

Does written corrective feedback make a difference in EFL academic writing?

Author: Dr. Pia Köhlmyr  
Affiliation: Gothenburg University, Sweden  
Corresponding author: Pia Köhlmyr  
Email: [pia.kohlmyr@sprak.gu.se](mailto:pia.kohlmyr@sprak.gu.se)

*The Author*

Pia Köhlmyr, PhD is currently senior lecturer and teacher educator at the Department for Languages and literatures, Gothenburg University, Sweden. Her research interests include L2 teaching and learning, (formative) feedback on student L2 writing, contrastive and error analysis focusing English written proficiency and grammar. She has extensive experience from teaching English both at high school and higher education level.

Does written corrective feedback make a difference in EFL academic writing?

Abstract

This study investigated the effect of feedback on written assignments in English, analysing texts from student teachers of English to see whether their academic writing would improve in subsequent new writings, in relation to type (direct or indirect) and syntactic form (e.g. questions or statements) of written feedback. Also, the teacher intent of the feedback (e.g. giving directives or positive encouragement) was analysed for this purpose.

Although results showed improvement in several targeted areas it was also found that some linguistic and textual problems had instead increased, despite the feedback provided. Significant differences were noted regarding feedback characteristics, which raise questions in relation to learner up-take. The findings merit further study, especially since previous research has shown that corrective feedback generally does make a difference (Hartshorn et al., 2010; Jönsson, 2013; Truscott & Hsu, 2008) and that feedback is central to the development of student learning (Hattie, 2009; Hattie & Timperley, 2007).

**Keywords:** EFL writing, feedback types, formative assessment, ESL writing, teacher feedback, written corrective feedback

---

## 1. Introduction

Feedback as an integral part of teaching and learning, and its possible impact on students' writing skills, has been debated since Truscott's claim that feedback in the form of error correction should be abandoned (cf. Chandler, 2003; Ferris, 1999; Truscott, 1996). Since then, research has shown that there is more to feedback than simply correcting linguistic errors; it can (and should) be a formative tool for teaching and learning. Despite this, feedback is still an area in need of more investigation to understand its full potential for supporting and improving learning (Hyland & Hyland, 2006; Jönsson, 2013; Sadler, 2010; Weaver, 2006). When feedback is timely, appropriate and understood, it may improve written production and also provide the relevant tools for the learner to be able to work independently (Jönsson, 2013; Nicol & Macfarlane-Dick, 2006). Furthermore, if feedback is "directed at the right level it can assist students to comprehend, engage, or develop effective strategies to process the information intended to be learnt" (Hattie, 2009, p. 177). Studies on students' understanding and use of feedback are still fairly limited in number (Jönsson, 2013) and many of them show inconsistent results. However, some point out that students neither understood the feedback given, nor understood what they were expected to do with it (Goldstein, 2004; McMartin-Miller, 2014). It seems obvious, then, that students first have to be made aware of the feedback and then "be given ample opportunities to apply the corrections" (Guénette, 2007, p. 52), how else can they learn from it? Hattie (2009) states that "when feedback is combined with effective instruction in classrooms, it can be very powerful in enhancing learning" (p. 178).

Researchers have called for more "research in naturalistic classroom contexts to explore the real needs of teachers and students" (Lee, 2014, p. 1) as well as students'

role in self-monitoring their own learning. Such research will add new knowledge to the current research base on feedback in writing. There is also comparatively little research done in the EFL context, and Lee (2014) states that such studies “will be a welcome addition to the field” (p. 2).

Although writing is seen as a generic skill, most feedback research has focused on the editing of drafts of text(s), i.e. revision, rather than investigating its effect on the production of new texts. Thus, there is a need for more research on the effect of feedback on new writings. In the present study, revisions were not required to be resubmitted. Instead, new texts were produced, whereby the generic skill of writing academic texts in English was focused, while the topics differed.

Student teachers of English in training, provided a unique possibility to work with feedback from a pedagogical point of view, while also catering for more direct individual needs. The purpose was twofold: by working on their own writing skills, being exposed to and themselves using feedback, they were in a position to learn from teachers and peers how to work with feedback in their future classrooms. It is important that teachers have the pedagogical skills to provide “sustainable feedback” (Hounsell, 2007) so as to help students not only develop their writing skills but also develop learner autonomy through self-regulatory processes (Parr & Timperley, 2010) and this can be achieved, e.g. by working hands-on with feedback, experiencing it first-hand.

## 2. Previous research

A substantial amount of research investigating feedback within the narrow field of error correction, e.g. studies in ESL writing, have generally focused one particular grammatical area (Bitchener & Knoch, 2010; Bitchener, Young, & Cameron, 2005; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007; Sheen, Wright, & Moldawa, 2009). This model has been supported by e.g. van Beuningen, de Jong and Kuiken (2012), claiming that there seem to be limitations as to how comprehensive feedback can be before any effects may be lost, indicating that feedback may work better if it focuses on more than one, but not too many, areas. However, others suggest that this might be challenging (Guénette, 2012) since teachers feel obligated to give comprehensive feedback lest they be accused of being “incompetent or lazy if they had not flagged the error” (p. 120). Van Beuningen et al. (2012) point out that “evidence on the language learning potential of unfocused CF [corrective feedback], which involves comprehensive correction of every error in students’ writing, is scarce” (p. 5). Only a handful of studies have actually investigated if comprehensive feedback may yield a learning effect on new texts (cf. Ellis et al., 2008; Sheen et al., 2009; Truscott & Hsu, 2008; van Beuningen, de Jong, & Kuiken, 2008). There are now several studies that focus on the long-term effects of corrective feedback looking at results on pretests and comparing this to delayed posttests, but they are generally about focused feedback, thus targeting only one or two error types at a time. This can be criticized for not being how teachers actually work in real life. Therefore, using comprehensive feedback seems more realistic. Other approaches (e.g. Ellis et al., 2008; Lee, 2011), again, discuss the advantages of more selective feedback. This could be timesaving for the teacher and by highlighting typical error patterns the responsibility for self-correction of unmarked errors is transferred to the learner (McMartin-Miller, 2014, p. 25), thus promoting learner self-efficacy. However, “little research has examined how teachers actually practice selective feedback in the classroom, and how students perceive their teachers’ approach” (Lee, 2014).

Today, there is an increasing number of studies looking into whether written feedback may help improve writing, some focusing on the impact on new texts (Bitchener, 2008; Bitchener & Knoch, 2010; Ellis et al., 2008; Sheen, 2007) but there is still uncertainty as to the actual positive long-term effectiveness (Truscott & Hsu, 2008). The fact that learners sometimes do not understand comments or concepts used by their teachers (e.g. Lea & Street, 2000; Parr & Timperley, 2010; Weaver, 2006; Wingate, 2010) or that they may have misunderstood the need for their own active part in feedback (McMartin-Miller, 2014) for it to be efficient, should be considered in relation to student use of feedback.

