more experienced learners have developed their own learning and studying strategies, thus the depth of processing factor mentioned above may apply to a greater degree to elementary learners who are grappling with new systems and ways of thinking. However, higher level learners have already understood the underlying principles of language learning. This may lead corrected elementary students to look at their errors and try to treat all error types the same way, thus leading to repeated, but not necessarily deep, processing which has a confounding effect on untreatable errors. On the other hand, the non-correction elementary group may have made such a marked improvement with untreatable errors because of the extra deep processing which was required to produce new essays, considering the fact that essays were separated by one week, allowing time to reflect (consciously or subconsciously) on their work, before being required to write again on the same broad subject. This could push them into the type of deep processing which is required for acquisition of morphological forms. This type of deep processing may be blocked by the processing of corrections.

4.2 Questionnaire results

4.2.1 General observations

As noted above (page 21) the questionnaire results from the correction and non-correction groups confirm previous research (Enginarlar, 1993; Ferris, 1995; Hedgcock & Lefkowitz, 1994; Leki, 1991) which found that students want their errors to be corrected. It is interesting to note that the only two students who liked non-correction were from the pre-intermediate group, and in fact were among the most able students in the class. This begs the question: are higher level groups more likely to accept non-correction? Truscott (1999:116) says that his students are happy without being corrected. These are university level students, who are probably at a much higher level than the learners in the present

study. This hypothesis is in conflict with the statistical evidence of the present study, which showed that untreatable errors improved the most without correction and deteriorated the most with correction in the elementary groups (See page 26). This is an aspect to be taken up in pedagogical considerations (see page 32).

Higher level students are already competent language learners, with their own strategies for dealing with difficult language features and so are more likely to accept non-correction, whereas elementary students probably feel the need of the reassurance that correction brings. This has important pedagogical implications.

4.2.2 Avoidance

Avoidance of difficult structures caused by correction is difficult to look at with statistical methods. Chandler (2003; see Chandler 2004:346) uses a holistic rating method to examine this aspect and Kepner (1991) analysed “higher level propositions” in order to gauge ideational quality, which could be considered as a way, if not of looking for avoidance, for analysing negative effects of avoidance, in that replacing a problematic structure with another complex form would not be noted in this analysis (see one student's observation, page 21), whereas simplification of ideas as a result of avoidance would be noted by this measure.

The questionnaire is a more direct method of judging avoidance, although students may not be aware of it themselves, especially in the long-term. As noted above, 21% of students admitted avoidance of difficult structures. Considering the nature of the students (teachers themselves) this is probably reliable. If students avoid features in the short-term, this effect will probably be passed on to future essays, even though the students were not aware of this. If, as noted above (page 21) this avoidance does not lead to simplification, but searching for another valid way of expressing the same concept, this could be a useful technique for students to foster, not in the context of grammar instruction, but in that of communication instruction. This too has pedagogical implications (see page 32).

4.2.3 Affective factors

Truscott (1996:354) talks about the “inherent unpleasantness of correction” and mentions that students who are not corrected are more relaxed and tend to write more freely than those who are corrected. As noted above (page 22) 25% of students reported feeling worried about errors as a result of non-correction. Thus the majority of students fit in with Truscott's position, but a sizeable minority do not. These students affective needs must be taken into consideration when thinking about methodological implications. However, 58% of students reported feeling freer to experiment with new forms, without worrying about being corrected. Even though a large minority did not recognise this, it could be a reflection of the deeper processing necessary for the acquisition of morphological forms, as noted above (see page 25).

There are many definitions of written fluency used in the literature (see Wolfe-Quintero et al., 1998:13-32 for a discussion), generally concerned with rate, length and sometimes complexity of writing (ibid: 14); thus the non-correction groups wrote more fluently in that they produced much more original writing, and, although it has not been measured in this study, this reported experimentation with new forms could also lead to an increase in complexity, hence fluency.

4.3 Pedagogical Implications

Bringing together all the above points in the discussion section, it is possible to consider some tentative pedagogical implications from these findings, which lie somewhere between Truscott's (1996:361) conclusion that “[g]rammar correction has no place in writing classes and should be abandoned” and Ferris' (2004:59) position that “[e]rror treatment, including error feedback by teachers, is a necessary component of L2 writing instruction.”

