

Grading, no longer an obstacle to learners' attendance to teacher feedback

Majid Nemati

(Assistant Professor, University of Tehran, Islamic Republic of Iran)

nematim@ut.ac.ir

Masoud Azizi

(PhD Candidate, University of Tehran, Islamic Republic of Iran)

m.azizi@ut.ac.ir

(Received: 10.04.2013, Accepted: 28.05.2013)

Abstract

Learners are often reported not to be motivated enough to attend to teacher feedback. Teachers also tend to grade learners' writing samples when providing them with corrective feedback though they know it may divert their attention away from teacher feedback. However, not grading learner writings does not seem to be an option due to both learners' demands for it and institutional regulations that require teachers to have summative evaluation. In order to overcome such contradictions, a new technique called Draft-Specific Scoring (DSS) was devised in order to use grading as a motivating, rather than demotivating, device in order to encourage learners to attend to teacher feedback and apply it to their first drafts to improve the quality of their writing accordingly. DSS is a grading system in which learners can improve their received grade by applying teacher feedback to their writing samples in order to improve its quality. The score they receive will improve as a result of the improvement in the quality of the revisions they make. They have two opportunities to go through this procedure. Their final score will be the mean score of all the grades they receive in their last drafts submitted. This experimental study was an attempt to check the effect of the use of this technique in error feedback provision on three measures of fluency, grammatical complexity, and accuracy. The results showed that DSS could help learners improve in all three measures while the control group receiving only error feedback without DSS could only improve in fluency.

Keywords: Draft-Specific Scoring, corrective feedback, writing, Fluency, grammatical complexity, accuracy

Introduction

The effectiveness of teacher feedback has been so controversial that the majority of the publications in L2 writing have been devoted to this subject for the past two decades. While some scholars, the most prominent of whom Truscott (1996), argue against grammar correction and believe that it does not help learners improve, others such as Ferris (1999) and Chandler (2003) argue *for* the practice. The literature on the subject is full of studies in support of both

parties making it impossible to come up with a definite answer.

However, no matter what literature says about its effectiveness or ineffectiveness, students demand teacher feedback because they believe it is necessary and helps them improve (Lee, 2008). Surface-level errors are so important to learners that ESL teachers may lose their credibility among learners if they do not correct all such errors in their students' writings (Radacki & Swales, 1998). ESL students were reported

to believe that a good writing is one which is error-free (Leki, 1990). Also, surveys regarding students' attitudes toward feedback in ESL context (e.g., Ferris, 1995; Satio, 1994) and EFL context (e.g., Diab, 2005; Enginalar, 1993) indicate that learners are concerned about accuracy, and to them, an effective feedback is the one in which teachers pay attention to linguistic errors. The present study was developed as a response to such demands while minimizing the obstacles in the way.

The grammar correction debate so far

After Truscott questioned the effectiveness of grammar feedback in 1996, there has been a very hot debate among scholars and researchers regarding the effectiveness or ineffectiveness of providing students with error correction. This debate has been mainly between Truscott (1996) on the one hand, and Ferris (1999), Chandler (2003), and Bruton (2009) on the other hand.

Truscott (1996) argues that most writing regarding corrective feedback has simply taken the value of grammar correction for granted. All practitioners practice it because they assume it is effective. Moreover, the side effects of such a practice, like its effect on learners' attitude and the energy and time it consumes in writing classes, are often neglected. He cites Cohen's review in which he had concluded that L1 students often pay no attention to corrections. Even if motivated enough to look at and understand the corrections, students may still not be motivated enough to incorporate them in their future writing. Truscott also argues that the students who do try to write in accordance with the feedback they receive may not do so for long, and as soon as they leave that particular class or write in a different context for a different teacher with different concerns, they may ignore the original advice.

Truscott believes that grammar correction is harmful. Relying on research carried out in L1, he argues that students who do not receive corrections have a more positive attitude toward writing. They may not be better writers in comparison with those receiving corrections, but they have been observed writing more. He claims that even in L2, grammar correction has harmful effects. He believes that it is so because of the "inherent unpleasantness of correction." They do not learn as well as uncorrected students do because they shorten and simplify their writing in order not to be corrected (Truscott, 1996, p. 355).

Ferris (1999), responding to Truscott's (1996) review of the research on grammar correction, regards Truscott's conclusion that grammar correction has no place in writing instruction and it should be abandoned as "premature and overly strong" (p. 2). Unlike Truscott, Ferris believes that, if not all, many students can improve their writing as a result of appropriate teacher feedback, so instead of abandoning such a practice, she believes that we should make our corrections more effective. In her opinion, the individual student variables affecting their willingness and ability to benefit from teacher feedback need to be explored. Also, one needs to investigate which methods or techniques in corrective feedback provision can lead to short term and long term student improvement. Only when these variables are explored enough, one can decide on the effectiveness or ineffectiveness of grammar correction.