Feedback can be delivered in many different forms and previous research has generally focused on two major categories; *direct*, providing a correct solution, or *indirect*, indicating an error or error area but not providing correction (e.g. Sheen, 2011). Results have been inconsistent, for example Chandler (2003) concluded that direct correction was preferred by students and “for producing accurate revisions” (p. 267), whereas indirect feedback by underlining errors, was felt to be better for promoting learning since students had to self-correct. Sheen (2007) also found that direct corrective feedback seemed to work well, at least for specific grammatical features. Others suggest that direct written feedback (WF) combined with metalinguistic comments may promote long-term improvement of language accuracy (e.g. Bitchener & Knoch, 2010; Bitchener et al., 2005; Wingate, 2010). On the other hand, Eslami (2014) found that “the indirect feedback group outperformed the direct feedback group on both immediate post-test and delayed post-test”, and Ferris (2006, 2011) also suggests that indirect WF has more potential to promote long-term learning. However, in van Beuningen et al. (2012), both direct and indirect comprehensive CF led to improved accuracy compared to two control groups, thus, results still seem inconclusive.

Other studies have investigated the impact of different linguistic forms of teacher comments. These discuss the use of *imperatives*, *questions* and *statements* (e.g. Ferris, 2003; Gascoigne, 2004; Mehr, 2013; Sugita, 2006), and the way comments are formulated may decide how or if the feedback is used (cf. Ferris, 2003). Lea and Street (2000) suggest that “imperatives cause difficulty in interpretation, which confuse or upset students, and the opportunity for learning is thus lost”. On the other hand, several studies report that students engage better with imperatives, resulting in improved revisions (e.g. Gascoigne, 2004; Mehr, 2013; Sugita, 2006). Others advocate the use of questions as a means to promote learner self-regulation (Gu nette, 2012). The use of questions has been debated, proposing that it could be difficult for L2 learners to interpret the meaning of such implicit pointers (Ferris, 1997, 2007; Hyland & Hyland, 2001) and therefore students would not act upon the feedback given.

Thus, it seems that, despite ongoing research, the essential question of how we can help students develop their writing skills “has remained elusive” (Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, & Anderson, 2010). Therefore, to find out more about the use and usefulness of feedback on written assignments in English, the present study carried out in 2012 had a twofold aim: one part focusing learner beliefs and self-assessment, and the other looking into whether feedback would lead to improvement regarding content and language in the writing of subsequent new texts. This paper reports on the latter part, focusing on:

- whether there would be an improvement in writing from pre- to posttest after feedback
- whether any improvement would apply to all, or only specific, feedback areas

- whether a relation between type of feedback and improvement could be discerned

The project was set out as a pilot study using a relatively limited material, thus no major generalizations can be made from the results. However, several interesting tendencies were found relating to intent, type and form of the feedback, giving rise to further ideas as to what a future large- scale study should focus on.

### 3. Method and Material

The study is quasi-experimental in that participants were enrolled in a program and randomly divided into two classes. No control group was used since feedback is part of the regular teaching of the course and thus, depriving one group feedback, was neither realistic nor conducive to motivation in this type of research. It was also considered unethical (cf. Bitchener & Ferris, 2012; Evans et al., 2010) since participants, albeit part of two groups, are very likely to confer and compare the course content and teaching since they are all part of the same educational program.

It was longitudinal, running over one semester, providing a pre- and a posttest. This naturalistic, classroom-based research intended to explore relationships between feedback and the process of writing.

#### 3.1 Participants

The original cohort consisted of 32 student teachers training to become English teachers in the upper secondary school, chosen by convenience sampling, i.e. the researcher had easy access to these groups enrolled in the teacher education program. A major advantage was that these students attended a five-year educational program, and therefore the same students would be available for delayed posttests later on. All students were informed that participation was voluntary (a) at the introduction seminar and (b) through a flyer handed out at the same seminar. Confidentiality was guaranteed since only the researcher could match the written material to each individual for comparisons of results.

The participants formed two classes (A and B) and the members were numbered (A 1-21 and B 1-11) from the start. These originally designated codes were retained by each participant despite some late drop-outs. After the posttest was collected, 25 participants remained (class A=16 and B=9) having taken the pre- and posttest, and attended the writing course classes for feedback. The dropout rate was thus 22%, totalling 25 student teachers and their two teachers.

In order to be eligible for the teacher education program, a general level of Swedish is required equivalent to a pass mark from upper secondary school level. Thus, students are expected to master Swedish to this level, irrespective of original L1. Among the participants, there were four other L1s—Chinese, Bosnian, Persian and Arabic— but they all mastered Swedish very well and therefore no problems were foreseen with the course assignments (writing English summaries of Swedish original texts).

The level of English was monitored through a diagnostic test given to all first-year students of English at the department. This is a multiple-choice test of general vocabulary size and applied basic grammar knowledge, resting on many decades of statistically stable and reliable results and, although it is not a production test, it was considered useful in establishing the overall proficiency level of the participants. The

participants in the study were well set at, or above, the pass mark of 10 points, with a mean of 14.8 points (SD=2.9), i.e. pass with distinction at this level, suggesting a good to high level of proficiency in English for all of them (Table 1).

Table 1. Group results on the Diagnostic proficiency VOC/MCT test.

Points	Group A	Group B
<10 pts (=F)	-	-
10-12 pts (=P)	5 (31%)	1 (11%)
13-15 pts (=PwD)	3 (19%)	4 (44.5%)
16-18 pts (=PwD)	7 (44%)	4 (44.5%)
19-20 pts (=PwD)	1 (6%)	-
<b>Total</b>	<b>N=16</b>	<b>N=9</b>

Although all participants achieved either a pass (P) or pass with distinction (PwD), it seemed that there was a higher proportion in the top range (Figure 1) in group B, where 89% of the students reached the PwD level, compared to 69% in group A.

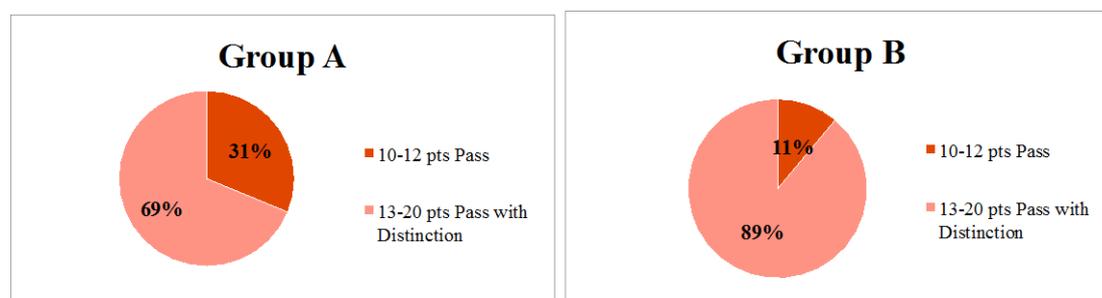


Figure 1. Comparison of diagnostic test results in the two participant groups.

Still, these results suggest a relatively similar level of proficiency in vocabulary size and basic grammatical knowledge.

The two teachers in the study were female university lecturers aged 40-55 with Swedish as their L1s, both with longstanding experience in teaching English writing and grammar courses in school and at university level. One of them was also the researcher. They were coded according to the class they taught (Teacher A and B). Having cooperated for many years, providing feedback in these courses, the teachers also had norming sessions comparing assessment before grading. The posttests were anonymous and marking divided evenly between the two teachers. This was considered to ensure an acceptable level of interrater reliability for this small-scale study.

### 3.2 Material

To ensure enough material to work with, a *pretest* (summary), several *practice texts* (text 1-2-3) and an immediate *posttest* (exam summary) were collected from the participants. Written feedback was provided to all texts except the posttest. Apart from the pretest summary, all practice texts constituted regular course work and thus participation did not affect the workload in any way for students or teachers.

