
The Effects of Form-focused Feedback on Quality of Writing and Performance on Accuracy

JOHN PELOGHITIS

Abstract

There has been much debate concerning the impact of form-focused feedback in helping second language (L2) students improve the accuracy and quality of new pieces of writing. A growing body of research on error correction in L2 writing classes has addressed how students improve in accuracy from one draft to the next. However, there is a lack of adequate studies examining the most critical issue in this debate: does form-focused feedback help students to improve their writing over time? The following study reports on the impact that indirect form-focused feedback has on quality and performance in accuracy. 35 lower intermediate Japanese students were divided into two groups; a group receiving form-focused feedback and a control group that did not. The treatment group received indirect feedback on three types of grammatical errors (verb errors, word errors, and noun ending errors) to trace the effects on quality and accuracy between two argumentative essays written at the beginning and end of a year-long course. Independent samples t-tests and paired samples t-tests were applied and the data revealed that both groups made significant gains but differences existed in the how each group improved. The control group actually made greater progress in the overall quality of their essays but the group receiving form-focused feedback significantly improved in performance on accuracy for two of the three targeted error categories. The research suggests that feedback incorporating indirect error correction as compared to no correction does not appear to help L2 learners produce better writing. This finding is consistent with many studies that have compared groups receiving form-focused feedback with a control group. The performance in accuracy in the feedback group also reveals what many have assumed, that longer treatment periods yield a greater effect on improvement in accuracy. An important implication from the following study is that practitioners need to carefully decide on a method of feedback (i.e. content-only or content and form) that best applies to their students – one that considers students' proficiency level, previous writing experience (in both L1 and L2) and motivation – and incorporate reading and revision strategies to complement it.

INTRODUCTION

Over the past several decades the role of corrective feedback in second language (L2) writing has been a constant source of interest and debate among teachers and researchers. Research studies examining this issue have made the distinction between feedback on form which addresses grammatical errors and punctuation and feedback on content which often consists of message-related comments to help learners develop and organize their ideas. Much of the debate in L2 pedagogy has involved form-focused feedback and whether it leads to improvement in accuracy and quality in writing.

In his critique of form-focused feedback, Truscott (1996) examined previous studies, and reported that practical problems exist in how teachers provide feedback to their students. In one such study, Cohen and Cavalcanti (1990) observed that teacher bias was a large factor in determining the type of feedback a student was given. Teachers targeted certain types of problems in their comments which were based on preexisting judgments rather than addressing more predominant problems. In addition, the researchers found inconsistencies in the teachers' feedback; teachers often did not notice many of the students' errors. Similarly, Zamel (1985) reported that teachers were not consistent or systematic in giving feedback and that the quality of the feedback was often poor. In addition, other studies have revealed that students' responses to feedback have been problematic. One reason is that students may have limited knowledge of how to edit or lack strategies for dealing with teacher feedback on essays. Cohen (1987) found that students often had difficulty making revisions because they did not understand the teacher's comments. For these reasons, experts have argued that the role of form-focused feedback be severely limited (Krashen, 1984; Zamel, 1985) and some have claimed that it is even harmful because it models poor priorities concerning the writing process and diverts attention away from more important writing issues (Truscott, 1996, 1999).

Despite the lack of evidence on this issue, studies confirm that teachers tend to provide more form-focused feedback than content-focused feedback, and that grammatical accuracy plays a significant role in assessment. As a consequence, L2 learners are more likely to direct their revisions to mechanical features and lexical accuracy (Hedgecock & Lefkowitz, 1994). Other researchers warn that the negative aspect of overt feedback on grammar may affect the quality of subsequent essays (Hendrickson, 1980; Semke, 1984).

Proponents of form-focused feedback have argued that there are convincing reasons for L2 writing instructors to continue the practice of providing feedback in their classes. A primary reason is that studies comparing students who received different types of feedback treatments (i.e. direct feedback, indirect feedback, teacher-student conferencing, error logs) have shown improvement in accuracy (Fathman & Whalley, 1990; Ferris, 1997; Lalande, 1982). Another equally important reason is that students value form-focused feedback and they think it is helpful to improve their writing (Cohen, 1987; Cohen & Cavalcanti, 1990; Ferris & Roberts, 2001; Leki, 1991). Although student perceptions about what teachers should do may not give credence to the argument that form-focused feedback is useful, if teachers refuse to provide comments on grammatical errors, then student motivation and confidence may decrease (Ferris 2003; Brice & Newman, 2000.) Finally, Ferris (2005) argues that just because the practice of form-focused feedback has been problematic in the past does not mean that it should be abandoned. Future L2 research designs need to examine the effects of various feedback processes more adequately to help teachers implement them effectively. The following section surveys some of the important findings from studies that have examined this issue.