Chandler (2003) did a thorough study on the efficacy of various types of error feedback and their influence on students' fluency and accuracy in writing. The two groups were found similar in error rates prior to the study. On the other hand, the experimental group's change was statistically significant

at the end of the instruction. Regarding fluency, both groups significantly improved over the 10 weeks between the first and fifth assignments, and they did not differ from each other over the semester. Chandler (2004) believes that although she did not calculate any measure of syntactic complexity, the results of her holistic rating are an indication, not proof, that the writings did not become simpler. The study by Robb Ross, and Shortreed (1986), who did have a measure of syntactic complexity, also showed that all of their groups receiving corrective feedback improved in syntactic complexity.

Truscott (2007) did a meta-analysis on corrective feedback. He found a positive effect for corrective feedback in uncontrolled studies, which he attributed to either bias in the setting of testing or the use of avoidance strategy by learners. He believes that corrected students write shorter and simpler texts in order to avoid making mistakes. As such, even the observed improvement in accuracy may be due to learning how to avoid structures about which they are not sure.

Bruton (2009), looking at the research and argument in error correction, questions Truscott's anti-correction position by drawing three basic conclusions: first, he believes that research into this topic should recognize that "language focus in L2 writing should be seen within a framework of pedagogical options, including minimally differing pedagogical purposes, writer goals and writing tasks, in relation to writer characteristics and context" (p.600). Second, the effect of language focus in L2 writing should not be limited to the issue of grammatical accuracy. Third, even in such a limited view, common sense and intuition defies that correction is harmful to developing accuracy and lack of correction

or simply more writing practice can result in improvement.

Bruton (2009) views the ongoing debate about correction as a "rather tedious sterile academic debate" which has damaged the field by giving researchers a narrow perspective and line of attention. Truscott (2010) objected to and rejected this view; however, Bruton (2010, p. 491) insists on his position and explains that he does not mean that the issue of grammar correction in L2 writing is not important or is less important than it was in the past; however, "the debate is tedious because the same points are reiterated; it is sterile because most of the research central to the argumentation against correction remains the same, with the numerous recognized flaws...; it is academic in the sense that it does not really have much relevance for most mainstream L2 writing contexts or practices." Bruton (2010) also expresses his concern about the fact that "sometimes academic debate uses research results and instruments to convince non-academics of their arguments, when the design of the research cited are far from sound" (p. 491). He also emphasizes the role of factors such as instruction, tasks, and grades in affecting learners' success:

If corrective feedback recognizes interest in the content of tasks, which are within the students' capabilities, is supportive and constructive, while rewarding improvement, reflected in the grading system, the conditions might be propitious for improvement... If teacher response emphasizes the defects (in red), shows a lack of interest for the content and offers criticism, reinforced by negative grades based on errors, the circumstances are hardly beneficial for improvement...Any grading system for L2 writing, probably needs to reward improvement, both in terms

of content and new language use, together with complexity/accuracy, and in terms of reducing recurrent errors. (pp. 496-497)

Teachers are also known to have their own beliefs about what constitutes good feedback and how it must be provided, which sometimes contradict those of students. For example, teachers tend to perceive their feedback more positively than students do. Tutors believe that they provide more detailed feedback than their students think they do. They also perceive their feedback to be more useful than students do. Finally, teachers tend to find their assessment to be fair while students are not sure about that (Carless, 2006). Lee (2009) also reports some discrepancies between teachers' beliefs and what they practice. For instance, they tend to focus more on language form while they believe they should not. They practice comprehensive error marking though they believe it should be selective. They also grade students' writings though they believe that grades draw learners' attention away from the intended feedback provided with the teacher.

No matter what conclusion research studies come up with, language teachers seem to continue providing their learners with corrective feedback mostly because they think they should. Leki (1990) asserts that although written comments to students' writings are time consuming, teachers still continue to provide them with these comments because they believe that that will help the writers improve. He also believes that teachers do so because their job not only requires them to evaluate students' writings but also needs them to justify their evaluation.

