

The Long-Term Effect of Selective Written Grammar Feedback on EFL Learners' Acquisition of Articles

Ahsan Pashazadeh*

MA Student, English Department, University of Tehran, I.R. Iran

Hamideh Marefat**

Associate Professor, English Department, University of Tehran, I.R. Iran

(Received: 12 Feb. 2009, Accepted: 10 Aug. 2009)

Abstract

The long-term efficacy of written grammar feedback (WGF) has escaped rigorous empirical investigation, partly due to its strong intuitive appeal. The few studies that have investigated the long-term effect of grammar correction have failed to find any positive long-term effects for WGF (Truscott, 2007). Nevertheless, a number of recent studies have reported a positive effect for WGF. The present study investigated the long-term effect of selective grammar feedback on a reasonably complex feature of the English grammar, attempting to shed some light on the factors that may explain the conflicting results of previous studies. A group of low intermediate EFL learners (N = 22) participated in this study. They received WGF on the use of articles and a limited number of various other grammatical categories. The results showed that selective WGF can produce large short-term gains for functionally complex grammatical features, but that it may prove to be detrimental in the long run. The study also found that learners tend to avoid the grammatical feature on which they have received corrective feedback. The theoretical and practical implications are discussed.

Key Words: Grammar Feedback, Accuracy, Item Learning, System Learning, Articles.

* Tel: 021-61119026, Fax: 021-88634500, E-mail:pashazade@ut.ac.ir

** Tel: 021-61119026, Fax: 021-88634500, E-mail:marefat@ut.ac.ir

1. Introduction

Several decades of research on grammar feedback have barely yielded any satisfactory evidence regarding the long-term efficacy of providing second language writers with corrective feedback (for reviews of the related studies see Hyland & Hyland, 2006; Norris & Ortega, 2000; Truscott, 1996). The strong claims concerning the positive effects of corrective feedback may have great intuitive appeal; however, these claims have been rebutted on the grounds that the primary studies carried out so far, which have found a positive effect for WGF, have not been designed well enough to provide convincing evidence for grammar feedback.

Believing that feedback is essential to help students recognize their linguistic shortcomings, language teachers spend a great deal of time offering corrective feedback on the written products of their students (Ferris, 1999, 2003; Lee, 2004; Truscott, 1996; Zamel, 1985). Writing teachers' unfailing faith in the efficacy of corrective feedback is shared by the majority of ESL theoreticians, a belief fuelled by the surge of the cognitive accounts of second language acquisition. Researchers working within the cognitive framework invoke the noticing hypothesis and the role of conscious learning to claim that short-term gains in accuracy can be taken as a sign of noticing, which is interpreted as the beginning of learning. Nevertheless, short-term gains may not translate into long-term grammatical accuracy.

In 1996 Truscott published a controversial critical review of feedback studies in which he cited evidence from various studies (such as Frantzen, 1995; Kadia, 1988; Kepner, 1991; Lalande, 1982; Robb, Ross, & Shortreed, 1986; Semke, 1984; Sheppard, 1992; Zamel, 1985) and concluded that feedback is not, and should not be expected to be, effective in the long term. Several researchers (Chandler, 2003; Ferris, 1999, 2003; Ferris & Roberts 2001, to name but a few) criticised him for his strong claims. However, Truscott has not withdrawn his case against grammar correction. In several subsequent papers, Truscott (1999, 2004, 2007) has made it clear that feedback studies do not support any positive effects for grammar correction.

Several years later, the debate still continues between the advocates and

opponents of corrective feedback. In fact, since the publication of Truscott's review article a number of other studies have found corrective feedback to be ineffective (e.g., Fazio, 2001; Polio, Fleck, & Leder, 1998). Nevertheless, Bitchener (2008), Bitchener and Knoch (2008), Ellis, Sheen, Murakami, and Takashima (2008), Rahimi (2009), and Sheen (2007) have provided some evidence that corrective feedback can be effective in improving the accuracy of L2 writers.