Research on the Issue of Improvement and Accuracy

Relatively few studies have directly addressed whether L2 students who receive form-focused feedback improve in accuracy and quality in new pieces of writing, as opposed to those who do not receive any feedback on form. Kepner's (1991) study has often been cited by critics as convincing evidence that form-focused feedback does not lead to improved accuracy (Polio, Fleck, & Leder 1998; Truscott, 1996). Kepner's study examined journal entries from 60 intermediate Spanish non-majors divided into two treatment groups; a group receiving direct error correction and a group receiving message-related feedback. The results indicate that no significant differences in accuracy existed between the two groups. In fact, the group receiving the message-related comments had more "higher level propositions" in their writing than the students who received direct error correction.

Sheppard's (1992) study also confirmed that no significant differences existed in accuracy between a group receiving comprehensive coded feedback on errors and a group receiving marginal comments regarding clarity. Sheppard compared the use of verb forms and sentence boundary markers in two narratives written by 26 college-level ESL students over a ten-week period, and used teacher-student conferences to complement each feedback treatment. Sheppard's findings indicated that the group receiving marginal comments performed better on sentence boundaries and produced more complex texts. Sheppard attributed the lower complexity in the group receiving coded feedback to avoidance – students were opting to choose simple structures over more complex ones to reduce the possibility of error.

In an early study Semke (1984) investigated the effect of different feedback treatments on 141 German foreign language students over a ten-week period. Students were divided into four groups; comments only, direct correction, direct correction with comments, and indirect coded correction. Using t-scores as a measurement, no significant differences in written accuracy were observed between the four groups which led Semke to conclude that error correction does not improve written accuracy.

Polio, Fleck, and Leder (1998) analyzed the effect feedback had on journal entries written by ESL college students over a 7-week period. The students were divided into two groups; a control group in which no form-focused feedback was given and a group receiving direct correction. The latter group had time allotted in class to review grammar points and learn about editing techniques. Despite writing four journal entries each week, students were only asked to submit one revised entry per week. Effects were measured by comparing the differences between a pre-test in-class essay and a post-test in-class essay. The authors determined that no significant contrasts existed between the groups.

In one of the few studies that examined EFL students, and over a lengthy period (9-months), Robb et al. (1986) observed changes in error-free T-units to total T-units on compositions in four distinct feedback groups of Japanese university students (direct feedback, coded feedback, highlighting, and error counts noted in the margins). The data revealed that all four groups significantly improved in accuracy over time; however, no significant improvement in accuracy was found between the four groups. Given the results, the authors concluded that providing a more explicit or direct feedback on student errors is not justified. There has been some confusion about whether the group receiving error counts in the margins is a true control group. Ferris (2003) categorizes this study as one without a control group because all groups had help in locating the source of their errors. Truscott (2007) counters by making a case that knowledge gained by the error counts written in the margins was so limited that this group can be considered a control group.

Although studies that compared control groups with groups receiving a type of form-focused feedback suggest that correction does not impact improvement in accuracy, research design issues have led some to question many of the findings. For example, Ferris (2003) argues that Kepner's (1991) study is not reliable because it is "so riddled with design and validity issues" (p.60). Kepner's study does not contain a pretest measure of students' errors and proposition counts and there was no apparent requirement for students to revise their journals. Sheppard's (1992) study has also been criticized on methodological grounds. For instance, no outside rater reported on the analysis of errors and, even though comprehensive coded feedback was utilized, no information was reported on the improvement of (or lack of) other error categories. The study by Semke (1984) has been questioned on several grounds particularly because no interrater reliabilities were reported and the types of writing in which her conclusions are based were journals and a free writing post-test exercise – genres which are traditionally not "corrected" (Ferris 2003). Polio et al.'s study used a different measurement, that is, an in-class essay to determine improvement in grammatical accuracy rather than a journal entry which had been the type of writing the teacher responded to throughout the treatment. In addition, the duration of the treatment may have been too short to measure improvement. Al-

though Robb et al.'s study is one of the few to examine improvement in accuracy over a lengthy period, the data is unclear and the pretest scores among the treatment groups showed that the students were at considerably different levels of proficiency before the treatment initiated. Thus, there is clearly a need for more valid research that not only compares the effects of receiving form-focused feedback and no such feedback, but also one that traces the long-term effects of such treatments (Ferris, 2002, 2004; Truscott, 1999).

A different group of studies which analyzed various feedback treatments (i.e. direct feedback, indirect feedback, teacher-student conferencing, error logs) provides evidence that feedback improves accuracy when revision occurs over an extended period of time (Chandler, 2000; Ferris 1995, 1997; Ferris et al., 2000; Lalande, 1982). Truscott claims, however, that evidence from this body of research is misleading because in all of these studies, there is no control group that received no correction on form:

The limits of such evidence are clear: in the absence of a control group, one cannot determine whether observed gains resulted from correction or from other factors. Thus, even if corrected students consistently showed significant improvement in their accuracy, this finding in itself would tell us nothing about the value of correction. The way to draw implications from uncontrolled studies is to quantify the gains they find in a way that does allow comparison with general standards and with gains expected in the absence of correction (Truscott, 2007 p.263).

