



Reinforced Teacher Corrective Feedback and Learners' Use of Subordination Clauses

Masoud Azizi^{1*}, Majid Nemati²

¹ Assistant Professor, Department of Foreign Languages, Amirkabir University of Technology, Tehran, Iran

² Associate Professor, Department of English, Faculty of Foreign Languages and Literatures, University of Tehran, Tehran, Iran

Received: 2021/12/19

Accepted: 2022/05/14

Abstract: With numerous variables mediating the way learners interact with teacher corrective feedback (TCF), one may not comment on its efficacy before such intervening variables are adequately addressed and learners' attendance to TCF is ensured. Among those variables, motivation is one of the most prominent ones largely affecting the degree to which students benefit from TCF. Draft-Specific Scoring (DSS), a scoring system giving learners' a reason to notice TCF by rewarding them for the revisions they make through a flexible system of scoring, was implemented to investigate if TCF could help learners improve their use of subordinate clauses. To do so, two groups of students studying English Language Literature at the University of Tehran, consisting of 55 participants who were randomly assigned as treatment and control groups, were studied. The results of the gain analysis indicated an improvement for both groups over time in the total number and accuracy of the subordination clauses used; however, the treatment group significantly outperformed the control group. While the two groups did not differ in their use of noun clauses, the DSS group was found to outperform the non-DSS group regarding the adverb and adjective clauses. This indicates that motivation to attend to teacher feedback is of great importance if TCF is to be effective.

Keywords: Draft-Specific Scoring (DSS), Subordination, Focused Feedback, Writing, Assessment.

* Corresponding Author.

Authors' Email Address:

¹ Masoud Azizi (mazizi@aut.ac.ir), ² Majid Nemati (nemati@ut.ac.ir)



Introduction

Learners' attendance to teacher corrective feedback (TCF) is of great significance if one is to comment on its effectiveness. If learners' attention to teacher feedback and the application of that in their future writings are not ensured, one cannot guarantee the efficacy of the feedback (Azizi & Nemati, 2018b). There are also several variables and individual differences, with motivation being the most important one (Bruton, 2009, 2010; Ferris, 1999; Lee, 2008), that can mediate and affect such an attendance and hence the effectiveness of teacher corrective feedback.

What is often observed in the literature on the studies on TCF is that learners' attendance has by default been presumed though nothing has been done to ensure or at least check it (Azizi & Nemati, 2018a; Guenette, 2007). The existence of a great number of research studies with contradictory implications about the effectiveness or ineffectiveness of TCF in the literature can be attributed to such an unwarranted assumption (Bitchener & Ferris, 2012; Diab, 2015). Students are often reported not to be driven enough to pay attention to TCF (Lee, 2014; Truscott, 1996) and the existence of a number of distracting variables affecting their attention, such as grading their writing samples (Lee, 2008), demands the conduction of studies in which learners' attendance to teacher feedback has been first ensured before one can reasonably commend on the efficacy or inefficacy of TCF (Azizi & Nemati, 2018a, 2018b, Nemati & Azizi, 2013). The present study was an attempt to address such a gap in the literature.

Literature Review

Over the past few decades, a great portion of the publications in L2 writing has been in response to Truscott's (1996) thesis concerning the usefulness of TCF. After he questioned the effectiveness and even the necessity of providing students with teacher corrective feedback, other scholars such as Ferris (1999, 2004), Chandler (2003), and Bruton (2009, 2010), responding to his objections, argued for the continuation of the practice. The abundance of studies with results in favor of both parties and the large variation in the variables they investigated has made it very difficult to draw any clear and firm conclusion regarding the effectiveness of TCF (Ferris, 2004; Guenette, 2007).

In Truscott's (1996) opinion, TCF has been practiced because it is assumed to be effective and the value of grammar correction has been taken for granted. He believes that learners do not attend to teacher feedback or incorporate it in their future writings as they lack the necessary motives. In addition, the negative impacts of correcting learners' errors, such as the adverse

effect of that on learners' attitudes and the amount of energy and time it demands, are often ignored. To him, grammar correction does more harm than good. He claims that uncorrected students enjoy a more positive attitude and may write more in comparison with the corrected ones, and the ones receiving TCF often write shorter and simpler texts so that they can avoid being corrected. As a result, when they are observed to improve in accuracy in the studies carried out on TCF, it is because they try to hide their weaknesses through avoiding the structures they are not sure of or might be corrected for (Truscott, 2007). In his opinion, grammar correction cannot be part of writing instruction and needs to be stopped.

Ferris (1999), calling Truscott's thesis premature and overly strong, argues that many students, if not all, do and can benefit from TCF. In her opinion, we need to continue such a practice as learners strongly demand and highly value teacher corrective feedback because students often hold that TCF is necessary for them and can be instrumental for their improvement (Lee, 2008). Learners are often so concerned about their surface-level errors that the instructor's credibility can be adversely affected if he fails to indicate or comment on all errors in the writing samples (Radacki & Swales, 1998). As a result, to learners, good and successful teacher feedback is the one in which adequate attention has been devoted to the surface-level errors (Diab, 2005; Ferris, 1999).