All the data was collected during the first semester of English studies when the students took the academic writing course *Written and oral proficiency* running over

one semester of 20 weeks. All students' texts were coded so that participants could later be matched individually to each text produced.

Prior to studying English, the participants had had one semester of core courses covering pedagogy, subject didactics and practical training in schools. During the second semester, they took English consisting of 3+3 parallel courses of which academic writing was one; the others covered English grammar, literature, culture, phonetics/ pronunciation, and didactics. The students were thus exposed to more reading and writing input than from the academic writing course alone.

In general, everything in the classroom was done in and through English as the target language. The few exceptions relate to these students training to become teachers in the Swedish school system, so some exercises target contrastive differences, at this first level mainly between Swedish and English linguistic areas. One example of this was the summary where the original text was in Swedish and the assignment was to transform it into a summary in good, idiomatic English, using their own words – a type of 'geared free writing'. This text type was targeted throughout the course, facilitating comparison of the collected material over time.

The project was launched at the beginning of the semester in order to make sure the pretest was as 'clean' as possible, i.e. the students' writing skills and proficiency could be tested before they had been given any teaching regarding text type, format and style at this level. The participants were given the pretest before the course began. They were asked to summarize the content of a Swedish article at home and send it in to their group teacher using the university learning platform. Once the writing course began, groups met once every other week for ten 2-hour-classes, over the 20 weeks (Fig. 2).

The same procedure was followed throughout the course: setting the assignment, students writing at home, handing in text, individual feedback from teacher, work in peer response groups, in-class oral general feedback and second round of peer response, revisions (done but not handed back to teacher), new assignment given, thus:

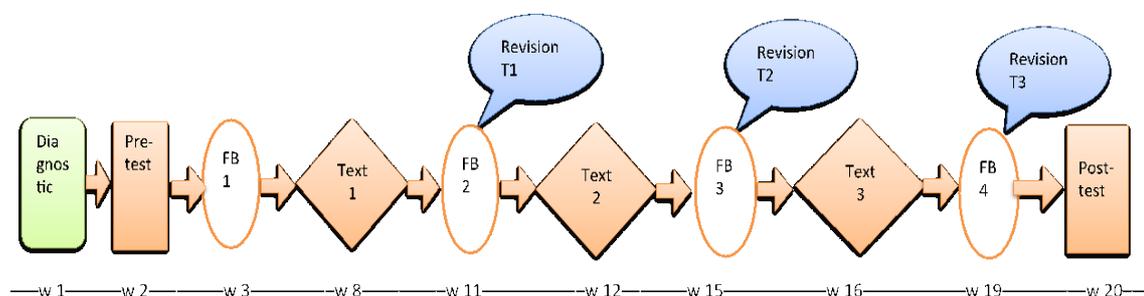


Figure 2. Procedure: collection of material during the autumn 2012

The *pretest* was a full-length summary, started off in class week 2, with minimal information about the genre as such. The writing was finalized at home and the text (approx. 350 words) was sent in two days later. Feedback on this text was provided in week 3 (FB1). Students were expected to revise within two weeks but no further feedback was given to revisions. After this, the class worked with vocabulary and paragraphing, and in week 7 a new assignment was introduced. *Text 1* (the *paragraph*, 50-75 words) was then written at home and sent in a week later (week 8). This text was returned with feedback in week 11 (FB2). *Text 2* was a *mini-summary* (approx. 100 words) started in class week 12, then finished at home and sent in three days later. Feedback was returned after three weeks (FB3). The last practice text, *Text 3*, was a *full summary* (approx. 350 words) written at home and sent in a week later followed by a

final feedback session in week 19 (FB4). After this, the immediate *posttest* (the actual exam) was taken. This was a full summary (approx. 350 words) written under different conditions since it was timed (3hrs) and written by hand in an exam hall.

Revisions were not collected, instead improvement was measured by comparing the pre- and posttests, investigating whether feedback on error areas and strong points in the practice texts would lead to improved *new* texts, reducing errors and strengthening students' writing skills.

Although the Swedish original texts in the pre- and posttest were not identical topic-wise, the texts were comparable regarding level of difficulty and complexity. Writing summaries is a generic skill that should be applicable to any topic and new text. It is the general language use and text structure that are assessed. The original Swedish texts functioned as prompts for students' free writing. This made the set up useful for finding out whether the feedback might have had a favourable impact on the students' proficiency development when producing new texts.

### 3.3 Feedback

In order for the teachers to feel comfortable with the teaching situation, comprehensive feedback was encouraged. This approach most closely resembled the regular teaching practice in the writing courses at the time. It is also supported in research reporting on learners expecting to receive comprehensive feedback, as well as teachers feeling that it would be unprofessional not to give language learners information and comments on all mistakes (e.g. Evans, Hartshorn, McCollum, & Wolfsberger, 2010; Ferris, 2011). Theoretical issues surrounding the corrective feedback debate were investigated by closely considering the pedagogical value of giving comprehensive error correction. It was then decided by the teacher-researcher in collaboration with the other participating teacher that comprehensive, rather than selective feedback focusing on only one or two specified areas, was to be used.

In the study, feedback was provided in three ways: written individual teacher comments, in-class oral teacher feedback, and peer response (in class and in student led workshops). The teachers provided written feedback covering both content and language on each text in the form of marginal comments delivered to students electronically.

When the texts contained an abundance of spelling, vocabulary and stylistic errors, teacher A used in-text colour coding instead of marginal comments. These colour codes were explained to the students early on. Teacher B, instead used a system whereby she carried out some of the suggestions and the system displayed the change made so the student simply had to accept or reject them in the returned document. Thus, the student did not necessarily have to consider or find out about the type of change required, but could simply click OK and the change was done for him/her. This difference between the teachers and the different amount of feedback provided on the pretest were the reasons for keeping the two groups separate throughout the study.

The individual feedback was sent to the students in time to browse through before class, and they were expected to bring this along for classroom work. In class followed a feedback run-through of a more general kind by the group teacher, and then, each student got a clean copy back (no feedback notes) to swop with peers and work in small peer response groups in class before taking on their own text for revision.

The 'two stars and a wish'-model was used for peer response, i.e. the reader wrote two positive things about the text and one suggestion for improvement or clarification before handing it back to the writer. For the practice texts the students were also asked

to work with peer response in their study groups before sending their final versions to the teacher in order to practice giving and receiving response. The writer could then compare the peer response with the teacher feedback brought to class and they were then encouraged to rewrite at home, making use of both. No revisions were returned to the teacher, instead a new assignment was introduced (see Fig. 2). In addition, the students had access to their course literature (Fawcett, 2011, *Evergreen. A Guide to Writing with Readings* and an exercise compendium). This material was referred to in the general feedback in class giving concrete tips on suitable exercises.

None of the practice texts were graded and the teachers emphasized that these texts were for improving their writing skills and becoming independent learners able to monitor their own writing, but also learning how to give and receive feedback themselves. Grades were only given on the final posttest, which was the actual exam paper. However, for research purposes the pretest was unofficially given a grade known to the researcher but never disclosed to the participants.