While these studies confirm that receiving form-focused feedback improves the quality of the texts after revision took place and causes writers to become more attentive to patterns of error, there is still a lack of compelling evidence that the presence or absence of revision makes a difference in the long run (Ferris, 2003).

Research on the Issue of Error Type

A growing body of research has examined the impact of feedback on accuracy by looking at improvement in how students use specific linguistic constructions between drafts and in new pieces of writing. A majority of these studies has reported that consistent differences exist in how students progress on certain linguistic categories after revision (Chaney, 1999; Ferris et al. 2000; Ferris and Roberts, 2001; Frantzen and Rissell, 1987; Lalande, 1982; Sheppard, 1992; Bitchener et al., 2005). Ferris (1999) suggests that much of this variability depends on whether the error is "treatable" or "untreatable." Treatable errors are linguistic structures that occur in a rule-governed way such as errors in verb tense and verb form, and noun endings. Untreatable errors are idiomatic or idiosyncratic and require an acquired knowledge of the language. Examples of untreatable errors include most word choice errors and problems with syntax. It is important to note that the term "treatable" does not necessarily imply that the errors are easier to revise as some treatable errors such as articles have very complex rules which give even highly proficient L2 learners difficulties.

Three recent studies have operationalized and explored this dichotomy. In a study by Ferris et al. (2000) data revealed that students were more successful in reducing certain treatable errors (verb tense and verb form) than in others (noun endings). Regression actually occurred in sentence structure errors (untreatable) and in articles (treatable). The study also showed that teachers prefer to provide indirect feedback on treatable errors and direct correction on untreatable errors. Ferris and Roberts (2001) conducted a short controlled experiment comparing how students progressed in accuracy on an in-class essay after receiving either indirect feedback or none at all. The results indicate that all three groups were able to correct the treatable errors better than the untreatable errors. Bitchener et al. (2005) compared three different feedback treatments (indirect, indirect with conferencing, and no feedback) on separate linguistic categories in new pieces of writing. By calculating the percentage of correct usage on the tested constructions, they found that students who received feedback significantly improved in accuracy on the use of definite articles and

past simple tense verbs both of which are treatable errors. No significance, however, was found in the use of prepositions which are less treatable and more idiosyncratic.

In evaluating the research to date, a number of questions remain concerning the value of form-focused feedback on L2 writing. Further research is needed to explore if form-focused feedback is useful in helping students improve the accuracy and quality in new pieces of writing. There has also been a lack of research examining lower-level students in EFL contexts, and over a prolonged period on instruction. In order to start addressing these needs, the following 26 week study was undertaken with 35 lower-intermediate EFL students at a Japanese university to investigate the extent to which corrective feedback on targeted linguistic forms helped students improve the accuracy and quality in new pieces of writing.

Research Question

What is the impact on improvement in overall essay scores and on the use of three types of grammatical forms between students who receive form-focused feedback and students who do not receive form-focused feedback?

METHODOLOGY

Participants and Instructional Context

Unlike most error correction studies which have focused on advanced ESL students over the short period of time, the participants in this study were enrolled in a two-semester compulsory writing course for first-year intermediate students. The participants were all Japanese females studying at a private women's university located in Tokyo, Japan. Although the students were not English majors (American Culture and European Culture), they were required to take a set amount of courses in two foreign languages to fulfill the requirements of their major. The participants had no experience in process-oriented expository writing, and while most (89%) have been studying English for approximately 6 years, their exposure to L2 composition was limited, and their previous instruction was heavily focused on grammar, vocabulary and reading for the purpose of passing university entrance exams.

The primary goal of the writing course was to familiarize students with a process-approach to writing, and teach basic academic writing skills with particular attention on generating topics, writing cohesive paragraphs, and organizing ideas into clear, logical compositions. There were two types of writing assignments; expository writing (four 5-paragraph essays) to acquaint students with the organizational styles common to academic environments and journal writing (13 entries) to enhance fluency and self-expression.

Both writing classes were taught by the same instructor. The instructor had been teaching English in Japan for six years at the time of the study and has implemented various methods of feedback over the years in his writing courses. The evaluators were two instructors who taught at the same university – an American with over nine years experience in teaching EFL writing and a Canadian with six years experience in writing instruction. Written consent was obtained from each student at the outset of the study, and to ensure anonymity, student numbers were used in place of names during the rater(s) evaluation of the essays.

Research Design and Procedure

Students from two classes met once a week for two semesters from April to July and from September to January. Each class was 90-minutes and students met 26 times during the year; thus students received 39

hours of classroom instruction. All the writing classes were taught by the same instructor and four five-paragraph expository writing assignments (argumentative, compare/contrast, descriptive, and argumentative) were required. Both groups followed the same schedule and the same process of feedback was followed on all of the essays.