A number of feedback studies have investigated the effect of corrective feedback on several grammatical and non-grammatical categories (e.g., Ferris, 2006). However, as Norris and Ortega (2000) have pointed out, studies that try to handle too many variables are not likely to add much to our knowledge base. In the same vein, the effects of error correction have sometimes been measured by grammaticality judgement tests or multiple-choice grammar tests, which clearly favour students who have received formal instruction (see Ellis, 2005). As far as L2 writing is concerned, writing original essays is deemed to be the most authentic measure of writing abilities.

Almost all feedback studies suffer from similar shortcomings – namely, lack of truly equivalent groups (e.g., Bitchener, Young, & Cameron, 2005), inappropriate definition of error categories (e.g., Ferris & Roberts, 2001), lack of a pre-test to measure initial differences (e.g., Kepner, 1991), lack of a delayed post-test to measure long-term effects (e.g., Fathman & Whalley, 1990), and use of inappropriate testing devices (e.g., Cardelle & Corno, 1981).

Ellis *et al.* (2008) studied the differential effects of focused and unfocused WGF, including a control group to investigate Truscott's (1996) thesis regarding the inefficacy of corrective feedback. The results of the picture story tasks revealed that providing the learners with WGF helped them to improve their use of two functions of the English article system – namely, first- and second-mention use of articles to refer to unknown and known entities. The experimental groups outperformed the control group and maintained their level of accuracy after 4 weeks.

The present study was inspired by Ellis *et al.*'s (2008) research; however, there were a number of shortcomings in Ellis *et al.*'s study which the present research has

attempted to eliminate. First of all, the elicitation device used in this study was a picture story task, a measure that could understandably have introduced a memory factor into the students' writing. In other words, the participants were given a set of pictures (as well as a constructed story on separate sheets of paper, which were collected before the students started writing their narratives) and were instructed to rewrite "the story with as much detail as they could remember" (p. 358). In the present study, controversial topics were given to the students, asking them to express their own ideas on the topics. This measure was taken in order to enhance the validity of the writing tasks.

Second, Ellis *et al.*'s study used various elicitation devices, including a grammaticality judgement test as well as a questionnaire. These devices were used to triangulate the results. However, the use of a measure like a grammaticality judgement test may give away the purpose of the study. Ellis *et al.* failed to detect any significant differences between their focused and unfocused treatment groups. Probably, the two treatment groups had similar performances due to the use of the grammaticality judgement test, which made both groups equally conscious of the aim of the study.

Third, Ellis *et al.*'s study used only one delayed post-test to assess the accuracy of the participants' writing. Given the U-shaped course of interlanguage development, it is befitting to use a second post-test to better trace the possible long-term effects of the treatment. The English article system is too complex to be significantly affected in a short period of time (Bitchener, 2008; Butler, 2002; Master, 1997). From a cognitive perspective, formal instruction reveals its effects through a process called "restructuring" (see McLaughlin, 1990). This cognitive process is believed to take place as a delayed result of instruction and negative evidence.

Fourth, the study did not account for the overuse of articles. When learners are taught a rule, they tend to overgeneralise it (or avoid the structure altogether). The results of feedback studies have generally been based on the calculation of *obligatory occasion analysis*, which does not take overuse into account. Sheen

(2007) is an exception in this regard. *Obligatory occasion analysis*, therefore, leads to overestimation of the accuracy index. *Target like use* is a measure that can help reduce the possibility of this overgeneralization pitfall (Ellis & Barkhuizen, 2005).

As Ferris (2004) maintains, due to lack of well-designed studies to address the efficacy (or inefficacy) of WGF, "we are virtually at Square One" (p. 56). The present study is a longitudinal attempt to shed some light on the role of grammar correction in helping EFL writers to improve their writing accuracy. The scope of this discussion will be limited to the role of teacher written feedback mainly because it is the major type of feedback offered in EFL contexts.

2. Purpose of the Study

Studies are normally designed to provide an answer to a particular question of some theoretical importance. Before the publication of Truscott's (1996) controversial article in *Language Learning*, few researchers questioned the role of grammar feedback. Therefore, there are a limited number of studies that have addressed the questions raised by Truscott in a direct and careful manner.