### 3.4 Analysis of data

To investigate possible links between feedback type/form and the quality of the summaries, the teacher feedback was analysed and compared to the students' linguistic quality of and grades on pre- and posttest. Students' overall development was measured comparing their summaries in pre- and posttests to see if the feedback might have influenced the writing of new texts. Grades (fail–pass–pass with distinction) provided at the end of the course depended on more than just the number of linguistic problems in the texts, it was also about quality and accuracy of content and structure. The summaries were graded by the two teachers, first separately, then together in a norming session to ensure consistency and agreement. This assessment was based on criteria for task and course fulfilment in the curriculum, considering language accuracy, structure and content coverage required in summary writing.

To compare results in pre- and posttest, two-tailed *t*-tests with the alpha level set at .05 were used. However, due to the limited size of the material, it is important to view the results only as tendencies pointing out a direction for further investigation. It is also important to remember that since the participants were both EFL learners and language teachers-to-be, the goal of the course was to develop more effective L2 writers by working with participants' language, but also to improve how they could teach writing themselves, consequently much focus was on language accuracy and text structure.

In the analysis, 'comment' referred to any feedback entry, or part of such an entry, relating to a specific area of grammar, content, structure etc. Therefore, a long comment giving feedback on several areas counted as more than one comment. Thus, an entry like *Word order plus verb form* was counted as two 'comments' since it deals with two areas. It should be noted, however, that most comments were not of this kind. Furthermore, differences in text length were controlled for by calculating the number of comments per 100 words.

A total of 1,927 feedback comments from all texts were analysed. All comments were divided into four more manageable categories; 'text structure', 'verb mastery', 'sentence structure' and 'other areas' (Table 2). These are sometimes jointly labelled 'targeted areas' comprising the following subcategories:

Table 2. Categories of targeted areas used in the analysis.

Category	Content
TEXT structure	paragraph structure, summary format and main content coverage
VERB mastery	subject–verb agreement, tense/aspect usage, verb form (regular vs. irregular form; verb + infinitive/gerund/bare infinitive)
SENTENCE structure	coordination/subordination, word order
OTHER areas	spelling, vocabulary, article usage, punctuation etc

The decision what to include in these categories rests on the researcher’s long teaching experience as to what seems to be the most common problem areas in EFL writing courses. These are also areas that are important for the overall function of texts.

The feedback provided was then categorised as *direct* or *indirect* feedback using Sheen’s (2011) terminology while incorporating positive as well as corrective comments (Table 3).

Table 3. Examples of direct versus indirect written feedback comments.

Direct feedback comments	Indirect feedback comments
a) “ <i>Good coherence through parallel structure with the –ing form: being, seeking out, meeting, being</i> ” B4	e) “ <i>Very good introduction</i> ” B1
b) “ <i>Cut ‘the’...</i> ” A1	f) “ <i>Incorrect expression – check and rephrase</i> ” A1
c) “ <i>No article here</i> ” A13	g) <b>Colour code</b> (blue=incorrect use of contraction) A11
d) “ <i>Inserted: -ly</i> ” B3 (=correction made)	h) “ <i>Check rules for punctuation for coordinated sentences</i> ” B8

Comments of the type found in (b) and (c) have to be acted upon and performed by the student according to suggestion, while (d) only has to be accepted or rejected by clicking the comment online and the correction is made. However, both types were treated as examples of direct feedback since the student did not have to find the solution. Colour-coded ‘comments’ (g) were treated as indirect feedback since they did not give the solution but simply implied what type of error had been made. The students only knew what the colours stood for so these were considered similar to giving a short comment like “*Check vocabulary*”, i.e. indirect metalinguistic WF.

The teachers’ feedback was further analysed using Ferris, Pezone, Tade, & Tinti’s (1997) categories relating to the *intent* of the feedback, i.e. either giving *directives* (ask for information, make a suggestion, inform on mistake) relating to the content, commenting on *grammar/mechanics* (language or text structure) or providing *positive* comments. The comments were also analysed in light of *linguistic features* leaning on Ferris’ (1997) model for syntactic form – *question, statement, exclamation* and *imperative*. Each comment was counted as one if it dealt with what could be reasonably interpreted as one (syntactic) type. As Ferris et al. (1997) point out, this is a delicate and not always easy task. There is of course a risk of being subjective in classifying comments but it is difficult to see how else this could be done but trusting the researcher’s good judgment as long as ample examples of how categorization was generally done are provided.

Comments varied in length, and in cases where they contained more than one clearly delimited clause these were defined as separate comments if they referred to different areas, as in the following example where the first part (in bold) was defined

as *direct* WF and the latter as *indirect* WF, targeting vocabulary and spelling, respectively:

**“First I think you could remove ‘amount of’ since ‘sleep’ can stand on its own – also there’s a spelling mistake in the phrase – can you spot it?”** (A2-1)

Furthermore, looking at the intent of the comment and the linguistic features used, a very long comment like the following was analysed thus:

*“(a) First, check vocabulary use/meaning of ‘mean’, and (b) don’t use ‘folk’. (c) Then I’m not sure you have quite understood what’s said about FB and Google check facts again and (d) also clarify what you mean in the final sentence - the text deals with the change that’s taken place in how people handle FB these days and what they upload, doesn’t it?”* A13-1

Although this comment clearly has two sentences separated by a full stop, the analysis resulted in four ‘parts’ relating to (a) *grammar* (use of ‘mean’), (b) *grammar* (word choice due to style), (c) *directive* (suggestion as to writer’s interpretation of facts), and, finally, (d) *directive* (suggestion as to writer’s misinterpretation of fact).

The *linguistic forms* of this comment were marked as (a) imperative, (b) imperative, (c) statement + imperative, and (d) imperative + question. Even a short comment like *Incorrect expression – check and rephrase* (A3) would have to be treated as one comment using *indirect* WF, treating one area ‘Other’, intending to achieve a change in language (*grammar*), rather than content but using two linguistic forms (*statement* and *imperative*) – it would be problematic to do it differently. This means that one feedback comment could deal with one or several areas (in grammar, structure, content), and this could be expressed using several linguistic features.

Thus, applying these models for analyses at several different levels, comments were sometimes treated and counted as one or several parts, and therefore the sums between different kinds of analyses did not always match up.

#### 4. Findings and discussion

The main research questions dealt with whether feedback would help develop students’ writing from pre- to posttest, and if there was a link to any particular type of feedback provided. This was measured first by comparing results in the pre- and posttest followed up by investigating the feedback from several perspectives.

In total, 104 texts were collected; 25 pretests, 54 practice texts and 25 posttests. For the posttest (i.e. the exam), grades were given on a three-point scale; fail (F), pass (P) and pass with distinction (PwD). Of the 25 participants, 56% (N=14) improved their grade by one or two steps (from F to P, F to PwD, P to PwD). The remaining 44% (N=11) either retained their original grade or, in one case, did worse going from pass to fail in the posttest (Table 4).

Table 4. Assignments handed in and grades for pre- and posttest.