The participants were separated into two treatment groups. Group 1, the control group, consisted of 18 students receiving only feedback on content. In Group 2, 17 students received feedback on content and indirect feedback on three categories of errors. Each group was required to submit three drafts for each of the four essays. The first drafts of the two argumentative essays were submitted in week five and week twenty-two of the course, and were selected to measure the quality and accuracy in student writing over the course. The topics of both essays were open, and models were provided to help students understand the organizational patterns and rhetorical structures commonly used in argumentative writing. The raters evaluated the essays on a rubric adapted from Ferris and Hedgecock (1998 p.310) for lower intermediate students. See Appendix A for details of the scoring rubric. Spelling and punctuation were eliminated from the rubric because they are not considered to be grammatical elements, an issue that Truscott (2007) mentioned in his critique of earlier studies that implied a linked improvement in spelling with accuracy. The students were evaluated from 1 to 5 in five categories; main ideas, organization/coherence, supporting ideas, grammar, and vocabulary.

Students were encouraged to write over 400 words for each draft. The first draft was returned with an attached handout that contained both positive and constructive comments as well as a preliminary score in each of the five categories. Many of the comments targeted specific problems in the essays and made suggestions on improvement. In some cases, students were referred to a particular page in the textbook to help them improve a feature in their essays (i.e. an unclear main idea). The same procedure was followed for the second draft. However, selected errors were underlined for the group receiving form-focused feedback. It is important to note that although the raters did score the essays, the comments and indirect correction were provided by the instructor, and that the scoring and commenting were done independently. When drafts were returned, 30 minutes were set aside in class for students to review the comments and corrections, and make revisions. To help both groups improve their content and grammar during the revision sessions, six mini-lessons (30 minute sessions – 3 in each semester) were introduced throughout the course. The mini-lessons introduced self-editing techniques and provided practice on identifying and correcting problems in sample essays. In addition to the revision sessions in class, students had time to revise their essays outside of class and were given three to four days to submit the next draft.

Rather than providing a system of comprehensive feedback (the marking of all errors), three groups of errors were selected based on frequency and error type. Three categories of error were selected from a short in-class writing assignment which took place in the second class. Both raters were trained on a coding scheme to identify grammatical errors and a simple percentage agreement was used to measure inter-rater reliability. The raters reported that sentence structure errors were the most common (23.4% of the total errors) followed by the use of verbs (19.4% of the total errors). Word choice errors (17.8% of the total errors) and noun ending errors (11.2% of the total errors) were the next most common error categories. Although sentence structure errors (i.e. fragments, unnecessary words or phrases, omitted words or phrases) were the most frequent error type, they were not included in the study. The primary reason for the exclusion is that such errors are generally considered to be quite complex and therefore better attended to at a later stage of language acquisition. In addition, the simple percentage agreement test revealed that the raters had a low level of agreement (73%) in identifying and categorizing this type of error which would have caused inconsistencies with data collection. To deal with this issue, regular sentence building activi-

ties and reading assignments were assigned to help students develop more accurate and complex sentence constructions.

Analysis

This research study was designed to measure how students improved over time in the composition of new pieces of writing. Improvement in quality refers to gains made in essay scores and improvement in accuracy is assessed by using the three grammatical forms selected in this study correctly. To analyze if significant improvements occurred in essay scores between the two groups of students, two independent samples t-tests on the first drafts of both argumentative essays were performed. These tests also established a pre-treatment and a post-treatment measure which were used to compare mean differences between the groups. A paired samples t-test was then applied to each group to report on any significance found within each group over the treatment period (26 weeks). This process was repeated for accuracy in using the selected grammatical forms. The alpha for achieving statistical significance was set at .05. Additionally, effect sizes using Cohen’s d were calculated on the both kinds of t-tests to help evaluate the stability of the research across samples and the strength of significance. A value of .2 is generally considered a small effect size, .5 a medium effect size, and .8 or more a large effect size.

Because two raters were used throughout the study, interrater reliability measures were conducted. The reliability measures were first established through a random sample of essays (approximately 22% of the total) before any evaluation or marking was performed on the two essays. First, a Pearson product-moment correlation coefficient was performed to determine interrater reliability between the scores given in the two essays. The overall computed Pearson correlation coefficient r on the scores was .740 for the first essay and .728 for the second essay. Correlation is significant at the .01 level, two-tailed. The results show (r= .740, p= .000 and .728, p= .000) indicates that there was a significantly positive and a strong relationship between the scores assigned by Rater 1 and Rater 2 on both essays. In fact, all of the categories scored by the raters had a strong correlation (higher than .643, p< .01). After reliability was achieved, student essays were evaluated by a single rater. Second, the same procedure was used to test interrater reliability for determining the percentage using the selected grammatical forms correctly. On the first essay the Pearson correlation coefficient r was .782 (p= .000) for verb errors, .458 (p= .17) for word errors and .703 (p= .000) for noun ending errors. The weaker correlation coefficient was found on word, .458 (p= .017). This may be due to the fact that word choice is more susceptible to multiple interpretations, which was likely to create less agreement between the raters in determining correctness. The second essay showed similar results. Although the values of coefficient r varied between each grammatical form, the coefficients were convincing. The Pearson correlation coefficients for the essays scores and grammatical forms are listed in Table 1.