In fact, all the recent studies (i.e., Bitchener, 2008; Bitchener & Knoch, 2008; Sheen, 2007) that have found a significant positive effect for grammar correction in the long term have focused on the use of first- and second mention functions of the English article system. Given the fact that the focus of these studies was formally and functionally simple (Liu & Gleason, 2002; Truscott, 2001), it is difficult to see how the results contradict the thesis advanced by Truscott (1996).

The present research was an attempt to address the shortcomings of Ellis *et al.* and other recent WGF studies. Specifically, the current study was conducted to investigate the long-term effects of providing grammar feedback on students' original writing¹. A second post-test was included to help better trace the U-shaped course of interlanguage development. Moreover, the entire article system of English was selected to guard against the possibility that formal and functional simplicity of the targeted error categories may render the results theoretically and practically insignificant.

3. Research Questions

This study was designed to answer the following questions:

- 1- Does WGF help EFL writers to produce more accurate forms of the treated categories?
- 2- Does WGF produce long-term effects?
- 3- Does provision of WGF affect avoidance of the treated grammatical category?

4. Methodology

4.1. Participants

This study was carried out in a private language school in Tehran. 22 graduate and undergraduate students – 13 male and 9 female – from an intact low intermediate class² participated in this study. They were enrolled in a "conversation class" taught by the researcher. The students met two times a week for 105 minutes. The course had a writing component, although the main focus was on developing conversation skills. The age of the participants ranged from 19 to 34, with the mean age of 23. Out of 22 students, only 15 completed all the writing tasks. Students with incomplete data sets were excluded from the final data analysis.

4.2. Instrumentation

The instruments used for this study consisted of four argumentative writing tasks. Care was taken not to include any topic that would dissuade the use of articles. Moreover, the rubrics were carefully written to encourage the students to use articles in their writing. However, it should be noted that it is very difficult, if not impossible, to force students to use certain structures. Even if it were possible to design such a task, it would not be very desirable, as such a task would pose serious threats to the originality of students' writing.

Moreover, due to the nature of the research questions and practical limitations, it was not possible to include several measures of accuracy. It would have been highly

desirable if the results of the writing tasks had been triangulated through other measures of linguistic accuracy. However, it should be noted that the inclusion of such measures would have introduced other confounding variables into the design of the study.

4.3.Procedure

In the second session of the course, the students were given a reading text on a controversial topic and were instructed to read it in approximately 5 minutes. It was explained to the students that they would be required to write their own ideas about the subject introduced in the reading text. The students were allowed to ask questions in case they had difficulty understanding the reading passage. The passages were constructed in a careful manner to convince the students that it was possible to argue without using low frequency words and complex grammatical structures. After the reading passages had been collected, the students received the writing task instructions at the top of a separate sheet of paper. The students were required to write an argumentative essay in 25-30 minutes on the same topic as that of the reading text.

The essays were photocopied and subjected to data analysis, which yielded the pre-test scores. The original essays were corrected and returned to the students the next session. The participants were provided with the correct forms as well as a general content comment at the end of their essays and were asked to take a look at the form and content comments for about 5 minutes. Afterwards, they received the second reading passage, and the same procedure was adopted for the next two essays. The third essay was analysed for accuracy, the results of which are reported as the immediate post-test. Another task (i.e., delayed post-test 1) was administered with a four-week interval. The students were not told when they would be required to write another essay so that they could not prepare themselves for the writing task. As the second delayed post-test, the same pre-test writing task was administered four weeks after the first delayed post-test (the significance of this measure will be explained below). On the two delayed post-tests, the students received neither

content nor grammar feedback. The timetable is displayed in Table 1.