Improved grade					
Student	Pretest grade	Practice texts done			Post-test grade
		Para-graph	Mini Sum	Full Sum	
A2	F	1	1	1	PwD
A6	F	1	1	0	PwD
A16	F	1	1	1	PwD
B7	F	1	1	1	PwD
A15	F	1	1	0	P
A18	F	1	0	1	P
A21	F	0	0	0	P
B3	F	1	1	1	P
B4	F	1	0	0	P
B8	F	0	1	1	P
B11	F	1	1	1	P
A3	P	1	0	1	PwD
B1	P	0	1	1	PwD
B6	P	1	1	1	PwD
<b>total</b>	<b>14</b>	<b>11</b>	<b>10</b>	<b>10</b>	<b>14</b>

No change (or lower) grade					
Student	Pretest grade	Practice texts done			Post-test grade
		Para-graph	Mini Sum	Full Sum	
A4	PwD	1	1	1	PwD
A9	P	1	1	1	P
A11	P	0	0	1	P
A17	P	1	0	1	P
B5	P	1	1	1	P
A1	F	0	1	1	F
A10	F	0	1	0	F
A13	F	1	1	1	F
A14	F	1	1	1	F
B10	F	1	1	0	F
A5	P	0	0	0	F
<b>Total</b>	<b>11</b>	<b>7</b>	<b>8</b>	<b>8</b>	<b>11</b>

F = fail, P = pass, PwD = pass with distinction

Taking a closer look at those who improved (N=14) in the posttest, only 44% (N=7) were from group A but 78% (N=7) from group B (Fig. 3). (One student in group A actually did worse in the posttest than in the pretest.

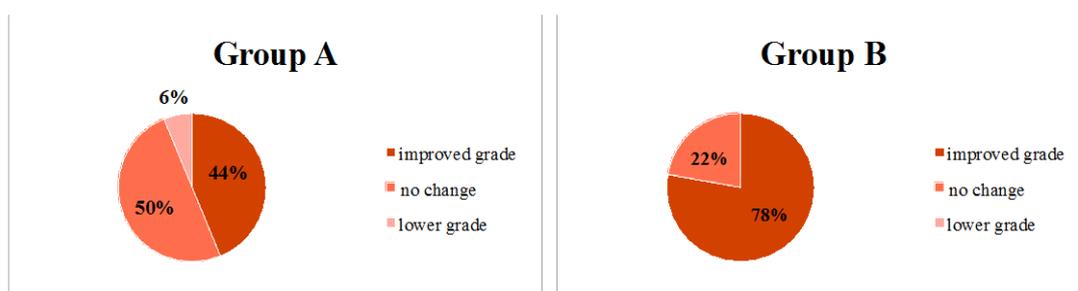


Figure 3. Grades development from pre- to posttest for students in group A and B.

This result could depend on students' proficiency levels, however, the diagnostic test indicated that the groups were very similar (A = mean 14.7 pts, B = mean 15.1 pts). A majority of the participants, 20 of 25 (80%), handed in at least two of the practice texts apart from the pretest, and received feedback on these. Thus, the number of texts written were basically the same in both groups, offering all students the same opportunity to receive feedback. The average number of words/student text in the pretest was 337 (M=337, SD= 31.7) and in the posttest 350 (M=350, SD=47.9).

First, pre- and posttests were compared to monitor any change in results. Unless they were clearly positive comments, all feedback comments were considered an indication of problems or errors in need of attention to improve the text(s). In the pretest, there were a total of 625 WF comments, i.e. 7.5 comments/100 words, having decreased to 6.5 in the posttest (Table 5).

Although the overall results indicated fewer comments/100 words in the posttest (pointing to a decrease in problems) compared to the pretest, this reduction was not significant at the .05 level, nor did it necessarily result in improved final grades. For

instance, a couple of students had fewer comments in the posttest, however, the final grade remained unchanged. This suggests that the grades provided at the end of the course depended on more than just the number of problems in the texts, it was also a question of quality and accuracy of content and structure. However, in this study, the number of comments/100 words was used as a way to measure what happened between the pre- and posttest in order to provide an overview of whether feedback might have affected the writing.

Table 5. Difference in number of WF comments/100 words in pre- and posttest.

Student	PRETEST				POSTTEST				Diff. Comments /100 words**
	Words / text	Total comments	Comments /100 words	Grade*	Words / text	Total comments	Comments/ 100 words	Grade*	
A1	354	23	6.5	F	348	46	13.2	F	+6.7
A2	297	16	5.4	F	339	15	4.4	PwD	-1.0
A3	379	6	1.6	P-	499	22	4.4	PwD	+2.8
A4	263	6	2.3	PwD	399	10	2.5	PwD	+0.2
A5	311	7	2.3	P-	423	25	5.9	F	+3.7
A6	365	14	3.8	F	362	9	2.5	PwD	-1.3
A9	342	19	5.6	P-	335	22	6.6	P	+1.0
A10	294	34	11.6	F	357	72	20.2	F	+8.6
A11	317	24	7.6	P-	356	32	9.0	P	+1.4
A13	362	22	6.1	F	320	24	7.5	F	+1.4
A14	347	23	6.6	F	372	32	8.6	F	+2.0
A15	378	15	4.0	F	271	17	6.3	P	+2.3
A16	319	23	7.2	F	300	16	5.3	PwD	-1.9
A17	297	5	1.7	P-	351	27	7.7	P	+6.0
A18	357	39	10.9	F	351	31	8.8	P	-2.1
A21	344	23	6.7	F	350	12	3.4	P	-3.3
B1	312	12	3.9	P-	341	10	2.9	PwD	-0.9
B3	349	38	10.9	F	350	26	7.4	P	-3.4
B4	320	38	11.9	F	335	20	6.0	P	-5.9
B5	291	19	6.5	P-	235	11	4.7	P	-1.8
B6	330	41	12.4	P-	340	13	3.8	PwD	-8.6
B7	339	30	8.9	F	337	21	6.2	PwD	-2.6
B8	365	50	13.7	F	357	15	4.2	P	-9.5
B10	339	53	15.6	F	374	30	8.0	F	-7.6
B11	359	45	12.5	F	348	14	4.0	P	-8.5
<b>Total</b>	<b>8330</b>	<b>625</b>	<b>7.5</b>		<b>8750</b>	<b>572</b>	<b>6.5</b>		<b>-1.0</b>

\* Abbreviations: Fail = F, pass = P and pass with distinction = PwD

\*\* Shows the difference in number of feedback comments/100 words from pre- to posttest

In the posttest, group A received more comments on average ( $M=7.3$ ) than group B ( $M=5.3$ ), and although this difference was not statistically significant ( $p > .05$ ), it is still worth keeping in mind (Table 6) as it could suggest more problems than in the other group. Comparing the amount of WF provided in the practice texts, it was clear that, on average, group A received significantly ( $p < .05$ ) less feedback. The difference was noticeable comparing posttest results, where those who did better had received significantly fewer WF comments ( $p = 0.01$ ). This could indicate fewer problems. Since 78% from group B and 44% from group A improved maybe the feedback provided in group B worked better.

In the practice texts, group B received significantly more WF ( $p < .05$ ) compared to group A, suggesting they had more problems to begin with, but these had been taken care of by the time of the posttest, possibly pointed out through feedback. It was beyond the scope of this study to investigate the type of errors in detail, instead it focused on the types of feedback to try to find a pattern and perhaps suggestions for providing feedback.

Table 6. The amount of WF comments/100 words in pretest summaries and all practice texts compared to posttest summaries.