TABLE 1. Interrater reliability tests on scores and error type using Pearson product-moment correlation

1 st argumentative essay				2 nd argumentative essay			
For essay scores		For performance in accuracy		For essay scores		For performance in accuracy	
Main Idea	.755*	Verb errors	.782*	Main Idea	.789*	Verb errors	.712*
Organization	.780*			Organization	.728*		
Support	.834*	Word errors	.458**	Support	.810*	Word errors	.527*
Grammar	.738*			Grammar	.660*		
Vocabulary	.643*	Noun ending errors	.703*	Vocabulary	.655*	Noun ending errors	.768*
Overall	.750*			Overall	.728*		

*p< .01

**p< .05

RESULTS

This section presents the results investigating the impact to which the control group and the group receiving corrective feedback improved in quality and accuracy when producing new compositions.

The mean scores and standard deviations on the scores for the first and second argumentative essays can be seen in Table 2. Independent samples t-tests were performed on each category and revealed that no significant differences existed between the groups on the first essay. Although the control group did have slightly higher scores in all of the categories except vocabulary, the test shows that both groups exhibited comparable writing skills at the start of the course.

TABLE 2. Independent sample t-tests on the 1st drafts of both argumentative essays

Category	Group	Number	1 st argumentative essay (N = 35)				2 nd argumentative essay (N = 35)			
			Mean	SD	T (df = 33)	P	Mean	SD	T (df = 33)	P
Main Idea	Control	N=18	3.00	.91	.410	.685	3.55	.51	1.88	.243
		N=17	2.88	.78			3.29	.77		
Organization	Control	N=18	3.05	.72	.487	.629	3.72	.57	2.920	.006*
		N=17	2.94	.66			3.18	.53		
Support	Control	N=18	2.67	.59	.097	.924	3.44	.51	1.267	.214
		N=17	2.65	.61			3.18	.73		
Grammar	Control	N=18	2.72	.46	.414	.681	2.94	.54	-1.682	.102
		N=17	2.65	.61			3.29	.69		
Vocabulary	Control	N=18	2.78	.55	-.228	.821	2.94	.54	.015	.988
		N=17	2.82	.64			2.94	.75		
Overall	Control	N=18	14.22	1.93	.409	.685	16.61	1.46	1.077	.289
		N=17	13.94	2.13			15.88	2.45		

* $p < .05$

The same procedure was followed for the second argumentative essay. The results showed that the control group received higher scores on organization ($M = 3.72$, $SD = .57$) than the group receiving the feedback ($M = 3.18$, $SD = .53$). This difference was statistically significant, $t(33) = 2.920$, $p < .05$ indicating that the control group's higher average score on organization was more than what would have been expected due to chance. A Cohen's d analysis revealed that this difference had a large effect size (-0.981). The large effect in scores on organization ($d = -0.981$) signifies that about 83% of the control group's scores are at or above the mean of the form-focused feedback group. Although no other statistically significant differences were found between the other items, the control group outperformed the feedback group in all categories except for grammar and vocabulary.

To reveal if significant differences existed in the mean scores between the first argumentative essay and the second argumentative essay paired samples t-tests were conducted for each category. Both groups improved their scores in all aspects of writing and significant gains were made in several categories. For the control group statistical significance was reached over the mean gains for main ideas ($d = .745$), organization ($d = 1.031$), and support ($d = 1.396$) as well as overall. The feedback group made statistically significant improvements regarding main ideas ($d = .529$), support ($d = .787$), grammar ($d = .982$), and overall scores. The Cohen's d analysis indicates that many of the effect sizes were medium or large. Table 3 pro-

vides details of the paired samples t-tests.