Table 1: Data collection timetable

Session	Testing and feedback
2 (week 1)	The first writing task (i.e., pre-test)
3 (week 1)	feedback on the first writing task; the second writing task
4 (week 2)	feedback on the second writing task; the third writing task (i.e., immediate post-test)
5 (week 2)	feedback on the third writing task
13 (week 6)	The fourth writing task (i.e., delayed post-test 1)
21 (week 10)	the fifth (same as the pre-test task in terms of content and instructions) writing task (i.e., delayed post-test 2)

The students received up to seven corrections on each essay, two of which were always article errors. Of course, it must be noted that some essays did not contain two article errors or seven errors on the whole, in which case as many errors as possible were corrected. The participants also received a single general comment on the content of their writing – such as 'Good' or 'This is really persuasive' or 'Pay more attention to the ideas'. Moreover, care was taken not to provide any oral or written feedback on the use of articles during the two four-week intervals.

4.4 Scoring

Target like use (see Sheen, 2007) was utilized in order to answer the first and second research questions (the number of articles used correctly divided by all the contexts in which the target language would require a particular article plus the number of articles oversupplied in non-obligatory contexts). Based on this analysis, each essay was assigned a score, which was calculated as a percentage. These scores were fed into SPSS, and paired samples t-tests were run to see whether there were any significant differences at different testing points.

In order to answer the third research question, obligatory occasions were identified on the written products of the students (i.e., all the contexts in which the target language would require a particular article). If feedback causes learners to avoid the treated categories, then learners should avoid the contexts that necessitate the use of the treated grammar category. Therefore, the number of obligatory occasions can be used to shed some light on whether and to what extent the participants tend to avoid the target grammatical category (Ellis *et al.*, 2008).

All the essays from the four testing times were scored collaboratively with an MA student of TEFL. The essays were corrected independently by the two raters. Having discussed the points of disagreement, perfect agreement (i.e., 100%) was reached. Sometimes, due to numerous grammatical and lexical errors, it was not clear whether or not an article should have been used. Therefore, the contexts that did not provide any clear clue about the use of the correct articles were left out of the final analysis. This decision was made in order to reduce the effect of subjective scoring. All the 15 essays from the pre-test were rescored by the author 14 weeks after the initial scoring, and the Pearson Product Moment Correlation yielded a reliability score of .91.

5. Results and Discussion

The descriptive statistics from the four testing times are displayed in Table 2 and graphically illustrated in Figure 1 below. Table 2 shows that the mean accuracy scores increased substantially from time 1 to time 2 (i.e., from pre-test to immediate post-test), only to decline drastically at the two delayed testing points (i.e., delayed post-tests 1 and 2). The paired samples t-test revealed that there was a statistically significant difference between pre-test and immediate post-test ($t_{(14)} = 4.163$; $p = .001$). As has been demonstrated by other feedback studies, this finding corroborates the observation that corrective feedback is effective in the short run.

Table 2: Descriptive statistics: accuracy over time

	N	Mean	Std. Deviation
Pre-test	15	35.64	12.40
Immediate post-test	15	53.86	17.95
Delayed post-test 1	15	38.66	17.31
Delayed post-test 2	15	32.60	19.18

The fact that corrective grammar feedback is effective in the short run, however, does not guarantee that it will be effective in the long run. Two more paired samples t-tests were conducted to reveal the differences between pre-test and delayed post-test 1 on the one hand and between pre-test and delayed post-test 2 on the other. The results of these tests revealed that there were no significant differences between pre-test and delayed post-test 1 ($t_{(14)} = 0.698$; $p = .497$) and between pre-test and delayed post-test 2 ($t_{(14)} = 0.541$; $p = .597$). This finding lends further credence to Truscott's (1996) "correction-free approach" (p. 116).