Ferris (1999) believes that instead of abandoning feedback provision, we need to make sure it is effective. She holds that it is essential that we adequately probe into individual learner factors impacting their desire and ability to use TCF. It is essential that we try to identify the techniques and methods in error correction that can help improve learners' writing skill. Only then can we comment on the plausibility of Truscott's thesis.

Bruton (2009), arguing against Truscott's position, asserts that it is reasonable to assume that additional positive and negative evidence can result in higher levels of accuracy in learners' writing performance. Highlighting the close bond between learners' motivation and their effort to progress, Bruton (2010) states that variables such as tasks, instruction, and scoring can impact students' achievement and need to be noted. In case teachers merely focus on learners' errors, do not show any interest in the content of the responses, and only reinforce their criticisms with negative scores, no one can expect the outcome to lead to 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" (Bruton, 2009, pp. 496-497).

According to Bruton (2010), learners are often provided with no purpose or motive for

what they are asked to do with the feedback they receive. In addition, they are often provided with no grade on their writing, and even when they are, no reference is made to the content, which can enhance the possibility of avoidance. All these are demotivating (Bruton, 2010). He believes that learners need the motive to try to improve their accuracy. However, Truscott (2010) accuses Bruton of presenting not a single case where learners had a strong motive and TCF was observed to be effective.

While Bruton's position on the important role of grades in the efficacy of TCF seems fascinating, the literature on grading shows that teachers face a number of challenges if they wish to provide learners with grades (Nemati & Azizi, 2013). Grades are known to avert learners' attention away from teacher comments and feedback. Learners have often been reported to ignore teacher feedback as soon as they find the grade accompanying teacher feedback especially if they are not asked for a revision and the resubmission of their papers (Lee, 2009).

As Hamp-Lyons (2007) warned, writing assessment is starting to dominate and lead writing instruction in most contexts with growing attention being paid to scoring or grading learners' writings. Connors and Lunsford (1993), analyzing papers commented by the teachers, found that over 80% of the statements made by teachers on the papers were judgmental. In Li and Barnard's (2011) study on the degree to which instructors believed in the significance of grading while providing learners with corrective feedback, all teacher participants were reported to consider grades a central part of their feedback. According to Li and Barnard (2011), teachers' provision of feedback was mainly intended to justify the scores they had awarded to students' assignments rather than to seek to help learners improve.

Despite all the problems grading may impose and in spite of the fact that teachers are aware of the harm it does, they still continue the practice of grading mainly because it is essential for summative evaluations during a program, and this evaluation is the one often demanded by most educational institutions (Lee, 2009). Moreover, learners' beliefs and preferences regarding teacher feedback are of great importance since they may impact the way learners interact with it (Han, 2017, 2019). Learners strongly demand and highly value receiving teacher corrective feedback on their writing samples (Lee, 2009). In a study on learners' preferences for the type of received feedback, Lee (2008) reported that 72.2 % of high proficient participants and 40.9 % of low proficient learners selected 'mark/grade + error feedback + written comments'.

Besides what learners demand, one needs to consider how they feel when they are

engaged with TCF as their feelings and attitude can also influence how they may interact with it (Han & Hyland, 2019). The way teachers provide students with corrective feedback may provoke different emotional reactions and feelings in learners (Han & Hyland, 2019; Zhang & Hyland, 2018). Some may feel excited while others may feel indifferent (Han & Hyland, 2015). There might be some learners who may feel frustrated (Zheng & Yu, 2018). While some might feel honored (Ferris, Liu, Sinha, & Senna, 2013), others may feel self-confident (Storch & Wigglesworth, 2010).

Regardless of the myriad of contradictory results and the debate among scholars concerning the efficacy of teacher corrective feedback, still, teachers continue to provide their learners with such feedback though they find the practice quite time-consuming mainly because they believe they need to and that it leads to learners' improvement (Li & Barnard, 2011; Moradkhani & Goodarzi, 2020). Teachers also do so because they feel obliged by their job to not only offer their evaluation of students' progress but also be able to rationalize that evaluation (Leki, 1990; Li & Barnard, 2011). However, learners are often observed not to attend to teacher comments and feedback as they are not motivated enough (Lee, 2014). As a result, it is suggested that variables such as tasks and grades be manipulated so that they can stimulate and enhance learners' attendance and attention (Bruton, 2009). This is also quite challenging as grades have been found to adversely impact learners' attention to teacher feedback (Lee, 2008). Still, teachers, though aware of the harm grading does, continue to provide them for a number of reasons such as institutional requirements and learner demand (Lee, 2009).

Having faced these paradoxes, one needs to come up with a middle ground conciliating the contradictions and challenges (Guenette, 2007). We need to find a way to ensure students' attention to and engagement with teacher feedback while at the same time awarding them grades so that teachers are able to respond to both their sense of obligation for summative evaluations and students' demand without risking attendance. We need a solution that does not jeopardize students' attendance; instead, it provides them with a good reason to optimize it (Azizi & Nemati, 2018a,b). The solution we offered was an assessment protocol we named Draft-Specific Scoring or DSS (Nemati & Azizi, 2013).