TABLE 3. Paired samples t-tests for essay scores

Category	Essay	Control group (N = 18)				Form-focused feedback group (N = 17)			
		Mean	T (df = 17)	SD	P	Mean	T (df = 16)	SD	P
Main Idea	Essay 1	3.00	-2.755	.91	.014*	2.88	-2.384	.78	.030*
	Essay 2	3.55				3.29			
Organization	Essay 1	3.05	-4.761	.72	.000*	2.94	-1.725	.66	.104
	Essay 2	3.72				3.18			
Support	Essay 1	2.67	-4.507	.59	.000*	2.65	-3.497	.61	.003*
	Essay 2	3.44				3.18			
Grammar	Essay 1	2.72	-1.719	.46	.104	2.65	-3.395	.61	.004*
	Essay 2	2.94				3.29			
Vocabulary	Essay 1	2.78	-1.144	.55	.269	2.82	-1.461	.64	.163
	Essay 2	2.94				2.94			
Overall	Essay 1	14.22	-6.754	1.93	.000*	13.94	-4.576	2.13	.000*
	Essay 2	16.61				15.88			

*p<.05

Independent samples t-tests were also conducted on the mean percentages to ascertain how the two groups compared in the performance of accurately using the three types of grammatical forms. The results of these tests can be seen in Table 4. The independent samples t-test on the first argumentative essay revealed that both groups had a similar percentage in the performance of the three types of forms. The feedback group did have a slightly higher percentage in their performance of verbs and word choice, but they outperformed the control group in regard to noun endings. Independent samples t-test on the second essay shows that no statistically significant differences were observed between the groups in the performance of any type of grammatical form. However, the feedback group did make greater gains in their use of verbs and noun endings. Little or no progress in accuracy in the performance for word forms was found.

TABLE 4. T-test for independent groups on performance in accuracy in grammatical forms

Grammatical forms	Group	1 st argumentative essay				2 nd argumentative essay			
		Mean (%)	SD	T (df = 33)	P	Mean (%)	SD	T (df = 33)	P
Verbs	Control Feedback	76.39	8.33	.504	.618	79.05	5.70	-.990	.330
		74.94	8.66			81.18	6.95		
Words	Control Feedback	79.89	7.34	1.089	.284	80.39	6.33	1.503	.142
		77.29	6.71			77.18	6.32		
Noun endings	Control Feedback	81.17	5.72	-.739	.465	83.78	6.26	-1.071	.292
		82.76	7.04			86.23	7.30		

*p<.05

Table 5 includes the data from the paired samples t-tests. The students in the control group did not show any significant improvement in the performance of the targeted forms between the two essays, but they almost reached statistical significance in noun ending errors, $t(17) = -1.99$, $p = .063$. However, the form-focused group did reach statistical significance on their improvement with verb forms, $t(16) = -3.584$,

$p < .05$, and nearly reached statistical significance on noun endings, $t(16) = -1.815$, $p = .088$. A Cohen's d analysis also revealed a medium effect size on feedback group's level of improvement ($d = .794$).

TABLE 5. T-tests for paired samples on performance in accuracy on grammatical forms

Grammatical forms	Essay	Control group (N = 18)				Form-focused feedback group (N = 17)			
		Mean (%)	SD	T (df = 17)	P	Mean (%)	SD	T (df = 16)	P
Verbs	Essay 1	76.39	8.33	-1.99	.063	74.94	8.66	-3.584	.002*
	Essay 2	79.05	5.70			81.18	6.95		
Words	Essay 1	79.89	7.34	-.297	.770	77.29	6.71	.086	.933
	Essay 2	80.39	6.33			77.18	6.32		
Noun endings	Essay 1	81.17	5.72	-1.486	.155	82.76	7.04	-1.815	.088
	Essay 2	83.78	6.26			86.23	7.30		

* $p < .05$

DISCUSSION

The main purpose of this study was to evaluate the impact of form-focused feedback on the improvement in essay scores and the performance in accuracy in the use of grammatical forms. According to the results of this study, both the control group and the corrective feedback group improved the scores of their second argumentative essay. The paired samples t-tests confirm that both groups made significant gains in forming main ideas, providing support, and in overall improvement in their new essays. This finding should not be surprising considering that both groups of students were exposed to a prolonged period of instruction and were required to produce a substantial amount of writing (i.e. four five-paragraph essays and thirteen journal assignments). The major difference found between the two groups was that the control group had a statistically significant gain in organization scores while the form-focused feedback group reached a statistical significant gain in grammar. This finding suggests that focus on form was successful in helping students increase their scores in accuracy when writing new essays. Students undoubtedly benefited from the indirect method of correction which allowed them a greater opportunity to identify and monitor their mistakes throughout the course. It is important to mention, however, that the improvement in accuracy came at the expense of organization. Even though the form-focused feedback group did in fact improve in organization, the level of improvement was not commensurate to the gains observed in the control group. The same can be said for the control group; they also improved in their grammar scores but much less than the group receiving the attention to form. This finding is more easily explained though since the control group received no feedback on form.

The independent samples t-test on the second argumentative essay also revealed that scores on organization for the control group were significantly higher than the feedback group, and that the effect size was large. An obvious explanation for this finding is the control group received a smaller amount of feedback which in turn allowed them to focus more attention on looking at the sample models they received to address the teacher's comments on organization.