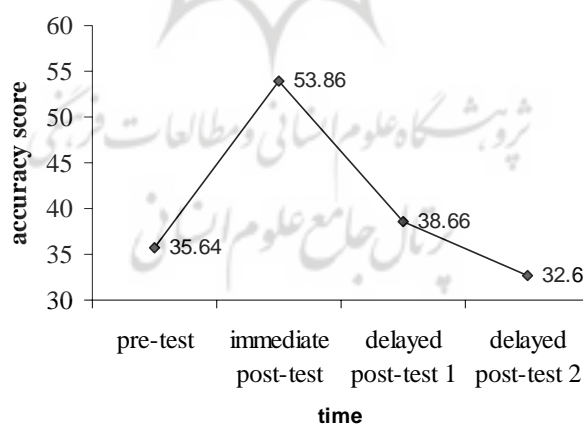


Figure 1: Accuracy over time

As pointed out by Norris and Ortega (2000) and Truscott (2007), tests of statistical significance do not show magnitude of the differences observed among different variables. Thus, Cohen's d was selected to test the effectiveness of the treatment in the present study. The results are presented as effect sizes in Table 3. Following Cohen (1992), the value of d will be interpreted as follows: small effect = 0.20 to 0.50, medium effect = 0.50 to 0.80, and large effect = 0.80 and higher.

Table 3: Effect sizes

Pre-test and immediate post-test	Pre-test and delayed post-test 1	Pre-test and delayed post-test 2
$d = 1.18$	$d = 0.20$	$d = -0.19$

According to Table 3, there is a large effect size for feedback on the immediate post-test, which almost disappears on the first delayed post-test. The effect size for delayed post-test 1 may be interpreted as evidence for grammar correction; however, it should be noted that this effect size is not substantial enough to support this interpretation. In fact, Norris and Ortega (2000) reported that the control groups in their meta-analysis displayed similar gains: $d = 0.30$. Thus, the gains in accuracy can be attributed to writing practice (see Rahimi, 2009), maturation, and other factors.

Table 3 also shows that there was a very small negative effect for feedback on the second delayed post-test. This finding is consistent with the results of some previous corrective feedback studies (e.g., Fazio, 2001; Sheppard, 1992). Corrective feedback – and instruction in general – causes drastic improvement in accuracy levels of learners in the short run. Nevertheless, as Krashen (1981) and Truscott (1996) argue, these changes are caused by learners' metalinguistic knowledge and conscious control over their output. These superficial changes do not reach the learners' competence, which may be one reason why they tend to disappear over time.

Considering the fact that the second post-test was the same as the pre-test

writing task in terms of the reading passage and the instructions, the negative effect size on the second post-test should be particularly alarming to the proponents of WGF. There is some evidence to suggest that learners cannot transfer instruction to other contexts (James, 2008). That is to say, task similarity (or difference) is an important factor that can affect learners' performance. For instance, instruction may help learners to use the English articles accurately in the narrative mode, but they may fail to use them accurately in another task that involves argumentative writing.

Ellis (1999) distinguished between system learning and item learning. He suggested that learners learn through both systematic internalisation of rules (system learning) and memorization of isolated items (item learning). The English article system is functionally complicated (Butler, 2002; Liu & Gleason, 2002; Master, 1997). It may be argued that the brief treatment in this study could not have contributed to the acquisition of the article system, which demonstrates several idiosyncratic, culture-based characteristics. In fact, the second delayed post-test was administered with this in mind. Assuming that grammar feedback is effective, learners should be able to avoid making the same mistakes on an identical task. The findings of this study are significant in that they fail to support any long-term effect for both system and item learning.

It has been argued that learners tend to avoid the categories that have been the subject of corrective feedback (Sheppard, 1992; Truscott, 1996, 2004). Semke (1984) also found that students who received correction simplified their essays in order to avoid being corrected. Truscott (1996) persuasively argued that this avoidance strategy thwarts natural language acquisition simply because grammar feedback encourages students to avoid risk-taking and experimenting with the grammatical forms that they have not fully mastered.

The third research question in the present study was formulated to provide some evidence on whether or not students avoid the treated grammatical categories. To this end, the obligatory occasions were counted. Following Ellis *et al.* (2008) and Truscott (2007), it was assumed that if feedback causes learners to avoid the categories that have been corrected, the number of obligatory occasions should

decrease over time. Table 4 shows the mean obligatory occasions in this study for the four testing times.