Ferris (2004:60) notes that “[d]ifferent types of errors will likely require varying treatments.” This has been borne out by this study, in that it has been seen that different classes of errors react in different ways to correction. Truscott's (2004:342) position is that *grammar* correction should be abandoned. The most correctable errors in this study were

not grammatical, but lexical, thus errors such as word-meaning and misspelling can be corrected effectively.

Truscott (2001) offered a theoretical analysis of which grammatical errors may be more or less correctable. He makes the point that correction should be based on “the possibility of success” (ibid:94), which contrasts with the traditional position as expressed by Lee (1997), who suggests that selective correction should be based on the level and needs of the learner. This study has shown that morphological errors do not respond to correction, so they should not be corrected, especially with low level students, as it possibly causes confusion, which leads to deterioration. Certain syntactic features, such as preposition errors seem to be slightly correctable, whereas others (definite article use) improved more without correction. This implies, firstly, that more research is necessary in order to establish which types of syntactic errors are more treatable than others; the verb tense error category could also be explored in more depth. However, one tentative pedagogical implication is to follow Truscott's (2004:95-99) idea of “simplicity”, that is correcting simple errors and leaving more complex ones uncorrected. This also has the benefit of being easy for writing teachers to apply, which is important considering that teachers sometimes have difficulty in correcting (see Lee 2004, which reports that about half of corrections were wrong in one study). This method is also simple for students to understand, which should lead to reduced anxiety, facilitating learning (Gardner and MacIntyre, 1993; MacIntyre and Gardner, 1989).

One of Truscott's (1996, 2001) criticisms is that correction produces stress and a negative learning atmosphere. Thus a selective correction method should make sure that not too many errors are corrected in order to reduce red ink and so improve the learning conditions. This also addresses the concerns of students who may feel worried about their errors with a total non-correction methodology, as they would know that their errors are being considered by the teacher in a selective correction methodology. This type of correction also serves to show the teacher where students weaknesses are, which can be responded to in other ways. For example, follow-up grammar instruction could be given in order to address specific grammatical problems identified as untreatable by correction. This instruction should try to stimulate deep processing, which has been seen to be helpful in the acquisition of morphological features. It could be in line with behaviourist thinking

as grammar exercises, or more naturalistic by exposing students to examples of correct target language in authentic contexts.

Research on sequences in SLA has already been cited as one reason for the ineffectiveness of correction (see page 7). Warnings have been given as to applying this research directly to the classroom (Bahns, 1990; Lightbrown, 2000); however, if course designers and textbook writers pay attention to these sequences, then in the context of writing in a general English course, if the writing follows the course (thus the developmental sequences) it is possible that the effectiveness of correction will be raised.

It is also important to consider the aspect of fossilisation. One argument in favour of correction is that it helps stop fossilisation (Ferris, 1999; 2004); however, research on errors which have already fossilised shows that correction alone is ineffective and that most short-term improvements backslide into the old errors (Han 2003; Selinker, 1972; Selinker & Han, 2001). This is another argument for rejecting the use of short-term studies into correction as an indication of long-term acquisition. The Multiple Effects Principle (Han and Selinker, 1999) implies that multiple methodologies working together are required to destabilise fossilised forms, therefore correction of fossilised errors could play a part in a de-fossilisation strategy if it is used in tandem with other specific methods, such as extra grammar instruction, to eradicate the poorly learnt form. On the other hand, in courses where this is not possible, it is best not to correct fossilised errors.

Newly learnt forms may be better targets for correction, although little research has been carried out into this area. It is interesting to note that Bitchener et al. (2005), which found beneficial effects for some error types, was a study carried out in a general English course where students were probably in the process of acquiring the forms under analysis, whereas most of the other studies in table 1 were in university writing course contexts, where it is possible that many of the errors were already fossilised.

An associated area which has received little attention in the studies on written error correction is the origin of the error. Johnson (1988) discusses the difference between 'errors' and 'mistakes'. He defines errors as faulty interlanguage caused either by a lack of knowledge or by incorrect learning. Mistakes, on the other hand, are seen as being

problems with performance under difficult conditions when the form could be produced correctly under non-stressful conditions. It is possible that this is simply the difference between metalinguistic knowledge and true acquisition, in that metalinguistic knowledge can be accessed under simple conditions but not when communicating under pressure. Another type of error is a 'slip', that is a form which can normally be produced correctly, but was as a one-off produced erroneously (Davies, 1983; Lennon, 1991).