In this technique, students receive both TCF and grades together representing the instructor's general evaluation of their work. No matter how many assignments they are asked to write during the program, their final score for that course will be the mean score for the grades they receive on all those assignments. However, the score they are provided with on

each assignment is not fixed and can be improved as the result of the revisions learners make based on the feedback they have received on that paper. The revisions might also be initiated by learners themselves and may include improvements in content, structure, style, topic development, or anything helping improve the quality of their writing. Based on the quality of the revisions on the mid drafts, the teacher will award new grades. Students' final score on the program will be the mean score for all the final grades they receive on the last revision they submit for each assignment. Often students are given up to two opportunities to revise their first draft, resubmit revisions, and improve their grades (Nemati & Azizi, 2013).

Nemati and Azizi (2013), investigating the effect of DSS on the fluency, accuracy, and grammatical complexity of the texts written by participants found no significant difference between the DSS group and the control group in the number of words written as the measure of fluency though the DSS group had written an average number of 55 words more. Interestingly enough, both groups had significantly improved in that index over time from pretest to posttest, which contradicted Truscott's claim (1996) based on which students receiving corrective feedback tend to write less. However, the difference between the two groups regarding the measure of accuracy was found statistically significant, with the DSS group being able to significantly improve over time while the control group could not display any significant improvement. Finally, regarding the measure of grammatical complexity, while the control group receiving corrective feedback did significantly decline, the DSS group showed no noticeable change, which contradicts Truscott (1996).

Azizi and Nemati (2018b) also examined the effect of DSS on changes in learners' overall writing proficiency, measured using the TOEFL iBT writing scoring rubric, as well as the same three measures of fluency, accuracy, and grammatical complexity. Their findings showed that while both groups receiving corrective feedback could improve over time, the DSS group could significantly outperform the control group. Checking the same measure of fluency, the two groups demonstrated a significant improvement from pretest to posttest, but this time, the DSS group had outperformed the control group. For the measure of accuracy, the results were similar to those obtained in 2013. Finally, regarding the measures of grammatical complexity, in the case of one of the two indices, i.e., the ratio of clauses to T-units, similar results as those in the 2013 study were obtained. However, for the second index, i.e., the frequency of the dependent clauses, which is a more straightforward index, a trend was observed between the gains in the scores from pretest to posttest, with a higher gain for the DSS group. However, the change over time for both groups was also found significant.

In another study, Azizi and Nemati (2018a) studied the effect of DSS on the change in learners' performance on the IELTS writing test in terms of both the total score and the four components included. Using a questionnaire and conducting interviews, they also checked the participants' motivation, attendance to teacher feedback, and their feelings and attitudes toward the course they went through. Regarding their improvement in IELTS total score in writing, while the non-DSS group could improve by half a band score at the end of the course, the DSS group's improvement was found to be a 1.5 band score, which was statistically significant. The same pattern was observed regarding all the four rubric components. In addition, exploring the DSS group's level of motivation, attendance, and attitudes, they reported very positive results. Participants reported a high level of motivation (a mean of 3.98 out of 5) in attending to teacher feedback when undergoing the DSS system of scoring. The participants' attendance was also found very promising (a mean of 4.24 out of 5). The level of attendance could also be inferred from the rate of submission of the revised drafts in the two groups. While the rate of submission for the first draft was 98.07% and 93.60% for the DSS and the control groups respectively, it dramatically changed for the next drafts. The rates changed to 73.56% and 5.64%, respectively, for the second draft, and 54.96% and 1.83% for the third draft. Finally, the DSS group's attitude and feelings were found to be quite positive regarding their experience with the new technique (a mean of 4.01 out of 5). The interviews could also confirm the results obtained using the questionnaire as one of the participants stated:

The advantage of this method in comparison with other writing classes was that in my last classes whenever I got my paper and saw my mistakes, the only thing was that it made me disappointed. I did not focus on my mistakes carefully, whereas in this class I used to look at my mistakes and try to find the correct form in in order to [sic.] prepare a revision. I knew that I will learn more, also get a better score... it was very motivating. What I liked a lot about this system was that I became very meticulous in my latter writings. ... I may perform the same method of teaching if I become a teacher as well (Azizi & Nemati, 2018a, pp. 17-18).

In all the studies mentioned above, the effects of DSS have been examined using a comprehensive approach to error correction. In other words, all the errors, irrespective of their type, were underlined and students were required to correct them in the revisions they subsequently made. In the present study, however, the effect of DSS was investigated using a focused approach to error correction, i.e., one specific structure, subordinate clauses, was examined. As a result, it was sought to see if DSS could make any significant difference in the number of subordination clauses learners used as well as the accuracy of such use.

Method

Participants

Two intact groups at the University of Tehran were selected using convenience sampling to randomly act as the treatment and control groups. There were 27 participants (10 male and 17 female) in the treatment group and 28 in the control group (13 male and 15 female). Their age ranged from 20 to 27 with a mean of 22 (SD = 1.79). They were all undergraduate students studying English Language and Literature taking part in the “Advanced Writing” course as part of their curriculum. The data were collected in two consecutive semesters.