Grading dilemma

Providing corrective feedback can result in a clash of roles on the part of the teacher. Leki

(1990) holds three roles for a writing teacher in responding to her students' writings: teachers as real readers (audience); teachers as coaches; and teachers as evaluators. Given the unequal power relation between a teacher and a student, Leki sees it unrealistic to accept that teachers can read learners' writings in the same way as they read texts they read on their daily life. A teacher may also act as a coach as well as an evaluator. This way she needs to cooperate with learners in that process. As such, she will be responsible if students fail to meet the criteria because it means she had not intervened enough when necessary. However, this being a collaborator and a judge at the same time is a contradiction which sounds difficult to resolve. Being an evaluator (the third role) also contradicts with another notion taught to students. Usually, students are encouraged to have in mind an audience for their writings, but simply knowing that the reader is not going to be a simple audience and is an evaluator distorts such a notion (Leki, 1990).

While being an evaluator can be in clash with other roles a writing teacher may have, performing such a role seems inevitable. However, being an evaluator is not as problematic as being an assessor. While an evaluator may evaluate a piece of writing by commenting on the weak points or specifying the parts or elements which need to be amended, she does not need to assign any score or grade to that piece of work. On the other hand, when acting as an assessor, a teacher is required to provide learners with a grade or score which can sum up her evaluation in the form of a single easily interpretable grade or score. However, such a practice may divert learners' attention away from teacher feedback and as a result do more harm than good (Lee, 2009).

Lee (2009), having administered a questionnaire to 206 secondary teachers and having conducted an interview with a few of them, explored their beliefs and their reported practices to examine the extent to which they correspond each other. She identified ten mismatches between teachers' beliefs and their written feedback practice. She found out that "teachers award scores/grades to student writing although they are almost certain that marks/grades draw student attention away from teacher feedback" (p. 16). She states that the feedback analysis shows that all the teachers give their students' writings a score. However, they do not believe that much in their usefulness because they think scores and grades divert learners' attention away from teacher feedback to the extent that some students may even ignore them particularly when they are not required to revise and resubmit their drafts for better grades. "One teacher remarked, 'The majority of students do not pay attention to the comments'. Another teacher even said, 'For students, they only look at the scores'." (p. 17).

This way, as Hamp-Lyons (2007) points out, in many contexts writing assessment is taking over writing instruction, that is, increasing attention is being paid to the issue of grading or scoring student writing. Connors and Lunford (1993), having conducted a discourse analysis of comments on 3,000 marked papers, observed that more than 80% of the comments had a judgmental tone. Such studies show that instructors read assignments for the purpose of grading and their feedback is mainly concerned with justifying the grades given (Li & Barnard, 2011).

One may wonder why teachers do not stop grading or scoring student writing if they are aware of the harm it does. Lee (2009),

quoting the same teachers, argues that grading is necessary for summative purposes. One teacher in the follow-up interviews emphasized the importance of grading by saying that he believes that compositions, except identifying students' difficulties in writing, serves another function, i.e., it serves for teachers to hand over score sheet. As such it seems that "the summative function of feedback has made teachers use scores/grades although they are fully aware of the harm that can be done to students" (p. 17).

However, that is not the only reason why teachers continue grading learner writing in conjunction with the corrective feedback they provide them with. Learners demand such a practice. Lee (2008) studying both high proficient (HP) and low proficient (LP) students of English during an academic year, examined their preference for the type of feedback they received. 72.2 percent of HP students and 40.9 percent of LP students chose the option 'mark/grade + error feedback + written comments.' In response to the question 'In the future compositions, which of the following would you be most interested in finding out?', 'teacher's comments on my writing' ranked first by 47.2 percent in HP students and 36.4 percent in the LP students. 'mark/grade' stood second by 38.9 percent in HP students and 36.4 percent in LP students.

The present study

Having been confronted with all such contradictions, we tried to find a middle ground compromising all such problems. In fact, it was tried to find a solution for motivating learners' to attend to teacher feedback while providing them with grades that can satisfy teachers' sense of obligation in having summative evaluation and learners' sense of need for such an evaluative feedback without jeopardizing

learners' attendance to teacher feedback. It not only does not divert learners' attention from teacher feedback, but it also gives them, at least for the majority of learners, a reason and the needed motives to attend to that.

The solution we came up with was a simple technique called Draft-Specific Scoring, based on which learners are provided with corrective feedback as well as a grade which represents the teacher's general evaluation of that piece of work. The final score would be the mean of all the grades learners have received for their assignments during the course. However, the grades learners receive are not fixed. Students can improve their grades by applying teacher feedback to their writings and revising their first and mid drafts. Usually, students are given two opportunities to go through this procedure of drafting and revising. The final score each student receives on any assignment is used to come up with the mean score. The present study was an attempt to check the effect of this newly-developed technique on the fluency, grammatical complexity, and accuracy of the texts learners write over the course of instruction. As such the following research questions were formulated:

1. Does the fluency of texts written by learners change over the course of instruction as a result of using DSS when providing teacher corrective feedback?
2. Does the grammatical complexity of texts written by learners change over the course of instruction as a result of using DSS when providing teacher corrective feedback?
3. Does the accuracy of texts written by learners change over the course of instruction as a result of using DSS

when providing teacher corrective feedback?