The fact that errors (in the more general sense of the word) are caused by different processes implies that different strategies are needed in treating them. More research is required to determine which, if any, are treatable with correction; however some reasoning is possible on the pedagogical implications of these differences. It is improbable that 'errors' caused by a lack of knowledge can be eradicated by indirect correction alone as the student does not know why it is wrong. Similarly a student who has incorrectly learnt a form, either recently or longer ago leading to fossilisation, will probably need more than simple underlining, or error codes, as they do not know how to correct the error. Direct correction may be more effective in the case of these 'errors' in interlanguage as it shows the student the correct form, even though it still does not explain the reason for the error. There is possibly no benefit in correcting slips as the student already knows and uses the form correctly most of the time. 'Mistakes' may be more promising candidates for correction as in these cases students are in the process of acquiring the form and correction, particularly codes and underlining could help students to use the form correctly under different circumstances, in the behaviourist paradigm (see page 6). It is therefore important for the teacher to understand the type of error made by the student in deciding whether to correct or not.

The majority of students in this study and others looked at (see page 21) reported wanting correction, so this selective correction methodology suggested would fulfil this desire; however, it is important to consider learner training, that is telling students what to expect in terms of error correction from the teacher and how to deal with the corrections received. This approach would let students know that they should not be worried about errors as they will not receive a sea of red ink on their texts, thus it should foster the type of experimentation that non-correction encourages; it could also nurture a feeling of confidence in the teacher which non-correction may not lead to. The corrected students in

this study were required to re-write their texts, incorporating the corrections, which was effective for the lexical errors in this study. It is necessary for students to engage with the corrections in order to produce the repeated exposure required for vocabulary retention, thus recording new or corrected words is important. The re-writing may not be necessary if students are encouraged to recycle the corrected vocabulary in different contexts over time. This has the advantage of saving class or homework time copying in order to re-write, which could be valid in a general English course, but is maybe less so in a writing course where several drafts of any text are usually written. More research is required to investigate this question.

Learner training is also necessary on the issue of avoidance. In the context of a general English course, learners should be trained not to avoid structures that they know to be difficult for them, as tackling these problem areas may push towards acquisition. In a writing course, avoidance could be a valid procedure, as the focus is on communicative competence which could be improved, not by simplifying, but by teaching strategies to look for alternative ways of expressing concepts. In this case, students should be made aware of the choices available to them in terms of trying to use difficult structures or looking for alternatives. This methodology fits in to a collaborative learning paradigm, with the teacher and students working together to decide which errors to work on (Dillenbourg, 1999).

The correction method used in this study was a simple error code, which was seen to have some success with lexical and simple grammatical errors, thus this or similar codes could be recommended to apply this suggested methodology for 'mistakes' (see page 31). In this aspect some of the short-term studies looked at in table 1 can be useful in determining which correction method to use. Greenslade and Félix-Brasdefer (2006) found a code to be effective, whereas Ferris & Roberts (2001) found both underlining and codes to be effective, thus either codes or underlining could be a valid way of applying a selective correction methodology. It has been observed that correction takes up a great deal of teacher time (Chandler, 2003; Ferris and Roberts, 2001; Goldstein, 2004; Lee 2004; Truscott, 1996), so underlining could be useful in that it is quicker and simpler for a busy teacher. It could, therefore, reduce the possibility of inaccurate correction. One possible argument in favour of the use of codes is that they could encourage deep processing;

however, it has been seen that not depth of processing but number of retrievals may be necessary for vocabulary retention (see page 25) and that the most correctable errors are lexical, so deep processing is not necessary. Any methodology aiming to correct morphological and complex syntactic errors should strive to encourage deeper processing. One possible method could be reformulation (Adams, 2003; Qi and Lapkin, 2001), although this is very time consuming, and impractical for most teachers (Sachs and Polio, 2007). Direct correction, that is the teacher giving the correct form, has also been suggested (Chandler, 2003). This method does not require deep processing, so should not be helpful in the case of morphological errors, but could help with lexical errors, and may be more useful in the case of incorrectly learnt forms (see page 31).