Instruments

The data used in data analysis were collected from learners’ writing samples at the pretest and posttest. To do so, the needed prompts were selected from among the list of sample TOEFL iBT writing prompts available on the ETS website. In addition, in order to check the homogeneity of the participants in the two groups and their comparability prior to the study, Oxford Quick Placement Test (OPT) was used. Due to the nature of the study and the fact that only intermediate learners of English were targeted for the purpose of the study, an English placement test was more appropriate than a proficiency one, and OPT was an appropriate choice both in terms of reliability and practicality.

Data Collection Procedure

Since the focus of the study, with a quasi-experimental design, was on learners’ use of subordination clauses, it was logical to select participants from a level of proficiency high enough for learners to be ready to master the structure and low enough not to have already mastered it. As a result, the intermediate level of proficiency was selected. After the administration of the Oxford Quick Placement Test, the intermediate participants in the two groups were identified, leaving the researchers with 55 participants in total, excluding 23 participants’ data from the final analysis, though it was not possible to exclude them from the whole study due to the nature of the program.

During the first three sessions of the study, both groups were taught the preliminaries of writing as well as a review of the grammatical points on subordination in English. The TOEFL iBT independent writing task was the reference for both teaching and assessing the quality of learners’ writings. The fourth session was devoted to gathering samples of the learners’ writing

as the pretest. A prompt was assigned to the participants to write about. They were given 60 minutes to plan and write at least 250 words on the given topic. TOEFL iBT task 2 writing rubric was used to score learners' writing samples. Two raters with at least 15 years of experience in teaching and assessing L2 writing rated the coded samples, which were already typed by the researchers in order to exclude any handwriting effect (Klein & Taub, 2005). The inter-rater reliability using Pearson-Product Moment Correlation Coefficient between the two raters was found to be 0.87, which is sufficiently high (Lange, 2011). Regarding the mismatches between the raters' scores, in cases where the discrepancy was not more than one score, a mean score was calculated, and in cases where the mismatch was more than one score, a third rater was employed and the mean between the two closest scores was used for data analysis. The scores were used to check the comparability of the two groups. No significant difference was observed between the two groups in their writing ability prior to the instruction, $t(53) = 0.81, p = 0.42$.

In each session, learners' samples were collected and students were asked to write about a new topic at home for the following session. The collected samples were graded, and corrective feedback was provided by the researchers. Only the grammatical mistakes made by the learners were commented on. If a sample had stylistic problems such as problems with topic development, cohesion, and coherence, it was only mentioned in the margin that the writer needed to improve it stylistically in terms of its topic development, cohesion, or coherence, for example. However, some of the samples which contained some kind of stylistic problems shared with most of the students were chosen and discussed with the whole class during the class hour in the following session.

For all kinds of mistakes but for their mistakes on subordination, direct corrective feedback was provided by the researchers. In other words, the mistakes were indicated in the form of underlining, and the correct forms were written above the wrong ones or in the margin. However, for their mistakes on subordination, indirect feedback was provided, that is, the errors were only indicated by being underlined. Correcting those forms was left to the learners themselves. The samples were then returned to the learners. This process continued for the whole course.

Up to this point, the procedure was the same for both groups. However, while the grades awarded to the essays written by the control group did not change as a result of the revisions made, the scores given to the essays written by the DSS group were temporary and draft specific, i.e., the treatment group could improve their grades through submitting the revised

version of their first draft based on the feedback the teacher had provided them with. The learners had two opportunities to improve their writing samples and improve their scores accordingly. Both groups were strongly advised to edit their works and hand the revised version to the instructor. The participants, in order to improve their essays, had to rewrite their first draft and correct their errors in subordination or improve them stylistically. In the last session, learners had 60 minutes to write about a new prompt as their posttest.

Participants' writing samples in the pretest and posttest were analyzed for the type, number, and accuracy of dependent clauses used. Two experts checked the coded texts for the number and accurate use of subordination clauses. The inter-rater reliability was found to be 0.89. The intra-rater reliability of the two were 0.87 and 0.93 and since the intra-rater reliability was more important in such studies (Chandler, 2003), the data of the rater with higher intra-rater reliability were used in data analysis.

In the case of the adverb clauses, 8 different types were checked: Time, Place, Reason, Purpose, Manner, Contrast, Condition, and Result. In the case of the adjective clauses, 4 different classifications were examined: the relative pronoun as the subject of the dependent clause, the relative pronoun as the object of the dependent clause, the relative pronoun as the object of preposition, and the adjective clause modifying the whole sentence. Finally, for noun clauses, four types were identified: noun clause as the subject, noun clause as the object, noun clause as the object of a proposition, and noun clause as the complement of an adjective.

Data Analysis

For data analysis, gain score procedure with followed up Mann-Whitney or Wilcoxon tests, where appropriate, were used due to the nature of the data which was in the form of frequency and its simplicity of presentation and interpretation. In a research design like the one used in the present study, the Within-Between Subjects Analysis of Variance was used. However, since that test had no non-parametric alternative and since the data in this study were of frequency type, gain analysis was employed (Salkind, 2010).

In addition, since the minimum number of words learners were supposed to write was pre-specified and since there was a time limit for it, it did not seem necessary to normalize the data in terms of length, i.e., to consider the number of dependent clauses in every 1,000 words, for example.