A likely reason for the gains in accuracy at the expense of content is that students who received form-focused feedback were more preoccupied with correcting grammatical errors as opposed to text-based errors. As discussed above, this is perhaps because revising grammatical errors is considered more manageable than revising content-based errors. Content-based revisions rely more on knowledge of writing genres,

organizational patterns and rhetorical structures, concepts that are often unfamiliar to L2 learners. A study by Paulus (1999) which investigated “think-aloud protocols” during revision, supports this premise. The study found that students took more initiative in correcting surface-level errors, and relied on teacher feedback to make content-based revisions. This also explains the control group’s progress. They focused more on attending to the teacher’s comments on content since they received no form-focused correction. This would explain why the control group had higher scores on the content areas of their essays.

In examining how students improved their overall essay scores, the results illustrate that the control group made greater gains (a mean difference of +2.39) than the form-focused group (+1.94). One might suggest that this strengthens Truscott’s (1996) claim that focus on form is harmful and should be abolished since the feedback group could have made greater gains. However, one could argue that error correction did lead to a greater improvement in accuracy scores, so the treatment had success in helping students to write more accurately. Success, in other words, depends on what outcomes teachers and students expect or desire.

The data concerning how students improved in their performance using the selected grammatical forms suggests that form-focused feedback contributes to greater accuracy. The independent samples t-tests found that the form-focused group made larger gains than the control group in the mean percentage for accurately using verbs (a mean difference of +6.24 versus +2.66) and noun endings (a mean difference of +3.47 versus +2.61). This finding contradicts many of the conclusions in the research to date that compared the effect on accuracy between a feedback group and a no feedback group. It is likely that the attention to form helped students notice their common errors and make the appropriate revisions on the assigned essays. Students may have also benefited from receiving indirect feedback which has more potential for developing problem-solving strategies in students.

It is important to note that neither group improved in performance for the word group. The group receiving form-focused feedback actually regressed in accuracy (a mean difference of -0.11) between the two drafts. This confirms previous studies that have found that students are more successful in correcting certain “treatable” errors than “untreatable” ones (Ferris et al., 2000; Ferris & Roberts, 2001; Bitchener et al., 2005). However, not all of the treatable errors improved similarly and form-focused feedback had a large impact on the revision of verb errors. Word choice errors are more likely to be problematic in lower-level students because their vocabulary is still quite limited and they may lack strategies or knowledge of the rules to make the appropriate lexical choices.

LIMITATIONS, FUTURE RESEARCH AND CONCLUSIONS

In order to shed some light on the value of form-focused feedback in L2 writing, the present study investigated the impact that form-focused feedback has on quality and performance in accuracy. This investigation tests the underlying assumption by many second language practitioners that students who receive form-focused feedback are better off than those who do not. The study found that lower-intermediate EFL university students benefit from both methods, but in different ways. Although the control group made a significant level of improvement in organization and in the overall score of their argumentative essays, the form-focused feedback group made a significant gain in grammar and performance in accuracy in two grammatical areas. Based on the results, one can argue that the overall strength of an essay is more important than grammar, and students, particularly lower-level students, lack the ability to deal with multiple forms of feedback when composing. However, the positive effects of form-focused feedback group cannot

be easily dismissed. Form-focused feedback did lead to significantly higher scores in grammar and helped students to use verbs and noun-endings more accurately in new pieces of writing. The results can also be seen as encouraging since they were not attained after a revision session; rather, the results suggest that students who received the form-focused feedback incorporated revision strategies that were built sequentially throughout the course.

One limitation to the following study is that the data was not triangulated. Multiple, independent methods of obtaining data in a study is important to support the validity and reliability of the findings. Although the use of two independent raters was used to determine reliable data collection, improvement in essay scores and performance in accuracy could have been better supported by including questionnaires concerning student perceptions of the revision process. Because no measures were incorporated in the study to investigate the students' decision-making process, the conclusions are based on post-hoc assumptions about what exactly students did during the writing process and why they chose to focus more or less attention on a particular aspect of their writing. It would be useful in the future to include "think-aloud protocols" when students revise their writing, as Ferris (2003) suggests. This would give researchers more opportunities to understand the effect that a certain comment or correction has on a student's decision-making process.

Based on the empirical evidence from this study, providing form-focused feedback does seem to help students improve the grammatical accuracy in their essays. However, the improvement in accuracy comes at the expense of focusing less on other aspects of writing which can have an adverse effect on quality. Writing teachers need to carefully assess whether such a sacrifice is helpful.

L2 learners often find it difficult to develop all aspects of feedback simultaneously due to the complex processes of writing in a second language. As a result, learners selectively attend to only those aspects that are automatic or have already been proceduralized (O'Malley & Chamot, 1990). In order to enhance or facilitate language production, students can develop particular revision strategies. This may or may not be true for more proficient L2 writers who have more strategies and linguistic competence to deal with multiple issues reported in the teacher feedback. It seems likely that incorporating mini-lessons targeting self-editing techniques and practice and by providing time in class to revise help students to focus on the process of improving their essays. The simple opportunity of allowing students time to read and reread their essays in class may have contributed to the improvement on accuracy and higher quality essays; a conclusion reached in studies conducted by Fathmon and Whalley (1990) and Russikoff and Kogan (1996).