The corrective feedback in this study included the possibility of students asking the teacher/researcher for clarification of the corrections. It is possible that the extra cognitive effort required by the student to try to explain and understand the problem in conversation with the teacher helped in acquisition of some of the corrected forms (Lindgren and Sullivan, 2003). Bitchener et al. (2005) also found the greatest improvements in their group with corrections and teacher conferences. This is in harmony with a Vygotskian view of cooperative learning, with teacher and student working together to achieve learning (Vygotsky, 1978). It could be concluded that effective correction should be backed up with personal teacher explanations, although this point needs more research. This could be effective because of the extra cognitive processing required to explain verbally the problem combined with personal attention from the teacher trying to ensure that the point has been understood.

The Vygotskian perspective can also help in deciding which errors to correct. The zone of proximal development (ZPD) is defined as “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers” (Vygotsky, 1978:86). The concept was extended by Wood et al. (1976) by introducing the term 'scaffolding', referring to how tutoring can help in problem solving. This could be applied to written error correction if the errors corrected are in the ZPD, which in practice would mean that they are not too conceptually difficult for the student.

This could be the case for 'mistakes' (see page 31) as the student is in the process of acquiring these forms, and understands their use generally, but fails on certain occasions.

Another aspect to consider is that of the level of the students. It has been seen that the strongest positive effects for non-correction and the strongest negative effects for correction of morphosyntactic errors were both with the elementary class. Thus selective correction is especially important at lower levels. The questionnaire results, however, show a tendency towards a stronger desire for correction at lower levels. This indicates that writing as part of a general English course at elementary level should not be considered as a vehicle for grammar instruction by correction. However, uncorrected free-writes may well foster the acquisition of more complex forms, so this should be encouraged, even at elementary level. This could be applied by having a five minute free-writing slot at the end of lessons in order to apply what was learnt during the lesson. This also encourages the acquisition of writing skills at this low level. As learners at higher levels are more able to deal with learning strategies the negative effects were less evident. However, there were no positive effects for morphosyntax, so complex errors should still not be corrected at higher levels. This suggestion should be treated with caution, as the present study only analysed elementary and pre-intermediate students.

Work on oral feedback shows uptake of spoken corrections to be scant. Panova and Lyster (2002) concluded that retrieval and production may be more effective than hearing correct forms in oral corrections. This could be comparable in written production as similar cognitive processes are at work when language is produced both orally and in written form. The difference being that students have more time to think before producing written language. This extra thinking could give students the possibility of accessing metalinguistic knowledge, which can be incorporated into written production, which could then lead to acquisition. More research is required to investigate this area, but this leads to the conclusion that learner training is required to encourage students to think about their metalinguistic knowledge before and during writing in order to use learnt but not yet acquired forms.

In the context of general English courses leading to exams, this metalinguistic knowledge may also help students with the type of grammar questions found in exams, so could be

helpful in terms of exam technique, even if the impact on acquisition is debatable. It is also possible that whilst innate interlanguage is tapped when writing first drafts, metalinguistic knowledge could be used when editing and writing second drafts, thus correction could be used to foster this metalinguistic knowledge along with learner training of how to re-draft using the metalinguistic knowledge thus acquired. This again needs more research to validate these suggestions. Terrell (1991) talks about some of the ways that this metalinguistic knowledge could be useful.

To summarise the pedagogical implications, the results of this study considered in the light of the literature on ELT and SLA lead to the following recommendations:

- Selective correction is a valid methodology.
 - Lexical errors can be corrected.
 - Simple grammar errors may be corrected, but not with elementary level students.
 - Complex morphosyntactic errors should not be corrected.
 - Fossilised errors should not be corrected, unless as part of a larger defossilisation strategy.
- Follow-up grammar instruction may be a way to help improve problems with complex forms.
- A positive atmosphere should be cultivated in class towards writing correction.
- Students should use corrected lexis repeatedly, possibly by re-writing and recording words in vocabulary books.
- Students should be encouraged not to avoid difficult language in the context of a general English course.
- Simple codes may be an effective correction system for lexical errors.