Compared groups	Feedback comments/100 words								
	Pretest	Para-graph	Mini Sum	Full Sum	Total		Posttest		
					<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Group A (N=16)	5.6	6.7	5.8	5.3	5.9	0.6	7.2	7.3	4.5
Group B (N=9)	10.8	7.2	11.0	6.6	8.9	2.3	5.4	5.3	1.9
Difference between group A and group B					<i>p</i> = 0.04		<i>p</i> = 0.23		
1) Improved (N=14)	8.1	7.9	9.3	5.6	7.7	1.6	4.9	5.0	1.8
2) No change (N=11)	6.6	5.4	6.4	6.1	6.1	0.5	8.6	8.6	4.8
Difference between 1) Improved and 2) No change					<i>p</i> = 0.62		<i>p</i> = 0.01		

The students who improved had received more comments in all practice texts ( $M=7.7$ ) except the full summary, compared to those who showed no change. Although not statistically significant ( $p > .05$ ), this difference could be an indication that more WF in the practice texts increased the chance of it having an effect on new writing in the posttest.

A closer look at areas targeted by the feedback revealed that six of nine subcategories had fewer comments in the posttest. As fewer WF comments were taken as a sign of improvement, the remaining categories, *subject-verb agreement*, *tense/aspect* and *other areas* (Table 7), seemed to pose continuous problems.

Table 7. The amount of WF comments in all written summaries in pre- and posttests distributed over target areas of analysis.

Target areas	Feedback comments/100 words			
	Pretest		Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
TEXT STRUCTURE: paragraphing	0.8	1.2	0.4	0.7
TEXT STRUCTURE: summary format	3.9	3.0	1.7	1.4
TEXT STRUCTURE: content coverage	1.7	1.5	1.0	0.9
VERB MASTERY: subject-verb agreement	1.5	2.4	1.5	1.9
VERB MASTERY: tense/aspect	0.5	0.8	0.5	1.2
VERB MASTERY: verb form	0.4	0.7	0.2	0.7
SENTENCE STRUCTURE: coord./subord.	0.7	1.0	0.4	0.8
SENTENCE STRUCTURE: word order	0.3	0.5	0.1	0.3
OTHER areas*	15.3	11.6	17.0	10.8

\* 'Other areas' subcategories include spelling, vocabulary and preposition choice, article usage, contracted form used, pronoun choice, genitive form.

Although all areas were targeted through comprehensive feedback, it seemed the impact was strongest for 'Text structure', where the most noticeable effect appeared in *summary format* and *content coverage*. In fact, these were the only subcategories with a result significant at the .05 level. This is not surprising since these areas match the main focus of the course. Minor improvements (not statistically significant) also appeared in 'Sentence structure' where problems with *coordination/subordination* and *word order* decreased somewhat. In 'Verb mastery', only *verb form* showed a small improvement (again not significant at .05). Many students still did not master verb agreement and tense in the posttest, where, individually, they either showed no change

at all or made even more mistakes than in the pretest. These areas commonly create contrastive problems for EFL learners, therefore they are regularly focused in writing classes. There was also an increase, albeit not significant, in comments targeting problems in the ‘Other areas’ category. A more detailed look revealed that more than half of these involved *vocabulary choice* and *spelling*.

In both teaching groups, most feedback comments targeted ‘Text structure’ and ‘Other areas’. However, group B was offered significantly more feedback than group A in both of these areas (Table 8).

Table 8. The amount of WF comments/100 words for targeted areas in the study groups and among those who improved or not in the posttest.

Compared groups	Feedback comments/100 words							
	Text structure		Verb mastery		Sentence structure		Other areas	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Group A (N=16)	1.7	0.6	0.7	0.6	0.4	0.3	3.0	1.7
Group B (N=9)	2.4	0.6	0.4	0.3	0.3	0.2	6.0	1.2
Difference between group A and group B	<i>p</i> = 0.01		<i>p</i> = 0.10		<i>p</i> = 0.32		<i>p</i> = 0.00	
Improved (N=14)	2.0	0.5	0.5	0.3	0.3	0.2	4.6	2.0
No change (N=11)	1.9	0.8	0.7	0.7	0.3	0.3	3.4	2.2
Difference between 1) Improved and 2) No change	<i>p</i> = 0.79		<i>p</i> = 0.25		<i>p</i> = 0.68		<i>p</i> = 0.18	

There were minor differences between those who improved and those who did not. In all areas, except ‘Verb mastery’, those with better results received more WF, although this is not statistically significant. Thus, it seemed that the targeted areas were the same irrespective of group.

Moving on to the intent of the WF, a clear majority of comments, 75% (N=1047), dealt with *grammar/mechanics*. A closer look at the entries revealed that these were comments on language errors that needed attention to improve the text(s). These were followed by *positive* remarks (14%, N=190), such as ‘*Good summary*’ or ‘*Neat paragraph*’, and *directives* about content (11%, N=153). There were clear contrasts between the teaching groups (Table 9). Group B received more comments on *grammar/mechanics*, a statistically very significant (*p* = 0.01) difference. The same group was also given significantly (*p* = 0.00) more *positive comments*, e.g. *Good paragraphing*, than the A group. These comments generally focused on text-type format and content.

Table 9. The amount of WF comments/100 words regarding the intent of comments in all practice texts distributed over study groups and those who improved or not in the posttest.

Compared groups	Directives		Grammar/Mechanics		Positive comments	
	M	SD	M	SD	M	SD
Group A (N=16)	0.8	0.4	4.6	2.2	0.7	0.4
Group B (N=9)	0.9	0.6	7.8	3.9	1.4	0.4
Difference between group A and group B	$p = 0.57$		$p = 0.01$		$p = 0.00$	
Improved (N=14)	0.6	0.4	5.7	1.8	1.0	0.5
No change (N=11)	1.1	0.5	5.8	4.6	0.8	0.6
Difference between 1) Improved and 2) No change	$p = 0.02$		$p = 0.99$		$p = 0.35$	

Although not statistically significant, *grammar/mechanics* comments were slightly more common for the group that did not improve. In this group, *directives* were also more common and this difference is significant ( $p = 0.02$ ). *Positive comments* were more frequent for those who improved, and although not statistically significant, this could indicate that maybe receiving more positive input than *directives* during the writing process might impact end results favourably. On the whole, *directives* (focusing content and organisation of ideas) were the least used type. A likely reason could be that language accuracy, covered through *grammar/mechanics*, was given a higher priority in this first writing course. Some research has suggested that *directives* do not affect revisions and might even have a detrimental effect on “students’ thinking” (Bruno & Santos, 2010; Goldstein, 2004). Ferris (1997), on the other hand, found the opposite. However, none of these refer to the effect on new texts as the present study does. Ferris (1997) also noted *positive comments* not resulting in any improvement, and Hattie and Timperley (2007) suggest that the focus of *positive comments* matters – personal comments (*Good girl!*) are not likely to enhance performance of the task at hand. In the material, *positive WF* generally focused on text structure, e.g. *Good title* or *Excellent paragraph* with or without further explanation, so maybe these were helpful.

If the areas targeted and the intent of WF seemed to be less influential on results, maybe the way it was delivered held the answer. In the practice texts, 61% (N=827) of the feedback comments were *indirect*, either in the form of very general positive comments without explanation as to why the marked part was good, or metalinguistic clues or codes referring to a specific type of error. The remaining comments, 39% (N=528), were *direct WF*, either positive comments clearly explaining what was good, or giving the correct solution in the comment or by deletion/insertion directly in the text.

Group A received significantly more *indirect* than *direct WF* ( $p < 0.00$ ). The B group was instead provided with significantly ( $p < 0.00$ ) more *direct WF*. Since only 44% from group A, but 78% from group B did better in the posttest, this suggests that maybe *direct WF* provided a better chance to improve (Table 10).