Method

Participants

There were 85 participants present in two groups from two different universities, namely University of Tehran and Azad University. There were 26 (10 male and 16 female) participants in the treatment group at the University of Tehran. Their age roughly ranged from 22 to 25. They were all high intermediate EFL learners studying English Literature. They were all Iranian but for one Chinese female student. For the control group, 57 participants were present, all studying English Literature and Translation at Azad University. After these participants were filtered, 31 (12 male and 19 female) participants with an age range of 21 to 27 remained. Since the participants at Azad University were more heterogeneous in language proficiency level in comparison with those studying at the University of Tehran, they were matched based on the Oxford Quick Proficiency Test they had taken as a requirement of their department and the results of the pretest in writing. As a result, out of the 57 participants for the control group, there remained only 31 for data analysis.

Procedure

During the first 3 sessions, the preliminaries of writing were taught to both groups, and using model essays, different parts and components of an essay were discussed and instructed. The base of the instruction was TOEFL iBT independent Task in writing which is very similar to IELTS task 2 in writing. In these tasks, test takers are given a prompt and are asked to write an essay on that in a limited time. The given time is 30 minutes in TOEFL iBT and 40 minutes in IELTS test. As such, learners were informed of the criteria based on which their writing

samples were supposed to be evaluated and scored. In the fourth session, samples of students' writing were collected as the pretest. Participants in both groups were given 80 minutes to plan and write about a given topic. The samples were scored and returned to the participants with teacher comments on them. They received scores given by their instructor based on the general impression and the quality of their writing. The two sets of scores given by expert raters were later contrasted for making sure that the participants in both groups were comparable in their writing proficiency. No significant difference was found between the two groups at the pretest: $t(55) = .11, p = 0.91$.

To prevent Halo and Hawthorne effects, both groups were kept blind to the fact that they were being studied. During class time, some of the learners' writing samples were chosen and discussed with the whole class, and their weaknesses and strengths were pointed out. Each session, learners' essays were collected, scored, and commented on by the teacher researcher. At the end of each session, the participants were assigned a new topic to write about for the following session. Their essays had to be at least 150 words long, typed and printed in an A4 paper. Learners' essays were read by the researcher, and for the grammatical mistakes, learners were provided with indirect corrective feedback, i.e., the errors were underlined but not corrected. To keep the conditions the same for all, no explicit feedback were given in the samples for the problems they had with the style of writing and issues such as topic development, topic relevance, coherence, and cohesion. Instead, some of those samples with such problems were identified and discussed with the whole class during the class time. However, for all essays, if necessary, it was commented that they need to be improved stylistically in

terms of topic development, for instance. The participants were required to revise the drafts they had submitted based on the feedback they had received and return them to the teacher the following session. The two groups were told that their final score would be the average score for all the scores they had received for their assignments during the course. Both groups wrote 9 assignments during the course including the pretest, and the posttest. However, they did not have the opportunity to revise their drafts for the posttest. As such, they received comments on only 8 assignments during the whole course. Their final exam was regarded as their posttest.

Up to this point, the procedure followed was the same for both the control and treatment groups. However, the two differed in one major aspect. The scores given to the essays written by learners in the control group were fixed, that is, they did not change after the revisions made by learners, but in the case of the treatment group, learners could improve their scores by the revisions they made. For example, a learner who had received 14 out of 20 for the draft she had submitted could revise her sample based on the feedback she had received and improve her score. She could receive 16, or 18 or any other score based on the quality of her revised sample. She could even receive the same score in case the revisions were not satisfactory. The revised samples were again commented on and returned. The learners had one more opportunity to revise their returned samples and undergo the same procedure. This is what we call *Draft Specific Scoring*.