Table 4: Mean obligatory occasions

	N	Mean	Std. Deviation
Pre-test	15	11.06	8.64
Immediate post-test	15	8.80	5.21
Delayed post-test 1	15	11.26	5.96
Delayed post-test 2	15	10.00	4.72

It should be noted that the important comparison here is between the pre-test and the immediate post-test, as the students received corrective feedback only on their first, second (not represented in the table above), and third essays. According to Table 4, students do tend to avoid contexts that necessitate the use of articles, although the difference did not reach statistical significance ($t_{(14)} = 1.065$; $p = .305$).

6. Conclusion

The results of the present study are not consistent with the recent feedback studies (e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Ellis *et al.*, 2008; Sheen, 2007). Nevertheless, the inconsistency can be easily explained by recourse to the differences between the elicitation device as well as the target structure of the present study and those of the recent studies mentioned above. It should be noted that all these studies used highly selective grammar feedback and focused on the use of only two functions of the English articles (i.e., first- and second-mention functions to refer to unknown and known entities), employing pictures as the elicitation device. Moreover, except for Sheen (2007), none of these studies took account of the possibility of overuse, which may have affected the results.

Rahimi (2009) investigated the effect of corrective feedback on the accuracy of

EFL writers. He offered extensive corrective feedback on the expository and argumentative essays written by Iranian English majors. Rahimi concluded that the results of his study did not show a significant effect for feedback alone, although he maintained that corrective feedback had revealed its effect in interaction with practice. Therefore, Rahimi's results are fairly consistent with other feedback studies.

In fact, most of the criticisms that have been levelled against Truscott are rooted in a misunderstanding of his 1996 case against grammar correction. For instance, Ferris (1999) has argued that Truscott's sources are out-dated, that he has ignored evidence that run against his thesis, that feedback has been studied in various contexts and cannot be generalised to other contexts, and that he has rushed to an erroneous conclusion. Truscott (1999) published a response to Ferris's criticisms in which he dismissed most of her arguments.

In terms of interlanguage development, Truscott's (1996) thesis is theoretically more appealing than the rival claims. From an affective point of view, corrective feedback may have disastrous effects on students' attitudes (Hendrickson, 1980; Semke, 1984). Moreover, the assumption behind feedback studies is that there is a linear relationship between feedback and L2 accuracy. However, ESL researchers generally accept that L2 proficiency develops in a complex, U-shaped manner (Kellerman, 1983 cited in McLaughlin, 1990). Therefore, one might wonder why corrective feedback could ever be effective. Feedback studies are based on the positive role of consciousness in learning and on the noticing hypothesis (see Schmidt, 1990, 1994). According to this hypothesis, learners' conscious attention is a necessary and sufficient condition for acquisition to occur. Truscott (1998) has argued that the noticing hypothesis lacks a theory of language and that it does not have a firm basis in psychological research.

Sometimes applied linguists and SLA researchers appear to ignore the early findings regarding the course of second language development. Applied linguists are aware of the non-linearity of the acquisition of grammatical rules and regularities. However, they tend to oversimplify language acquisition processes (see Larsen-

Freeman & Cameron, 2008). Moreover, almost all feedback studies assume that feedback can be effective regardless of the learners' preferences and cultural differences (Goldstein, 2006; Hyland, 2003). Given all these shortcomings, one may be justified in concluding that the entire feedback studies are heading in the wrong direction.

The present study has theoretical as well as practical implications. Theoretically, this study contributes to the debate on the role of negative evidence in promoting second language acquisition. From a practical point of view, this study may help dispel the misconceptions regarding the positive effects of (extensive) corrective feedback, which is a time-consuming practice for teachers. Teachers spend a great deal of time giving formal feedback on the written products of their students – in fact, busy teachers may provide inconsistent and careless feedback due to lack of time (Zamel, 1985). However, there are good reasons to believe that students do not benefit from extensive (or selective?) feedback: Negative affective impacts, lack of attention on the part of students, and learners' in-built syllabi (Corder, 1967).