Table 10. The amount of direct and indirect WF comments/100 words in all practice texts distributed over study groups and those who improved or not in the posttest.

Compared groups	Direct WF comments/100 words		Indirect WF comments/100 words		Comparison within groups
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Group A (N=16)	0.8	0.5	5.0	2.1	<i>p</i> = 0.00
Group B (N=9)	6.1	2.1	4.0	2.4	<i>p</i> = 0.00
Difference between group A and group B	<i>p</i> < 0.0001		<i>p</i> = 0.3138		
Improved (N=14)	3.2	2.6	4.2	1.5	<i>p</i> = 0.33
No change (N=11)	2.1	3.3	5.2	2.8	<i>p</i> = 0.01
Difference between 1) Improved and 2) No change	<i>p</i> = 0.35		<i>p</i> = 0.25		

Although more *direct* WF was given to those who improved compared to those who did not, this difference was not statistically significant ( $p = 0.35$ ). In fact, both groups received more *indirect* than *direct* WF, however, the difference is only significant for those who did not improve ( $p = 0.01$ ). Since there are studies suggesting that learners might have problems with indirect feedback as it might be misunderstood (Parr & Timperley, 2010) this could explain why they did not do better. However, it is not possible to statistically link either of these types to the success of some students and not others. If anything, maybe a balance between the types feedback is to be recommended. Research seems to point in that direction, e.g. Chandler (2003) suggesting that *direct* WF is useful for short-term correction, and van Beuningen et al. (2012) proposing it is suitable for grammatical problems, while *indirect* WF is better for non-grammatical problems. So, it seems the choice between explicit or implicit feedback to develop writing is less than clear-cut.

Next, focus was turned to linguistic form of the feedback. Most of the comments, 34% (N=566), were *statements* followed by *deletion/insertion* at 20% (N=330). After these, comments were fairly evenly distributed between *imperatives* 17% (N=280), *questions* 15% (N=239) and *colour codes* 9% (N=156), followed by *exclamations* 3% (N=43) and *abbreviations* 2% (N=38) last.

Although the overall pattern in groups A and B coincided, there were two noticeable differences. In group A, significantly more *questions* ( $p = 0.00$ ) and *colour codes* ( $p = 0.03$ ) were used than for group B. Another very significant difference was that group B received many more *statements* ( $p = 0.01$ ) and much more *deletion/insertion* ( $p < 0.00$ ) than group A. For the remaining categories, there was no statistically significant difference between the two teaching groups (Table 11).

Table 11. The amount of different linguistic forms of WF comments/100 words in all practice texts distributed over study groups and those who improved or not in the posttest.

Compared groups	Question		Statement		Exclamation		Imperative		Colour code		Abbreviation		Delete/Insert	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Group A (N=16)	1.8	1.0	2.3	1.2	0.2	0.2	1.5	0.7	1.4	1.7	0.2	0.2	0.1	0.1
Group B (N=9)	0.3	0.2	4.0	2.0	0.3	0.3	1.8	1.8	0.1	0.1	0.1	0.4	4.8	1.7
Difference between group A and group B	$p = 0.00$		$p = 0.01$		$p = 0.16$		$p = 0.62$		$p = 0.03$		$p = 0.59$		$p < 0.00$	
Improved (N=14)	0.9	0.9	2.8	1.2	0.2	0.2	1.2	0.5	1.1	1.8	0.2	0.2	2.3	2.3
No change (N=11)	1.7	1.1	1.7	1.1	0.3	0.3	2.1	1.5	0.8	1.0	0.4	0.2	1.1	2.7
Difference between 1) Improved and 2) No change	$p = 0.04$		$p = 0.80$		$p = 0.23$		$p = 0.05$		$p = 0.68$		$p = 0.64$		$p = 0.23$	

For those who improved their grades, the order of occurrence of the types of comments was more or less identical to that of the whole cohort. Although the use of *colour codes* and *deletion/insertion* were more frequent in this group, the difference was not statistically significant for either type. Instead, significantly fewer *questions* ( $p = 0.04$ ) and *imperatives* ( $p = 0.05$ ) were used compared to those who did not improve, where these types were instead clearly more frequent. Although the use of *questions* has been encouraged in different studies (Ferris, 2014) they can also be problematic since students may find them confusing and fail to recognize them as triggers to improve the text. Rhetorical and indirect questions might be misinterpreted and the learner might try to answer them instead (Bruno & Santos, 2010; Ferris, 2007; Lea & Street, 1998). In other studies, *imperatives* seem to have been more effective for revisions than *questions* and *statements* (Mehr, 2013; Sugita, 2006). Perhaps the use of more questions in the A group explains why fewer students from that group improved in the posttest.

It is worth noting, however, that the use of *deletion/insertion* was extremely significant in the B group, and 78% from group B improved in the posttest. This seems to suggest that there could be a link between direct correction of language problems targeted by this type of feedback and results. Previous research has shown that students receiving explicit teacher corrections made fewer errors in revisions, especially regarding ‘treatable’ areas although it might yield better short-term than long-term effects (e.g. Bitchener et al., 2005; Chandler, 2003; Ferris, 2003, 2014). The areas targeted here are mainly ‘treatable’ types (Bitchener et al., 2005; Ferris, 1999) from the ‘Other areas’ category.

## 5. Concluding remarks

The aim of this small-scale study was to investigate whether there would be a discernible improvement in writing after receiving feedback, and if such improvement could be linked to certain types of feedback provided. To do this, written work from two groups of student teachers in an academic EFL writing course, and the teachers’ written feedback to the students, were collected over a 20-week period. All in all, 56% of the participants had improved by the end of the course, suggesting the feedback may have had an effect at least for some of the students, although this general improvement was not statistically significant at alfa level .05.

Results from the initial diagnostic test suggested that the groups were similar in language competence from the beginning, and this seems to be supported by looking at the error areas targeted in the pretest which were very similar, too. So, the two original

groups could be seen as comparable and similar regarding their proficiency level. The distribution of feedback comments on the four targeted areas was also similar for those who improved and those who did not; the most frequently commented areas were 'other areas' and 'text structure'.

Focusing on the participants with improved results from pretest to posttest, the findings showed that they received more feedback comments in general on their practice texts. The intent of these comments was more frequently to point out language mistakes in need of attention, i.e. mainly grammar/mechanics comments, rather than giving directives on the content.

More direct WF (although not statistically significant) was provided those who improved compared to those who did not - but the interesting point is that within the improved group there was a more even balance between direct and indirect feedback – some research suggests that a balance between these two types is better since they may impact different types of errors areas. Direct feedback seems to work better for grammatical errors, while indirect is better for non-grammatical errors. Here more research into what areas were targeted by the respective feedback type is needed.

Looking into linguistic form of the direct WF comments, colour codes and deletion/insertion were more frequently used in the group who performed better on the posttest. The latter category seems to have been used for treatable errors, i.e. errors where language rules are involved. These results coincide with previous studies showing similar results in favour of direct feedback having yielded positive results at least in a short-term perspective. Therefore, it would be interesting to further investigate possible long-term effects of direct versus indirect feedback.