Both groups received a sample of the score profile in which the instructor would record their scores in order to come up with their final score at the end of the semester. Their final score would be the mean of all the

scores they received on their assignments during the semester. For the treatment group, the final score they received on the last revision they submitted was taken into account while for the control group the single score they received for each score were used to calculate their final score. They were also recommended to keep a similar profile for themselves. Here are the sample score profiles for both treatment and control groups:

	Student Name:									
Assignment	1	2	3	4	5	6	7	8	9	10
1st Draft										
2 nd Draft										
3 rd Draft										

Final score: the mean score of all assignments

Figure 1: Score profile for treatment group

	Student Name:									
Assignment	1	2	3	4	5	6	7	8	9	10

Final score: the mean score of all assignments

Figure 2: Score profile for control group

Performance measures: Fluency, grammatical complexity, and accuracy

Regarding the fluency measures, a number of measures were present to choose from. Chandler (2003) used the amount of time it took her participants to write an assignment. She did so because the length of each assignment was fixed. However, Truscott (2004) objected to that. Truscott believes that the number of words must be the measure used to assess fluency. The studies done before Chandler (2003) had also used the number written words as the measure of fluency. The same measure is also used in the present study as the measure of fluency.

In order to check for the complexity of texts written by students in both groups over time, two measures were examined as introduced by Wolfe-Quintero, Inagaki, and Kim

(1998) as some of the best measures used in the literature: the ratio of the number of clauses to the number of T-units, and the number of dependent clauses used. The second measure was also used by Robb et al. (1986) to check learners' change in grammatical complexity. Maybe this measure can be regarded as a more straightforward measure because it is in the form of frequency rather than ratio and can be more easily interpreted as it is affected only by one index not two as in a ratio.

In the case of checking the change in learners' accuracy level, the ratio of error-free T-units to the number of T-units was used as introduced as the best measure of accuracy by Wolfe-Quintero, Inagaki, and Kim (1998).

In order to be consistent and accurate in counting the number of different elements such as T-units, error-free T-units, dependent clauses, and the number of clauses in participants' samples, there had to be an operational definition for each. A dependent clause could be any type of adverb clauses, adjective clauses, or noun clauses. All reduced clauses were also counted. An independent clause was one which was complete in meaning and did not need any other clause to complete it. A T-unit was an independent clause with all the dependent clauses attached to it. As such, every sentence including only one independent clause was also a T-unit (Wolfe-Quintero, Inagaki, and Kim, 1998). An error-free T-unit was a T-unit which did not include any kind of error but for spelling and punctuation. All the writing samples were rated with only one rater for the measures in fluency, grammatical complexity, and accuracy. As Chandler (2003) states, in such studies, the intra-rater reliability is more important than the inter-rater reliability. The intra rater reliability for

all the measures was above .94. In order to check the change in learners' fluency, grammatical complexity, and accuracy, either the gain scores were checked or the SPANOVA was used.

Results

Due to the design of the study, SPANOVA could be the best statistical test for data analysis. However, this test has some underlying assumption which must be met. In this section for research questions in which such assumptions were met, the results of SPANOVA were reported. In other cases, the gain score analysis was used as a good substitute to the use of SPANOVA.

The first research question addressed the existence of any significant change in learners' fluency of writing and the difference between the two groups as a result of the intervention received by the treatment group. A SPANOVA was performed for the two groups across the two time periods (pretest, and the posttest). There was a significant interaction between time and group, Wilks' Lambda = .74, $F(1, 55) = 18.96$, $p < .0005$, partial eta squared = .26. There was a substantial main effect for time, Wilks' Lambda = .57, $F(1, 55) = 41.04$, $p < .0005$, partial eta squared = .43. However, the main effect for Group, comparing the effect of the intervention on the two groups, was not found statistically significant, $F(1, 55) = 1.02$, $p = .32$, suggesting a lack of benefit for any group over the other one and an improvement for both groups in the number of words written. It is worth mentioning that according to Cohen (1988, pp. 284-7), .01 eta squared shows small effect, .06 shows moderate effect, and .13 represents a large effect size.

Table 1 summarizes the descriptive statistics for the two groups across time.

Table 1: Number of words written by the 2 groups

	<i>Group</i>	<i>Mean</i>	<i>S.D</i>	<i>N</i>
<i>Pretest</i>	Treatment	280.58	77.99	26
	Control	289.42	73.12	31
<i>Posttest</i>	Treatment	359.08	119.54	26
	Control	304.39	84.27	31

The second research question addressed the change in learners' grammatical complexity of texts written across time from the pretest to the posttest. Since the picture looks somewhat blurred after using SPANOVA, it seems reasonable to analyze the data using another procedure. The comparison of the gain scores of the two groups from pretest to posttest is a good substitute to the use of SPANOVA and is mathematically the same as that (Anderson, Auquier, Hauck, Oakes, Vandaele, & Weisberg, 1980).