The collected data were analyzed at two levels. First, the total number of dependent clauses and the total number of accurate uses of dependent clauses were compared. Next, the

total number of adverb clauses, adjective clauses, and noun clauses and their accurate use were compared.

Results

Changes in the Total Number of Dependent Clauses

In order to examine the number and accuracy of the dependent clauses used by the participants, the gains in the total number of dependent clauses as well as their accuracy use were examined. Table 1 presents the descriptive statistics for both groups over time from pretest and posttest for both the number of clauses and the number of accurate instances of them used by learners as well as the gain in them over time.

Table 1. *Learners' Total Number and Accuracy of Dependent Clause Used and their Gain in Them*

	Group	N	Min.		Max.		Mean		Std. Deviation	
			No	Acc.	No.	Acc.	No.	Acc.	No.	Acc.
Treatment	Pretest	27	4	3	23	22	12.74	10.63	4.66	4.57
	Posttest	27	5	4	41	40	18.59	16.89	9.16	9.25
	Gain	27	-12	-10	24	25	5.85	6.26	7.68	7.99
Control	Pretest	28	5	5	24	20	12.71	10.54	4.38	3.80
	Posttest	28	7	5	27	25	14.32	12.11	5.19	4.66
	Gain	28	-6	-5	10	11	1.61	1.57	3.15	3.35

The pattern of results regarding the change in the number and accuracy of the dependent clauses used is almost the same. While both groups had started almost at the same level, the treatment group, undergoing DSS, finished the course with an average gain of 5.85 in the number of dependent clauses and a gain of 6.26 in the accurate instances of such a structure. However, this gain was much less in the case of the control group. This group could only show a gain of 1.61 in the number of clauses used and a gain of 1.57 in the accuracy of the used clauses.

The results of the Mann-Whitney tests which were run between the gain scores demonstrated a statistically significant difference between the two groups, with the treatment group outperforming the control group both in the number and accuracy of the dependent

clauses used. It is worth mentioning that no significant difference had been observed between the two groups in the pretest. Table 2 presents the results of the Mann-Whitney tests.

Table 2. *Mann-Whitney Results for Gains in the Number and Accuracy of Dependent Clauses*

	Pretest Total Use	Pretest Accuracy	Gain inUse	Gain in Accuracy
Mann-Whitney U	377.50	373.50	200.50	212.50
Z	-.008	-.076	-3.002	-2.797
Sig. (2-tailed)	.993	.939	.003	.005

Interestingly enough, the gains for both groups in the number and accuracy of the dependent clauses were found statistically significant from pretest to posttest according to the results of the Wilcoxon Signed Ranks Test, $z_{use} = -3.00$, $p = .00$, $z_{accuracy} = -2.8$, $p = .01$. However, the number of positive ranks in the treatment group from pretest to posttest was much higher than that of the control group, which indicates that DSS could result in a more homogenous and successful group. Table 3 presents the related statistics.

Table 3. *Wilcoxon Signed Ranks Test Results for the Two Groups*

Group		Posttest – Pretest No. of Clauses	Posttest– Pretest Accuracy
Treatment	Z	-3.460	-3.573
	Sig. (2-tailed)	.001	.000
Control	Z	-2.582	-2.334
	Sig. (2-tailed)	.010	.020

Changes in the Total Number of Adverb, Adjective, and Noun Clauses

At the second level, the gains in the use of adverbs, adjectives, and noun clauses were checked separately. The two groups initiated the instruction almost at the same level with no significant difference in the total number and accuracy of the adverbs, adjectives, and noun clauses at the pretest. The related descriptive statistics are presented in Table 4.

However, their improvement in each category was different for each group. While a significant improvement was observed in the case of almost all categories both in the number

and accuracy of use for the treatment group (the only exception was in the case of the total number of adjective clauses for which a trend was observed, $z = -1.89$, $p = .058$), for the control group, a significant improvement was observed only in the case of the number of accurate instances of adverb clauses ($z = -1.97$, $p = .05$) and the number of used noun clauses ($z = -2.13$, $p = .03$). For the rest of the categories, no significant improvement was observed for the control group. Table 5 presents the results for the comparison between the studied indices in the pretest and posttest for both groups.

Table 4. *The Number and Accuracy of Adverbs, Adjectives, and Noun Clauses Used*

Group		N	Min.		Max.		Mean		Std. Deviation	
			No.	Acc.	No.	Acc.	No.	Acc.	No.	Acc.
Treatment	Pretest Adv.	27	0	0	10	7	4.11	3.63	2.12	1.80
	Posttest Adv.	27	1	1	15	15	6.44	5.81	3.61	3.62
	Pretest Adj.	27	1	1	13	12	5.41	4.41	2.76	2.58
	Posttest Adj.	27	0	0	14	14	6.74	5.93	3.46	3.40
	Pretest Noun	27	0	0	8	8	3.22	2.59	2.31	2.27
	Posttest Noun	27	0	0	19	18	5.41	5.15	5.26	5.07
Control	Pretest Adv.	28	1	0	9	7	3.68	3.00	1.93	1.76
	Posttest Adv.	28	1	0	11	11	4.43	3.68	2.63	2.52
	Pretest Adj.	28	1	1	13	9	5.21	4.29	2.77	2.27
	Posttest Adj.	28	1	1	12	11	5.21	4.36	2.94	2.63
	Pretest Noun	28	0	0	11	10	3.82	3.25	2.33	2.12
	Posttest Noun	28	0	0	10	10	4.68	4.07	2.37	2.36