These implications must be treated with caution, considering several flaws noted in this study, as noted below, which may have confounded somewhat the results. It should also be noted that these points are from the point of view of fostering long term acquisition. If the goal is improvement over drafts of a single essay, then these points do not apply.

4.4 Strengths and weaknesses of the current study and ways forward for future research

Guénette (2007) analysed 32 different studies on the question of the efficacy of error correction, analysing their research design, and found many aspects which could lead to doubting the accuracy of many conclusions reached. The present study will now be examined in the light of Guénette's analysis.

Many of the studies considered in table 1 and also by Guénette (2007) were short-term, so did not address the question in terms of the effect of correction on underlying SLA. The design of the present study over two months, with seven 100-word essays (each different for the non-correction groups, and four original texts and three re-writes for the correction groups), should be long enough to show tendencies towards long-term effects. This compares with many studies in table 1, for example Bitchener, et al. (2005) whose study lasted twelve weeks, with four 250 word tasks, and Semke's (1984) 10-week study of 10 minute free-writes. Further research could perform a similar study over a longer period, possibly with less frequent and longer writing, as the frequency of one essay a week proved tiring for many students. This may have adversely affected their performance, thus the outcomes of the study. A larger sample, both in terms of longer writing and more students would allow more reliable conclusions to be drawn.

Guénette (2007:43) speaks about the importance of a “control group that is in every way comparable to the experimental groups in terms of proficiency level, writing conditions, and instructional context.” This is one of the major strengths of the present study, and in particular the participatory action research design, as the control groups (non-correction) were in the same class as the experimental (correction) groups, thus the writing conditions and instructional contexts were equal. On the point of proficiency level, the groups were selected to have approximately equal average errors per 100 words on the first essay to ensure homogeneity between the groups, with a balance of strong and weak students in each group. The six students who were taken out of the statistical analysis due to absence may have changed this balance, although the average errors per 100 words on the first essay after the absent students were removed was still similar for correction and non-correction groups at each level (NE=3.74, CE=4.66; NI=7.35, CI=6.79), An uneven mix of

high and low ability students between groups could potentially make one group more or less likely to respond better to either method (correction, or non-correction). This is in line with Gardner's (1985, cited in Myles, 2002) socio-educational model, which recognises that individual learner differences play a big part in language acquisition. Further research could therefore examine the question of whether higher or lower ability students respond better to correction.

Another criticism that Gu nette has of several studies (Fazio, 2001; Kepner, 1991; Semke, 1984) is that these are studies of journal writing, which would not normally be corrected, revised or graded, thus students would probably pay little attention to corrections made in this type of writing, therefore correction would be expected to have minimal effects. The present study avoids this pitfall as the writing studied was an integral part of the course, linked to the grammatical and lexical aspects of the syllabus.

Gu nette (2007:47) criticises Sheppard (1992) for being a study in a foreign language context, as opposed to the second language university contexts of many of the studies mentioned. Her criticism is that as the students only have to write in the classroom they may be poorly motivated to learn to write. Research is also needed in these contexts to be able to give reliable advice to teachers in a variety of teaching contexts. The students in the present study were all highly motivated to learn, both because of the final exam itself, and the job as English teachers that they would receive on passing the exam, thus Gu nette's criticism seems inapplicable to this study.

The choice of correction method is an important factor, as it could be argued that correction could have been effective with a better method. The simple code used in the present study was seen in class, and on correcting second drafts to have been effective. Elementary and pre-intermediate correction groups considered together made an average of 5.72 grammatical errors per 100 words on the first draft but only 1.65 on the second draft (essays 1 and 2), showing that the corrections were mostly understood and applied; however, it is possible that a different correction method would have led to different results. A future study could compare other correction techniques with non-correction with a similar student population. It would also be interesting to analyse what types of errors were not accurately corrected on second drafts.

The present study follows Semke's (1984) design, in that corrected students were required to re-write essays while non-corrected students wrote new texts. This was criticised by Guénette (2007:48) as the corrected students only wrote half as much original material as non-corrected students. However, this criticism seems to miss the point – if corrected students re-write and write new material then they receive more instruction than non-corrected students. In the present study both groups received exactly the same amount of time for both general studying and work on writing as they were actually in the same class, at the same time. Guénette (2007:49) also notes that different groups in Robb et al. (1986) received different classroom instruction. This is again addressed by the participatory action research design, in that both groups received exactly the same instruction. The potential confounding variable of instruction received leading to observed improvements is also removed by the two groups being in the same class and receiving the same instruction. Future studies could look into different ways of handling correction, for example re-writes, simply reviewing corrections but not re-writing, and the use of error correction charts.