Regarding the first measure of grammatical complexity, that is, the ratio of the clauses to T-units, there was no significant difference between the gain scores of the two groups at the end of the instruction, $t(55) = -.25$, $p = .79$. The paired samples t tests run between each group's pretest to posttest showed no significant difference for the treatment group [$t(25) = 1.33$, $p = .20$], but for the control group, it was found statistically significant, $t(30) = 3.86$, $p = .00$, eta squared = .33.

Table 2: Ratio of clauses to T-units

	<i>Group</i>	<i>Mean</i>	<i>S.D</i>	<i>N</i>
<i>Pretest</i>	Treatment	1.90	.41	26
	Control	1.82	.35	31
<i>Posttest</i>	Treatment	1.77	.38	26
	Control	1.72	.28	31

Regarding the second measure of grammatical complexity, that is, the number of dependent clauses used, a SPANOVA was performed for the two groups across the two time periods (pretest, and posttest). There was a significant interaction between time and group, Wilks' Lambda = .91, $F(1, 55) = 5.24$, $p = .03$, partial eta squared = .09. There was a substantial main effect for time, Wilks' Lambda = .79, $F(1, 55) = 14.80$, $p < .0005$, partial eta squared = .21. However, the main effect for Group, comparing the effect of the intervention on the two groups, was not found statistically significant, $F(1, 55) = .90$, $p = .35$, suggesting a lack of benefit for any group over the other one though they both had significantly improved over time. Table 3 summarizes the descriptive statistics for the two groups across time.

Table 3: Number of dependent clauses

	<i>Group</i>	<i>Mean</i>	<i>S.D</i>	<i>N</i>
<i>Pretest</i>	Treatment	13.19	6.08	26
	control	13.32	5.35	31
<i>Posttest</i>	Treatment	17.77	10.04	26
	control	14.48	5.52	31

The analysis of the gain score also shows the same pattern. The Mann-Whitney U test run between the gain scores of the two groups was not found statistically significant, $U = 288$, $z = -1.85$, $p = .06$. The Wilcoxon Signed Rank tests between the two groups' change from pretest to posttest showed significant differences for both the treatment group, $z = -2.63$, $p = .01$, and the control group, $z = -2.41$, $p = .02$.

All the above statistics indicate that as in the case of previous measure of grammatical complexity, no significant difference was observed between the two groups in the complexity of texts they wrote. However, unlike the previous measure, this measure

showed a significant improvement in both groups' complexity of texts they wrote.

The last question checked whether the two groups did not differ from each other in the accuracy of texts they wrote across time from pretest to posttest. The data were analyzed using gain score procedure. The independent samples t test run to compare the two groups' gain scores in accuracy was found significant, $t(55) = 2.48$, $p = .02$, Eta squared = .10 which is a large effect size. Tables 4 and 5 summarize the descriptive statistics for the two groups' gain scores. Moreover, the difference between the treatment group's mean of accuracy measure from pretest to posttest was statistically significant, $t(25) = -2.82$, $p = .01$ with a quite large effect size (Eta squared = .24). However, this change was not found statistically significant for the control group, $t(30) = 1.14$, $p = .26$, suggesting an advantage for the treatment group over the control group. This shows that while DSS was successful in improving the accuracy of texts written by learners across the course of instruction, the control group did not succeed in improving in accuracy.

Table 4: Change in accuracy over time

	<i>Group</i>	<i>Mean</i>	<i>S.D</i>	<i>N</i>
<i>Pretest</i>	Treatment	.80	.14	26
	Control	.78	.09	31
<i>Posttest</i>	Treatment	.86	.13	26
	Control	.76	.15	31

Table 5: The two groups' gain scores in accuracy

	<i>Group</i>	<i>N</i>	<i>Mean</i>	<i>S.D</i>
<i>Gain Score</i>	Treatment	26	.06	.11
	Control	31	-.02	.16

Discussion

In the present study, as an attempt to find a solution to the long-lasting problems grading and even corrective feedback were said to cause, it was tried to examine the effect of Draft-Specific Scoring on the fluency, grammatical complexity, and accuracy of the texts learners write. This was mainly a response to the previous research in the field which indicates that learners receiving corrective feedback write shorter and simpler texts due to the use of avoidance strategy while their accuracy does not improve.

While both groups significantly improved in fluency from pretest to posttest, the difference between the two groups was not found statistically significant even though the treatment group had outperformed the control group by 55 words in the posttest. This pattern of results suggests that what Truscott states about the disadvantage for the correction group in fluency is not true because even the control group improved in fluency.

In the case of change in learners' grammatical complexity of the written texts over time, the measure involving ratio showed no difference between the gain scores of the two groups. However, based on the descriptive statistics, both groups had a decrease in the complexity of their written texts. Although this decrease was statistically significant for the control group, it was not for the treatment group.