Table 5. *Wilcoxon Signed Ranks Test Results for the Two Groups from Pretest to Posttest*

Group		Total	Accurate	Total	Accurate	Total	Accurate
		Adverb Clauses	Adverb Clauses	Adjective Clauses	Adjective Clauses	Noun Clauses	Noun Clauses
Treatment	Z	-2.87	-2.81	-1.89	-2.38	-2.07	-2.47
	Sig. (2-tailed)	.004	.005	.058	.017	.039	.013
Control	Z	-1.78	-1.97	-.133 ^b	-.414 ^a	-2.13	-1.81
	Sig. (2-tailed)	.075	.048	.894	.679	.033	.069

Comparing each group's gain in each category, one can see that the treatment group undergoing DSS could gain more in each category in comparison with the control group both in the case of the number of the utilized dependent clauses and their accuracy to the extent that this gain was zero for the control group in the number of adjective clauses. Table 6 presents the related descriptive statistics. Table 7 presents the Mann-Whitney test results for the gain scores.

Table 6. Learners' Gains in the Number and Accuracy of Adverbs, Adjectives, and Noun Clauses

Group		N	Min.		Max.		Mean		Std. Deviation	
			No	Acc.	No.	Acc.	No.	Acc.	No.	Acc.
Treatment	Gain in Adv. C.	27	-5	-5	10	11	2.33	2.19	3.52	3.61
	Gain in Adj. C.	27	-6	-5	7	6	1.33	1.52	3.51	3.00
	Gain in Noun C.	27	-5	-5	15	14	2.19	2.56	4.77	4.77
Control	Gain in Adv. C.	28	-3	-3	6	5	.75	.68	1.92	1.68
	Gain in Adj. C.	28	-2	-2	4	3	.00	.07	1.54	1.49
	Gain in Noun C.	28	-4	-3	6	7	.86	.82	2.05	2.18

Table 7. Comparison of the Two Groups in their Gain Scores

	Adv. C.	Adv. C. Accuracy	Adj. C.	Adj. C. Accuracy	Noun C.	Noun C. Accuracy
Mann-Whitney U	256.50	270.50	257.00	250.00	334.00	310.00
Wilcoxon W	662.50	676.50	663.00	656.00	740.00	716.00
Z	-2.078	-1.840	-2.055	-2.194	-.747	-1.153
Sig. (2-tailed)	.038	.066	.040	.028	.455	.249

As it is evident above, the DSS group outperformed the non-DSS group in the number of adverb and adjective clauses they used. However, no difference was observed regarding the number of used noun clauses. Considering the number of accurate instances of such clauses, the results were similar but for the fact that in the case of accurate use of adverb clauses, a trend was observed, $U = 270.5$, $z = -1.84$, $p = .06$.

The change in the number and accuracy of different types of subordinate clauses in the adverb, adjective, and noun clause categories were also examined, but due to the overwhelming details of that investigation, it could not be presented in this paper.

Discussion

The results of the present study, which was an investigation of the effect of DSS on learners' use of subordinate clauses, showed that while both groups improved over time in terms of the total number and accuracy of instances of subordinate clauses in general, the DSS group significantly outperformed the control group. In the case of adverbs, adjectives, and noun clauses, it was observed that the two groups differed in the gains each could accomplish regarding the number and accuracy of the adverb and adjective clauses but not the noun clauses, with the treatment group outperforming the control group in both cases. All these indicate that Draft Specific Scoring has an advantage over the more traditional feedback provision approaches in helping learners improve their use of subordinate clauses (also see Nemati & Azizi, 2013; Azizi & Nemati, 2018a, 2018b).

However, the effect of DSS can be viewed from two different angles. One is the difference between the two groups in terms of gains they achieved in the number and accuracy of the dependent clauses used, which was what we did above. From the second point of view, it is also possible to examine each group's improvement from pretest to posttest in each type of dependent clauses separately.

Regarding the total number of adverbs, adjectives, and noun clauses used by learners, the treatment group could significantly improve in all the three categories over time. The fact that the DSS group could improve in more categories than did the control group shows the superiority of this technique or better to be called grading system. However, it implies other points. A question that arises is why neither group could improve in some of the categories. Different reasons could be speculated. First of all, the short duration of the course could be one determining factor. The low number of assignments learners could write (they wrote 10 essays including both the pretest and posttest) might have been inadequate to give them the chance to

practice such structures. In addition, for most types of dependent clauses, learners had started the course with a very low level of using such structures. When learners do not use a structure, they won't receive any corrective feedback on them, and as a result, they won't be able to improve them. For instance, further analysis showed that in the case of the noun clauses, participants in both groups used the object noun clause (e.g. I know *that you are right.*) much more frequently than the other three types both in the pretest ($M_{\text{treatment}}=2.89$; $M_{\text{control}}=3.29$) and posttest ($M_{\text{treatment}}=4.70$; $M_{\text{control}}=4.14$). Object noun clauses are the most common type among others in everyday use, and learners often master this category without even knowing it is a noun clause. However, the other three types are less frequently used and encountered in everyday language use, and they require instruction. The low number of use both in the pretest and posttest confirms that, with the means being very close to zero, ranging between 0.11 and 0.21 in the pretest and 0.29 to 0.33 in the posttest.