There were also significantly fewer questions and more statements used in the WF for those who improved. This suggests a clearer and more direct approach to what needs to be done. The use of questions has been discussed as being problematic with EFL learners specifically, because they often mistakenly interpret questions as being 'real' questions and therefore optional, rather than coercive, albeit polite, pointers indicating required attention to errors. Thus, no action is taken. If the feedback is not clearly understood, how are students supposed to be able to act upon it? This could, perhaps, partly explain why the students receiving more questions did not succeed as well as the others.

For those who did not improve, the intent of the feedback was significantly more often directives, generally asking for or giving information about the content. The feedback provided was also more indirect for this group, and they were given significantly more comments in the form of questions and imperatives than the group who improved.

These diverse findings have raised further questions about the possible impact of the type and form of feedback given, whether it was appropriate or not, and if and how it was actually used and understood by the learners (e.g. McMartin-Miller, 2014; Nicol & MacFarlane-Dick, 2006; Weaver, 2006). If the feedback did not make a lasting enough impression and subsequently did not help minimizing errors for some students - maybe it was not clear enough, or there was too much of it, or perhaps too little. Further investigation of learners' perception and interpretation of feedback could perhaps give a better idea about their use of it. It could also provide a better understanding of how teachers could direct their WF to optimize the learning effect.

It seems that it might also be worthwhile trying to work more intensively with fewer selected areas where it is more difficult for the students themselves to identify and remedy the mistake without feedback from the teacher. The findings of this study suggest that teachers should not avoid using explicit correction, at least for short-term

purposes. Instead, if used pedagogically, e.g. by combining it with advice on individually suited exercises (found in a course book, an exercise compendium, or handed out by the teacher as and when needed), the ‘other areas’ group errors could very well be dealt with through direct feedback. These are advanced learners, so this would be in line with Sheen’s (2007) suggestions to make sure direct correction is accompanied by meta-linguistic comments to raise language awareness in these areas. It also seems that it would make sense to try to give slightly more selective feedback to certain areas where errors seem to be most persistent. It is important to explore how teachers could use time efficiently without missing out on the most important feedback areas – focused feedback within given typical L2 ‘problem’ areas that impede progress might be a way forward without teachers feeling that they let their students down by not commenting on everything (cf. Ferris, 2010).

The results from this small-scale study point out tendencies suggesting there is a need for further research using a larger material looking at feedback type(s) and form in combination with monitoring how students perceive and make use of feedback. This should make for a better understanding of student uptake over time.

## References

- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102-118.
- Bitchener, J., & Ferris, D. R. (2012). *Written Corrective Feedback in Second Language Acquisition and Writing*. London: Routledge.
- Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten-month study. *Applied Linguistics*, 31(2), 193-214.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191-205.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267-296.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36(3), 353-371.
- Eslami, E. (2014). The Effects of Direct and Indirect Corrective Feedback Techniques on EFL Students' Writing. *Procedia - Social and Behavioral Sciences*, 98, 445-452.
- Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing Corrective Feedback in Second Language Writing Pedagogy. *Language Teaching Research*, 14(4), 445-463.
- Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31(2), 315-339.
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1-11.
- Ferris, D. R. (2003). *Response to student writing. Implications for Second-Language Students*. Mahwah: Lawrence Erlbaum.
- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In Hyland, K., & Hyland, F. *Feedback in L2 writing: Contexts and issues* (pp. 81-104). Cambridge: Cambridge University Press.

- Ferris, D. R. (2007). Preparing teachers to respond to student writing. *Journal of Second Language Writing*, 16(3), 165–193.
- Ferris, D. R. (2010). Second Language Writing Research and Written Corrective Feedback in SLA. Intersections and Practical Applications. *Studies in Second Language Acquisition*, 32(2), 181-201.
- Ferris, D. R. (2011). *Treatment of Error in Second Language Student Writing* (2nd ed). Ann Arbor: University of Michigan Press.
- Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, 19, 6-23.
- Ferris, D. R., Pezone, S., Tade, C. R., & Tinki, S. (1997). Teacher commentary on student writing: Descriptions & implications. *Journal of Second Language Writing*, 6(2), 155–182.
- Gascoigne, C. (2004). Examining the Effect of Feedback in Beginning L2 Composition. *Foreign Language Annals*, 37(1), 71-76.
- Goldstein, L. (2004). Feedback and revision in second language writing: Contextual, teacher and student variables. In K. Hyland & F. Hyland (Eds.), *Feedback in Second Language Writing: Contexts and Issues* (pp. 185-205). Cambridge: Cambridge University Press
- Gu nette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16(1), 40-53.
- Gu nette, D. (2012). The Pedagogy of Error Correction: Surviving the Written Corrective Feedback Challenge. *TESL Canada Journal*, 30(1), 117-126.
- Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, 44(1), 84-109.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J. & Timperley, H. S. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Hounsell, D. (2007). Towards more sustainable feedback to students. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment in Higher Education. Learning for the Longer Term* (pp. 101-113). London: Routledge,
- Hyland, F. (2010). Future Directions in Feedback on Second Language Writing: Overview and Research Agenda. *International Journal of English Studies*, 10(2), 171-182.
- Hyland, F., & Hyland, K. (2001). Sugaring the pill Praise and criticism in written feedback. *Journal of Second Language Writing*, 10(3), 185–212.
- Hyland, K., & Hyland, F. (2006). *Feedback in Second Language Writing: Contexts and Issues*. Cambridge: Cambridge University Press.
- J nsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63–76.
- Lea, M. R., & Street, B. V. (2000). Student writing and staff feedback in higher education: an academic literacies approach. In M. Lea & B. Stierer (Eds.) (2000) *Student writing in higher education: new contexts* (pp. 32-46). Milton Keynes, UK/Philadelphia, USA: Open University Press/Society for Research into Higher Education.
- Lee, I. (2011). Feedback revolution: what gets in the way? *ELT Journal*, 65(1), 1-12.
- Lee, I. (2014). Feedback in writing: Issues and challenges. *Assessing Writing*, 19, 1-5.
- McMartin-Miller, C. (2014). How much feedback is enough?: Instructor practices and

- student attitudes toward error treatment in second language writing. *Assessing Writing*, 19, 24-35.
- Mehr, S. H. (2013). The effects of statements, imperatives, and questions as teacher's comments on the revision of composition of Iranian EFL learners. *International Journal of Language Learning and Applied Linguistics World*, 4(3), 1-12.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning, and student progress. *Assessing Writing*, 15(2), 68-85.
- Sadler, D.R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255-283.
- Sheen, Y. (2011). *Corrective Feedback, Individual Differences and Second Language Learning*. New York: Springer
- Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37(4), 556-569.
- Sugita, Y. (2006). The impact of teachers' comment types on students' revision. *ELT Journal*, 60(1), 34-41.
- Truscott, J. (1996). The Case Against Grammar Correction in L2 Writing Classes. *Language Learning*, 46(2), 337-369.
- Truscott, J., & Hsu, A.Y-P. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17(4), 292-305.
- van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners' written accuracy. FGw: Amsterdam Center for Language and Communication (ACLCLC). Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA), 279-296.
- van Beuningen, C., De Jong, N., & Kuiken, F. (2012). Evidence on the Effectiveness of Comprehensive Error Correction in Second Language Writing. *Language Learning*, 62(1), 1-41.
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and Evaluation in Higher Education*, 31(3), 379-394. Retrieved from: <http://dx.doi.org/10.1080/02602930500353061>
- Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assessment & Evaluation in Higher Education*, 35(5), 519-533.