Since this first measure was in the form of a ratio, it was affected by two variables, the numerator and the denominator. The change in any one of these can have its own interpretation while the combination of the two makes it very difficult to interpret. Therefore, the second measure, the number of dependent clauses, can be a better index.

Maybe that was why Robb et al. (1986) also used this measure for checking grammatical complexity. The results of checking this measure indicate that not only did the complexity of learners' texts not decrease, but it actually increased over time. This increase was significantly different for both groups but not from each other. The observed pattern of results regarding grammatical complexity is in line with that in Robb et al. (1986). All in all, these findings indicate that at least even if the provision of corrective feedback plus DSS does not increase the grammatical complexity of the learners' texts, it does not let it decrease.

Regarding the final research question, examining the change in learners' level of accuracy, the results point to the superiority of DSS approach over the more traditional methods of feedback provision. While learners receiving corrective feedback alone did not improve in accuracy, the ones receiving corrective feedback plus DSS did improve in accuracy over time.

It seems that Truscott (1996, 2004, 2007) has been right to some extent regarding the behavior of learners receiving corrective feedback alone. The control group was observed not improving in accuracy. Regarding grammatical complexity, it showed a significant decline according to one of the measures and showed a significant improvement according to another measure more commonly used in the literature. The control group, however, improved in fluency, which contradicts Truscott's prediction. On the other hand, the treatment group receiving corrective feedback plus DSS proved to be more successful in improving in fluency, grammatical complexity, and accuracy. Even when learners receiving corrective feedback alone improved in a measure,

those receiving corrective feedback plus DSS could outperform them. This shows that DSS has the potential to overcome the weaknesses traditional methods of feedback provision have.

DSS also seems to be more consistent with the process approach to writing in which the emphasis is on mid drafts rather than final drafts. Feedback on mid drafts assume a much higher importance to the extent that Muncie (2000) states that if feedback is going to work, it does so on mid drafts. Moreover, many studies (Ellis & He, 1999; Ellis, Tanaka, & Yamazaki, 1994; Long, Inagaki, & Ortega, 1998; Mackey, 1999; Mackey & Oliver, 2002; Mackey & Philip, 1998; McDonough, 2005) have connected interactional feedback with L2 learning since it causes learners to notice L2 forms. They are all based on Long's interactional hypothesis (Long, 1996, 2006). He proposes that due to the role of interaction in connecting "input, internal learner capacities, particularly selective attention, and output in productive ways," interactional processes can facilitate language learning (Long, 1996, pp. 451-452). Such helpful processes can include the negotiation of meaning and the provision of recasts, both of which regarded as kinds of corrective feedback to help learners detect their problematic utterances. One process that can arise from such feedback is modified output (Swain, 2005), which can be helpful in language learning (Mackey, 2006). In addition, no matter in conversational interactions or in written interactions, learning will not occur if there is not a form of noticing on the part of learners. In case learners do not pay

attention or attend to the feedback the teacher provides them with, there will be no L2 development. In case they notice it, but it does not result in any modified output, again whether learning has occurred or whether the potential for learning has been fully fulfilled is questionable.

On the one hand, by motivating learners to attend to teacher feedback, DSS is a device to ensure learners' paying attention to teacher feedback and their noticing of that. On the other hand, by requiring them to revise their drafts, it helps them have modified output. Since understanding teacher feedback and teacher intention has not always been easy for learners, when they attempt to incorporate teacher feedback in such a system, there are times when questions are raised for them about teacher intention by, for example, underlining a sentence. It is also possible that they revise a sentence underlined by the teacher, but in the returned draft, they observe that the same sentence is underlined again. In usual systems of evaluation, this usually results in frustration on the part of the learners resulting in the abandonment of the draft by him. However, in DSS learners, having a good reason for it, consult with the teacher about his or her intention. This is what can be called the negotiation of meaning. As such, it can be observed that DSS has the potential to incorporate all the necessary processes for helping learners develop their L2.

Conclusion

Using DSS, teachers will not have to change the principles underlying their practice. Teachers are repeatedly reported to express

their belief in grading. Grading also helps teachers have a better overall assessment of their students at the end of the semester (Lee, 2009). Teachers, however, are aware of the harm grading may do to learners. They know when learners see grades on their paper, they will most probably ignore teacher comments and feedback (Lee, 2009), but still they continue to grading not only because of their belief in grading and actually their kind of obligation for it, but also because of their students' demands for that. Students strongly demand for grades because grades help them evaluate themselves easier. Grades are also more easily interpreted than sometimes elaborate comments all over their paper (Lee, 2008). If teachers continue grading, learners will pay less attention to their feedback. If they stop grading, they will face new problems.