As such, it seems that those structures occurring more frequently have a better chance for improvement. It seems that learners need at least a basic knowledge of a structure in order for corrective feedback to be effective. Moreover, it is plausible to conclude that different structures are affected differently by corrective feedback. While some show a considerable improvement, some others may show a low or moderate improvement, with some other structures showing no improvement at all.

A second point concerns the fact that the control group simply receiving corrective feedback could not improve that much and in fact, showed a significant decline in one case. This could to some extent confirm Truscott's (1996) thesis regarding the ineffectiveness or even harmful effect of TCF. However, since even the control group could show a significant improvement in some other cases, Truscott's claim faces counter evidence too. As such, it seems that a weaker version of Truscott's (1996) thesis should be put forward. It seems that TCF may affect different structures differently; in some cases, it may aid learners to improve, in other cases, it could turn out to be ineffective, and still, in rare cases, it might be harmful. It is a matter of interaction among different variables involved. That might be why the literature is full of studies each coming up with different results; while some find TCF quite effective in general or for some specific structures, others find it ineffective (Ferris, 2004; Guenette, 2007). However, the fact that the group undergoing DSS could improve in structures that the control group did not display any progress or even showed a decline in performance indicates that such a lack of positive effect for corrective feedback could be due to other external reasons such as

learners' lack of motivation to attend to teacher feedback rather than the unpleasantness of the corrective feedback as claimed by Truscott (2007).

One may consider DSS an offspring of the process approach to L2 writing. In the process approach, instead of concentrating on the final draft or product, the attention is centered on the mid drafts as in DSS. Mid drafts and the provision of feedback on them are so important that if TCF is supposed to work, it is believed that it will do so on mid drafts (Muncie, 2000). In addition, interactional feedback is often associated with second language learning as it motivates learners to notice second language forms (Long, 2006; Mackey & Oliver, 2002; McDonough, 2005). According to Long (1996, p. 451-452), interaction plays a significant role in associating "input, internal learner capacities, particularly selective attention, and output in productive ways". As a result, processes involved in feedback provision on the mid drafts can and may facilitate the process of language learning. Meaning negotiation and recasts provision are among such useful processes. These processes may also result in modified output (Swain, 2005), which is also instrumental in language learning (Mackey, 2006).

Although DSS may look like portfolio assessment in L2 writing, it is in fact quite different. "While, in portfolio, the focus is on the process of writing to reach the final product, DSS works with the products to strengthen the processes involved in developing" L2 writing proficiency" or learning an L2 structure. In addition, unlike portfolio assessment in which delayed evaluation is stressed, immediate evaluation is the cornerstone in DSS. Moreover, instead of collecting students' works over a long period of time as in portfolio assessment, no selection is carried out in DSS and each writing sample is packed away after at most three weeks. In other words, "instead of defining long term objectives, we invest on short-term objectives in DSS in order to achieve the long-term objective by the end of the course of instruction" (Azizi & Nemati, 2018b, pp. 104-105).

DSS is a technique in assessment to assure students' attention to and noticing of teacher feedback. Since learners are required to revise their first and mid drafts in DSS, it can also help encourage modified output, which is often the result of the negotiation of meaning between the learners and the teacher as it is not always easy to understand teacher feedback or her intention, for example, in underlining a piece of text. While making sense of obscure teacher comments on the margins or understanding her intention in some other cases may cause frustration and the abandonment of the writing sample on the part of the learners in usual methods of feedback provision, in DSS, since motivated enough, learners seek teacher intention on such cases which can strengthen the negotiation of the meaning between them. As a result, it is not unsafe to

claim that Draft Specific Scoring has the potential to implement most if not all the crucial processes involved in learning second language writing through benefiting from teacher corrective feedback.

Conclusion

DSS helps teachers carry on their practice based on the underlying principles they believe in. Those who argue for the negative effect of scoring may argue for its abandonment. However, instructors often express their strong belief in marking learners' writing samples. They consider grading a crucial element in the feedback provision process (Li & Barnard, 2011). Grading can aid instructors to come up with a better picture of their students' abilities by the end of the program (Lee, 2009). Although they know about the adverse effect grading may have on students' attendance to teacher feedback, teachers still continue their practice mainly due to not only their belief in the importance of grading but also their sense of obligation for summative evaluation. Students, too, strongly demand such an evaluation as they assist them in having a better evaluation of their own performance. In addition, the interpretation of grades is often much easier than teachers' elaborate comments (Lee, 2008). In case instructors keep on grading, they risk learners' attendance to teacher feedback, and if they stop doing so, they will have to face new challenges.