Guénette (2007:50) mentions the role of grades in student motivation, and the effects this may have had on performance in several of the studies examined. In the present study, no grades were given to individual essays in either group, so this did not have any effect.

The problem of students misunderstanding corrections is also noted (Guénette 2007:50). In the present study, the questionnaire results show that the majority of students (93%) found the corrections to be clear. This is reinforced by the improvement over two drafts, as noted above (page 38), which indicates that students understood most (but not all) corrections, in order to be able to correct them on the next draft. It is also important that the corrections are consistent over the course of the study, and indeed over an entire course in order for students to see clearly that a certain form is erroneous. The blind double marking carried out indicates consistency, as an intrarater reliability of 98% was found.

Another weakness of this study is that, although a sample of student writing was blind double marked, the researcher alone divided the errors into the categories in table 6. Although every effort was made to ensure consistency and impartiality, working with a second researcher would have ensured more reliable results. Thus the double marking

validates the conclusions of the first research question, but not the second. Subjective judgment was often required to decide if something was an error or not (see Chandler, 2003: 276), whether an error was grammatical or not (for example “shoot” instead of “shot” - is it a simple spelling error, or did the student not know that the past tense was required? Context, and surrounding errors were used to judge these questions), and in which category to place certain errors (i.e. simple or non-simple preposition errors).

The greatest weakness in the present study is the fact that as the essay questions used followed the general grammatical and lexical content of the course, there may have been little opportunity for students to have worked consistently on any given structure or lexis over subsequent essays (see appendix 1 for the essay questions). The follow-up essays for the non-correction groups were all designed to elicit similar grammar and lexis, and at the same time the correction groups worked on understanding corrections and re-drafting, so each two-essay cycle re-covered the same linguistic ground, but subsequent essays moved on to different areas. The last essays attempted to elicit similar language to the first questions in order to be comparable; however, the different language used (and the correction/non-correction used) in the intervening essays may not have had any direct effect on performance in the last essay. An analysis of specific grammatical points (as in Bitchener, et al., 2005), such as the third person 's', or the present perfect could have obviated this problem if all essays (not just first and last) were analysed, in order to track the change over time of the error-rates.

Another problem is that the grammar category analysis carried out may, for example, compare many third person 's' errors in one essay with many negative prefix errors in a later essay, as they are both morphological errors. This could be taken into account by an analysis of incorrect uses per total occurrences of a specific form; this would also remove the problem of some forms occurring many times in certain essays and less in others. During the preparation of this paper a more detailed analysis of individual error types was carried out, and the errors were observed to be evenly distributed across many of the 30 error-types looked at, which implies the validity of the chosen analysis, but it is beyond the scope of this study to analyse in this depth. A more detailed analysis of the data is, therefore needed in order to ascertain the validity of the data in this study. See appendix 2 for the breakdown of all errors.

Future studies should also try to elicit similar language in all questions set, and choose a consistent error analysis method. The category analysis developed in the present study could serve as a model for studies which elicit similar language to be analysed, such as journal writing, or ESP writing courses. It may be more appropriate for studies of longer texts over longer periods, where more errors in each category would be gathered, so the different error types in each category should average out over a large quantity of writing. An analysis of a particular grammar point may have been more appropriate with the present data. The chosen analytical method simply looked at the number of errors, so one error repeated ten times in one text would give the same error-count as ten different errors in the same category in another essay, whereas they are clearly different cases. Counting error-types per student, as opposed to total errors, would have alleviated this problem.

A related issue is whether more errors actually means worse writing. Myles (2002:9) makes the point that “the more content-rich and creative the text, the greater the possibility there is for errors at the morphosyntactic level.” It must be acknowledged that these students are in the process of experimenting with new language, and that errors are an unavoidable part of acquisition. It is interesting to note that the elementary students in the present study made an average of 6.51 errors per 100 words on the last essay. This compares with Chandler's (2003) upper intermediate/advanced students' error rate of 6.0 errors per 100 words (two groups combined). As students progress in proficiency they move on to tackle more complex language, so errors remain even though the quality of the writing improves (Myles, 2002). This being said, as students progress they need to overcome previous errors, so analysing a specific grammatical point with error counts over a particular course could still be a valid measure.