The following study indicates that although form-feedback did help students to successfully revise some of the targeted linguistic forms and led to higher scores for grammar, the no feedback group improved to a greater degree on most of the categories scored on essay quality. Future research should explore whether or not other not form-focused feedback influences other types of linguistic forms particularly untreatable ones where learners may have more difficulty, and if the impact is commensurate across different levels of proficiency. More research should also address how other factors such as time to revise in class and self-editing techniques and strategies, influence the revision process. Such studies may offer more insight into the complex nature of revision and offer teachers an alternative to form-focused feedback – one that addresses grammatical correctness without sacrificing improvement in the quality in content.

Appendix A. Rubric for scoring in-class essays (Adapted from Ferris and Hedgecock, 1998 p.310)

Score	Main Idea	Organization/ Coherence	Supporting ideas	Vocabulary	Grammar
5	Addresses topic and are concise and clearly presented	Logically organized and coherence is marked by transitions	Uses specific and appropriate supporting ideas and examples	Excellent word selection and usage	Minor and infrequent errors
4	Related to the topic and reasonably clear	Solid organization and uses coherence markers	Contains ideas and examples that support the main idea	Above average word selection and usage	Minor errors present but do not distract the reader
3	Related to the topic but could be presented more clearly	Lacks logic or coherence because of connectors or transitions are not used effectively	Supporting ideas do not adequately address the main topic or lack detail or clear examples	Average word selection and usage	Some major errors present that hinder comprehensibility
2	Marginally related to the topic or difficult to identify	Little organizational structure and lacks or misuses connectors and transitions	Ideas are either lacking in number or are unclear and irrelevant in supporting the topic	Weak word selection and usage	Distracts the reader because errors are frequent and major
1	Does not address the topic or lacks a main idea	Lacks both organization and coherence	Supporting the main idea is ineffective due to inappropriateness or the lack of development	Extremely weak word selection and usage	Difficulty understanding due to major and frequent errors

References

- Brice, C. & Newman, L. (2000, September). *The case against grammar correction in practice: What do students think?* Paper presented at the Symposium on Second Language Writing, Purdue University, West Lafayette, IN.
- Chandler, J. (2000, March). *The efficacy of error correction for improvement in the accuracy of L2 student writing.* Paper presented at the AAAL Conference, Vancouver, BC.
- Cohen, A. (1987). Student processing of feedback on their compositions. In A. L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp.57-69). Englewood Cliffs, NJ: Prentice-Hall.
- Cohen, A. & Cavalcanti, M. (1990). Feedback on written compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp.178-190). Cambridge: Cambridge University Press.
- Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp.178-190). Cambridge: Cambridge University Press.
- Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly*, 29, 33-53.
- Ferris, D. R. (1997). The influence of teacher commentary on student revisions. *TESOL Quarterly*, 31, 315-339.
- Ferris, D. R. (2002). *Treatment of error in second language writing classes*. Ann Arbor, MI: University of Michigan Press.
- Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum.
- Ferris, D. (2004). The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing*, 13, 49-62.
- Ferris, D. R., Chaney, S. J., Komura, K., Roberts, B. J., & McKee, S. (2000, March). *Perspectives, problems, &*

- practices in treating written error.* Colloquium presented at International TESOL Convention, Vancouver, BC.
- Ferris, D., & Hedgcock, J. S. (1998). *Teaching ESL composition: Purpose, process, & practice.* Mahwah, NJ: Lawrence Erlbaum.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161-184.
- Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity in second language writing. *Journal of Second Language Writing, 3*, 141-163.
- Hendrickson, J. M. (1980). The treatment of error in written work. *Modern Language Journal, 64*, 216-221.
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal, 75*, 305-313.
- Krashen, S. D. (1984). *Writing: Research, theory, and application.* Oxford: Pergamon Press.
- Lalande, J.F., II (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66*, 140-149.
- Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals, 24*, 203-218.
- O'Malley, J. & Chamot, A. (1990). *Learning strategies in second language acquisition.* Cambridge, England: Cambridge University Press.
- Paulus, T. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing, 8*, 265-289.
- Polio, C., Fleck, C., & Leder, N. (1998). 'If only I had more time': ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing, 7*, 43-68.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*, 83-93.
- Russikoff, K. & Kogan, S. (1996, March). *Feedback in ESL writing.* Paper presented at the 31st Annual TESOL Convention, Chicago, IL.
- Semke, H. (1984). The effects of the red pen. *Foreign Language Annals, 17*, 195-202.
- Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23*, 103-110.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*, 327-369.
- Truscott, J. (1999). The case for "the case for grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing, 8*, 111-122.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16*, 255-272.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19*, 79-102.