Draft Specific Scoring gives the language instructors the opportunity to adhere to their preferred practices while mitigating the adverse impact of grading and transforming its weakness into strength. DSS does not distract learners' attention from TCF; instead, it motivates learners to pay more attention to it. DSS combines assessment with instruction without overlooking any. In DSS, scoring students' writing is not the last step in the writing instruction but it is the beginning of the revision cycle. In DSS, learners learn that it is their responsibility to focus on the feedback accompanying the grade they receive in order to improve it. This way, Hamp-Lyons' (2007) concern is also addressed, and writing assessment cannot take over writing instruction.

This grading system has shown to be successful in motivating students to not only attend to the feedback they receive but also use it in their future assignments. It has also been shown that learners feel relaxed and more confident when DSS is used as part of their instruction. They have a more positive attitude toward it and enjoy writing when the writing assessment is accompanied by DSS (see Azizi & Nemati, 2018a, 2018b).

The result of the present study indicates that in case one wishes to achieve course objectives in a writing program and help students improve in accuracy in their writing samples, s/he cannot ignore the very significant role of motivation on the part of learners to attend to teacher feedback. No matter if teachers make use of DSS or any other technique, what is important is that they ensure students' attendance to the feedback they provide them with. This study indicates that there could be variables mediating the effectiveness of TCF, and motivation is only one of them. Future studies need to look for such variables and try to find a solution for each. Teacher feedback works. If it does not, one needs to look for what it is that hampers it (Bruton, 2009).

The present study was not without flaws. The low number of sessions participants could attend the program and the low number of writing samples they wrote could have affected the observed results. The more samples students produce, the more TCF they may receive on their mistakes, which enhances the possibility of learning as a result. Certainly, prolonging the program in future studies can provide a better picture of the effect of DSS. In addition, the inclusion of a larger sample may help with a better evaluation of the present technique. Finally, the focus of the present study was on subordination clauses. DSS may show a different effect on other language structures or under different conditions. Future studies may address these limitations.

Conflict of Interest

The authors declare no conflict of interest to report.

References

- Azizi, M., & Nemati, M. (2018a). Draft specific scoring and teacher corrective feedback: Hearing learners' voice. *Journal of Modern Research in English Language Studies*, 5(4), 1-26.
- Azizi, M., & Nemati, M. (2018b). Motivating the unmotivated: Making teacher corrective feedback work. *Issues in Language Teaching*, 7(1), 87-110.
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge.
- Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were. *System*, 37(4), 600-613.

- Bruton, A. (2010). Another reply to Truscott on error correction: Improved situated designs over statistics. *System*, 38, 491-498.
- 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.
- Connors, R. J., & Lunsford, A. A. (1993). Teachers' rhetoric comments on student papers. *College Composition and Communication*, 44(2), 200–223.
- Diab, N. M. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter?. *Assessing Writing*, 24, 16–34.
- Diab, R. L. (2005). EFL university students' preferences for error correction and teacher feedback to writing. *TESL Reporter*, 38(1), 27-51.
- 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. (2004). The “grammar correction” debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime?). *Journal of Second Language Writing*, 13(1), 49-62.
- Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22(3), 307-329.
- Guenette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16(1), 40–53.
- Hamp-Lyons, L. (2007). Editorial. *Assessing Writing*, 12(1), 1–9.
- Han, Y. (2017). Mediating and being mediated: Learner beliefs and learner engagement with written corrective feedback. *System*, 69, 133-142.
- Han, Y. (2019). Written corrective feedback from an ecological perspective: The interaction between the context and individual learners. *System*, 80, 288-303.
- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31-44.
- Han, Y., & Hyland, F. (2019). Academic emotions in written corrective feedback situations. *Journal of English for Academic Purposes*, 38, 1-13.
- Klein, J., & Taub, D. (2005). The effect of variation in handwriting and print on evaluation of student essays. *Assessing Writing*, 10(2), 134-148.
- Lange, R. T. (2011). Inter-rater reliability. In J.S. Kreutzer, J. DeLuca, & B. Caplan. (Eds) *Encyclopedia of clinical neuropsychology* (pp. 2368-2371). New York: Springer.

- Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(3), 144-164.
- Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal, 63*(1), 13-22.
- Lee, I. (2014). Feedback in writing: Issues and challenges. *Assessing Writing, 19*(1), 1-5.
- 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*(2), 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.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405-530.
- Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System, 30*(4), 459-477.
- McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition, 27*(1), 79-103.
- Moradkhani, S., & Goodarzi, A. (2020). A case study of three EFL teachers' cognition in oral corrective feedback: Does experience make a difference?. *Issues in Language Teaching, 9*(1), 183-211.
- Muncie, J. (2000). Using written teacher feedback in EFL composition classes. *ELT Journal, 54*(1), 47-53.
- Nemati, M., & Azizi, M. (2013). Grading, no longer an obstacle to learners' attendance to teacher feedback. *Applied Research on English Language, 2*(2), 129-143.
- Radacki, P. M., & Swales, J. M. (1998). ESL student reaction to written comments on their written work. *System, 16*(3), 355-365.
- Salkind, N. J. (2010). *Encyclopedia of research design*. Thousand Oaks, CA: Sage Publication.

- Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, 32(2), 303-334.
- 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(2), 327–369.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272.
- Truscott, J. (2010). Further thoughts on Anthony Bruton's critique of the correction debate. *System*, 38(4), 626-633.
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90-102.
- Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, 37, 13-24.