A number of factors, which have traditionally been included in other feedback studies, were excluded from the present study. This measure was taken to prevent the uncontrolled effect of these factors on the final results of the research. First, in order to guard against the consciousness raising nature of formal tests and the possibility that they may blur the differential treatment effects, no such instruments were included in the present study. Second, there is some evidence to suggest that different structures react differently to negative evidence (Ferris, 2006; Rahimi, 2009; Truscott, 2001). Following the recommendation by Norris and Ortega (2000), this study focused on the effects of selective grammar feedback on the development of only one grammatical category.

Like any other research, there were a number of factors that were not addressed by the present study. First and foremost, this study did not have a control group and the results should, therefore, be interpreted with caution. Second, this study did not include a proficiency test to determine the general proficiency of the students, which is a major shortcoming in most feedback studies carried out in the past. Third, it has

been suggested that various personality factors may differentially affect the success of grammar feedback (Guenette, 2007; Sheen, 2007). Nevertheless, due to practical limitations, it was not possible to measure the students' individual characteristics.

Notes

As Truscott and Hsu (2008) have demonstrated, successful incorporation of corrections into subsequent drafts is not a good measure of learning. Thus, revision studies (e.g., Maftoon & Rabiee, 2007; Sayyad Shirabad, 1999) have not been considered in the present article.

The results of a pilot study showed that intermediate (or more advanced) students demonstrate high levels of accuracy on the target structure (i.e., the English article system). Thus, less proficient students were selected for the purpose of this study in order to avoid a "ceiling effect".

References

- Bitchener, J. (2008). "Evidence in support of written corrective feedback". *Journal of Second Language Writing*, 17: 102-118.
- Bitchener, J., & U. Knoch (2008). "The value of written corrective feedback for migrant and international students". *Language Teaching Research*, 17: 102-118.
- Bitchener, J., S. Young, & D. Cameron (2005). "The effect of different types of corrective feedback on ESL student writing". *Journal of Second Language Writing*, 9: 227-258.
- Butler, Y. (2002). "Second language learners' theories on the use of English articles". *Studies in Second Language Acquisition*, 24: 451-480.
- Cardelle, M., & L. Corno (1981). "Effects on second language learning of variations in written feedback on homework assignments". *TESOL Quarterly*, 15: 251-261.
- 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.
- Cohen, J. (1992). "A power primer". *Psychological Bulletin*, 112: 155-159.
- Corder, S. P. (1967). "The Significance of Learner's Errors". *International Review of Applied Linguistics*, 4: 161-169.

- Ellis, R. (1999). "Item versus system learning: Explaining free variation". *Applied Linguistics*, 20: 460-480.
- Ellis, R. (2005). "Measuring implicit and explicit knowledge of a second language: A psychometric study". *Studies in Second Language Acquisition*, 27: 141-172.
- Ellis, R., & G. Barkhuizen (2005). *Analyzing Learner Language*. Oxford: Oxford University Press.
- Ellis, R., Y. Sheen, M. Murakami, & H. Takashima (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36: 353-371.
- Fathman, A., & E. Whalley (1990). "Teacher response to student writing: Focus on form versus content". In Kroll, B. (Ed.), *Second language writing: Research insights for the classroom*. (pp. 178-190). Cambridge: Cambridge University Press.
- Fazio, L. (2001). "The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students". *Journal of Second Language Writing*, 10: 235-249.
- 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.
- Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Erlbaum.
- 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: 49-62.
- Ferris, D. R. (2006). "Does error feedback help student writers/new evidence on the short- and long-term effects of written error correction". In Hyland, K., Hyland, F. (Eds.), *Feedback in Second Language Writing: Contexts and Issues* (pp. 81-104). Cambridge: Cambridge University Press.
- Ferris, D. R., & B. Roberts (2001). "Error feedback in L2 writing classes: How explicit does it need to be?" *Journal of Second Language Writing*, 10, 161-184.
- Frantzen, D. (1995). "The effects of grammar supplementation on written accuracy in an intermediate Spanish content course". *Modern Language Journal*, 79: 329-344.
- Goldstein, L. (2006). "Feedback and revision in second language writing: Contextual, teacher, and student variables". In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185-205). Cambridge: Cambridge University Press.
- Guenette, D. (2007). "Is feedback pedagogically correct? Research design issues in studies of feedback on writing". *Journal of Second Language Writing*, 16: 40-53.
- Hendrickson, J. M. (1980). "The treatment of error in written work". *Modern Language Journal*, 64: 216-221.