The context of the courses is also important, as most of the previous studies examined were in second language contexts. It is possible that correction could be more useful in foreign language settings where students have little opportunity to interact in the target language outside the classroom. There is a need for more research comparing SL and FL contexts to understand these differences.

The preceding criticisms notwithstanding, this type of writing is common in general English courses, so the present study still serves as a guide to teachers in similar contexts.

This analysis of the strengths and weaknesses of the present study has highlighted many ways to improve the study and many possible areas for future research. These can be summarised by Ferris (2004:49), when she asked the question “where do we go from here?” referring to future directions for research into the question of written error correction. She makes three main points:

- There is a need for more controlled longitudinal research on this question.
- Future studies must be consistent and replicable, in that they must clearly report all design parameters, and analyses carried out.
- Detailed studies are required into different aspects of error correction, such as:
 - Student handling of corrections.
 - The value of supplementary grammar instruction.
 - The relative correctability of different error types.
 - The long term effects of different types of error feedback.

(Adapted from Ferris 2004:56-58)

Ferris (2004:56) mentions a “catch-22 situation” where longitudinal studies often do not have non-correction control groups for ethical reasons, and controlled studies are not longitudinal, so do not show long-term effects on SLA. Participatory action research is, therefore, proposed as a solution to this problem in future studies, as students have the choice of which group to belong to and the confounding variable of different treatment of different groups is removed. Another way to alleviate the ethical dilemma is by giving content feedback to one group and form feedback to the other group.

It is also important to select an appropriate analytical method, which removes as many confounding variables as possible. The error category method developed in this study could serve as a model for future studies, but care must be taken to ensure that like errors are compared with like, by devising elicitation tasks which are comparable. A mechanism must also be designed to take into account the possible confounding variable of exactly the same error being repeated several times and counted each time as a new error. An analysis

of specific grammar points and lexical items would take into account the aspect of task difficulty, as harder tasks would produce more errors, but with harder structures, whereas correct or incorrect occurrences of a structure under analysis would remain more or less unaffected by the task difficulty if the form under analysis has been elicited in all tasks. Studies into longer writing with more students would produce a larger base of errors, so these problems would be averaged out with a larger population.

5. Conclusions

Many lessons have been learnt from this study, both in terms of the pedagogical and SLA impact of error correction on L2 writing, and how to conduct a study into this question.

Important factors to consider when designing any similar study are: how to ensure that the groups are identical in every way; and the choice of an analytical tool appropriate to the data. Participatory action research has been seen to be one promising method to ensure homogeneity of groups.

It has been suggested that an appropriate pedagogical response to written errors is one of selective correction, using an appropriate correction technique to correct only simple errors, such as lexis, spelling or simple grammar errors, whereas more complex morphosyntactic errors are best left uncorrected and responded to with other methods, for example grammar instruction or a more naturalistic exposure to target forms.

These two suggested responses go to the heart of the matter – behaviourist, or naturalist – one of the most fundamental dichotomies in TESOL, and in psychology as a whole. A behaviourist perspective sees correction as necessary and as grammar instruction as a valid tool in aiding acquisition (Skinner, 1957), whereas naturalists would advocate a non-correction methodology with exposure to authentic language as the natural root to acquisition (Chomsky, 1959).

Whilst the theorists argue over these questions classroom teachers must still respond to their students' errors. Guenette (2007:41) came to the conclusion that “no matter what

teachers did, some students would benefit from focused instruction and corrective feedback while others would not.” All students are individuals, and so one student may react in a different way to another in exactly the same class at the same level.

It behoves teachers to try to understand which students to help in what way. Action research, either participatory or reflective on the part of the teacher, is therefore one means to help teachers decide how to treat the specific errors of their individual students, combined with learner training to help students react in a useful way to the correction methodology chosen.

This dissertation can thus be concluded by re-phrasing the opening comments – to correct or not to correct? that is every teachers' personal question.

Chris Baldwin

Word count: 14,935 (excluding tables)