DSS lets teachers continue their preferred practices while minimizing the negative effect of grading and changing its weak point to strength. It uses grading as a motivating factor which not only does not divert learners' attention from teacher feedback, but it also ensures their attendance to it.

DSS also addresses Hamp-Lyons' (2007) concern. She believes that in most contexts, writing assessment is taking over writing instruction. As a result, grading and scoring student writing is increasingly receiving more attention. DSS changes the old practice in which grading was 'the end' in the story of writing instruction. It makes grading a new 'once upon a time' in each draft. It combines assessment with instruction without omitting any. It keeps both assessment and instruction in one go.

Learners do not only become aware of the teacher's evaluation of their work, but they also know that this is the beginning of the revision process. They know that when they receive a grade on their writing sample, it works like a compass to be used with teacher feedback in order to improve their writing skill and find their way to a better performance.

All in all, it seems that what is important is not whether teachers provide their students with corrective feedback. What is of utmost importance is whether learners' attend to the feedback they are provided with. Even mere attendance cannot be the end of the story. Learners need to attend and apply the corrective feedback they receive. In other words, learners need to notice the input and try to have an output based on the intake they had. This way, teachers' efforts are more likely to result in the desired outcome. Draft-Specific Scoring, as a technique ensuring such a process, can be quite helpful in pursuing such instructional objectives.

References

- Anderson, S., Auquier, A., Hauck, W.W., Oakes, D., Vandaele, W. & Weisberg, H.I. (1980). *Statistical methods for comparative studies*, New York: John Wiley and Sons.
- Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were. *System*, 37, 600-613.
- Bruton, A. (2010). Another reply to Truscott on error correction: improved situated designs over statistics. *System*, 38, 491-498.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219-233.

- 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*, 267–296.
- Chandler, J. (2004). Dialogue: A response to Truscott. *Journal of Second Language Writing, 13*, 345–348.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Connors, R. J., & Lunsford, A. A. (1993). Teachers' rhetoric comments on student papers. *College Composition and Communication, 44*(2), 200–223.
- Diab, R.L. (2005). EFL university students' preferences for error correction and teacher feedback to writing. *TESL Reporter, 38*, 27–51.
- Ellis, R. & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*, 285–301.
- Ellis, R., Tanaka, T., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning, 44*, 449–91.
- Enginarlar, H. (1993). Student response to teacher feedback in EFL writing. *System, 21*(2), 193–204.
- Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly, 29*, 33–53.
- Ferris, D.R. (1999). The case for grammar correction in L2 writing classes. A response to Truscott (1996). *Journal of Second Language Writing, 8*, 1–10.
- Hamp-Lyons, L. (2007). Editorial. *Assessing Writing, 12*(1.), 1–9.
- Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*, 144–146.
- Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal, 63*, 13–22.
- Leki, I. (1990). Coaching from the margins: issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 57– 68). Cambridge, UK: Cambridge University Press.
- Li, J., & Barnard, R. (2011). Academic tutors' beliefs about and practices of giving feedback on students' written assignments: A New Zealand case study. *Assessing Writing, 16*, 137–148.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. J. Bahtia (Eds.), *Handbook of second language acquisition*, (pp. 413–68). New York: Academic Press.
- Long, M. H. (2006). *Recasts in SLA: The story so far*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recast in Japanese and Spanish. *The Modern Language Journal, 82*, 357–371.
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*, 557–87.
- Mackey, A. (2006). Feedback, Noticing and Instructed Second Language Learning. *Applied Linguistics, 27*, 405–530.

- Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, 30, 459–77.
- Mackey, A. & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, 82, 338–56.
- McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27, 79–103.
- Muncie, J. (2000). Using written teacher feedback in EFL composition classes. *ELT Journal*, 54 (1), 47–53.
- Radacki, P. M., & Swales, J.M. (1998). ESL student reaction to written comments on their written work. *System*, 16, 355–365.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83–93.
- Saito, H. (1994). Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal*, 11, 46–70.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.): *Handbook of research in second language teaching and learning* (pp. 471–83). Mahwah, NJ: Lawrence Erlbaum Associates.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327–369.
- Truscott, J. (2004). Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing*, 13, 337–343.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Second Language Writing* 16, 255–272.
- Truscott, J. (2010). Some thoughts on Anthony Bruton's critique of the correction debate. *System*, 38, 329–335.
- Wolfe-Quintero, Kate, Shunji Inagaki and Hae-Young Kim (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.