- Hyland, K., & F. Hyland (2006). "Feedback on second language students' writing". *Language Teaching*, 39: 83-101.
- James, M. A. (2008). "The influence of perceptions of task similarity/difference on learning transfer in second language writing". *Written Communication*, 25: 76-103.
- Kadia, K. (1988). "The effect of formal instruction on monitored and spontaneous naturalistic interlanguage performance: A case study". *TESOL Quarterly*, 22: 509-515.
- Kepner, C. G. (1991). "An experiment in the relationship of types of written feedback to the development of second language writing skills". *Modern Language Journal*, 75: 305-313.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
- Lalande, J. F. (1982). "Reducing composition errors: An experiment". *Modern Language Journal*, 66: 140-149.
- Larsen-Freeman, D., & L. Cameron (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.
- Lee, I. (2004). "Error correction in L2 secondary writing classrooms: The case of Hong Kong". *Journal of Second Language Writing*, 13: 285-312.
- Liu, D., & J. Gleason (2002). "Acquisition of the article the by nonnative speakers of English". *Studies in Second Language Acquisition*, 24: 1-26.
- McLaughlin, B. (1990). "Restructuring". *Applied Linguistics*, 11: 113-128.
- Maftoon, P. & M. Rabiee (2007). "Responding to students' writing: EFL students' reaction to teacher's written/oral responses in one-draft only composition classroom". *TELL*, 1 (4), 75-101.
- Master, P. (1997). "The English article system: Acquisition, function, and pedagogy". *System*, 25: 215-232.
- Norris, J. M., & L. Ortega (2000). "Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis". *Language Learning*, 50: 417-528.
- Polio, C., N. Fleck, & N. Leder (1998). "If only I had more time: ESL learners' changes in linguistic accuracy on essay revisions". *Journal of Second Language Writing*, 7: 43-68.
- Rahimi, M. (2009). "The role of teacher's corrective feedback in improving Iranian EFL learners' writing accuracy over time: is learner's mother tongue relevant? *Reading and Writing*, 22: 219-243.
- Robb, T., S. Ross, & I. Shortreed (1986). "Salience of feedback on error and its effect on EFL writing quality". *TESOL Quarterly*, 20: 83-93.

- Sayyad Shirabad, J. (1999). *The impact of teacher's written comments on Iranian EFL students' revision*. Unpublished master's thesis, Islamic Azad University, Central Branch, Tehran, Iran.
- Schmidt, R. (1990). "The role of consciousness in second language learning". *Applied Linguistics*, 11: 129-158.
- Schmidt, R. (1994). "Deconstructing consciousness in search of useful definitions for applied linguistics". *AILA Review* 11: 11-26.
- Semke, H. (1984). "The effects of the red pen". *Foreign Language Annals*, 17: 195-202.
- Sheen, Y. (2007). "The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles". *TESOL Quarterly*, 41: 255-283.
- Sheppard, K. (1992). "Two feedback types: do they make a difference?" *RELC Journal*, 23: 103-110.
- Truscott, J. (1996). "The case against grammar correction in L2 writing classes". *Language Learning*, 46: 327-369.
- Truscott, J. (1998). "Instance theory and Universal Grammar in second language research". *Second Language Research*, 14: 257-291.
- Truscott, J. (1999). "The case for "the case for grammar correction in L2 writing classes": a response to Ferris". *Journal of Second Language Writing*, 8: 111-122.
- Truscott, J. (2001). "Selecting errors for selective error correction". *Concentric: Studies in English Literature and Linguistics*, 27: 225-240.
- 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". *Journal of Second Language Writing*, 16: 255-272.
- Truscott, J., & A. Hsu (2008). "Error correction, revision, and learning". *Journal of Second Language Writing*, 17: 292-305.
- Zamel, V. (1985). "Responding to student writing". *TESOL Quarterly*, 19: 195-202.