Iranian Journal of Applied Linguistics (IJAL), Vol. 15, No. 1, March 2012, 43-71

Combating Fossilization through Feedback: Does it Work?

Fahimeh Marefat [a]
Associate professor, Allameh Tabataba'i University, Tehran, Iran

Musa Nushi [b]
PhD candidate in TEFL, Allameh Tabataba'i University, Tehran, Iran

Received 23 December 2011; revised 3 February 2012; accepted 13 February 2012

Abstract

This *pseudo-longitudinal* study adopted *typical-error* and *corrective feedback* approaches to investigating interlanguage fossilization. The errors in the argumentative essays of 76 Iranian EFL learners within and across three proficiency levels were identified and classified using the model proposed by Gass and Selinker (1994). The learners were first provided with implicit and then explicit feedback to see if the two feedback types would improve written production, and if there were errors that persisted. The results indicated that *word choice, plural, word form* and *article "the"* were the most frequent types of errors, with *word choice* topping the list. The results also showed that, although providing learners with feedback could lead to a reduction in errors, explicit feedback was a more effective strategy. However, the t-test results demonstrated that the short-term impact of feedback could not be sustained over time, a finding congruent with those of Truscott (2007) who questioned the efficacy of error correction. Moreover, the findings revealed that the error categories of *pronoun, word order, passive* and

[a] *E-mail address:* fahimehmarefat@yahoo.com
Corresponding address: College of Foreign Languages and Persian Literature, Allameh Tabataba'i University, Tehran, Iran

[b] *E-mail address:* mosesnushi@hotmail.com

possessive were likely candidates of fossilization because, although for all the other error types implicit feedback could be beneficial, these showed resistance to correction and needed explicit feedback to be eliminated.

Keywords: Interlanguage; Fossilization; Implicit versus explicit feedback; Iranian EFL learners**;** Argumentative writing

Introduction

Fossilization, first introduced by Selinker (1972), is deemed to be a distinctive characteristic of second language (L2) learning, one which has been referred to as one of the most enduring and fascinating problems in L2 acquisition. Most pervasive among adult L2 learners (Han, 2003, 2004; Han & Odlin, 2006; Kellerman, 1995; Lee, 2009; Schachter, 1996), the phenomenon of fossilization has been characterized as premature cessation of learning, despite the learner's strong motivation to learn, abundant amounts of exposure to L2 input, and plentiful opportunities for practice (Selinker, 1972).

Despite an extensive literature on fossilization, SLA researchers (e.g. Long, 2003; Han, 2009; Han & Odlin, 2006) have repeatedly raised two problems regarding the construct. First, the term "fossilization" lacks a unified definition. In the absence of a comprehensive conceptual definition, the term has been (mis)used by many simply as "a protean, catch-all" term to describe *any* lack of progress in L2 learning, regardless of its nature (Birdsong, 2004, p. 87). The second problem relates to the fact that fossilization has not been adequately described empirically.

This research study, in fact, investigated whether provision of corrective-feedback on errors that typically occurred in the written output of Iranian L2 learners across different proficiency levels could help them move past these *areas of stability*. It also intended to see whether there are any error type(s) that showed resistance to feedback, which could constitute likely candidates for fossilization.

Literature Review

Selinker (1972, 1992) maintains that one of the central attributes of L2 learners' interlanguage is its fossilizability, that is, it can cease developing in any of its developmental stages. He is, in fact, of the belief that "a mere 5%" (p. 212) of second language learners achieve native-speaker competence. White (2003) goes

even further saying that for L2 learners "native-like performance is the exception rather than the rule" (p. 2).

Han (2004) reviews empirical studies done on fossilization and says these studies have typically adopted one, or a combination of more than one, of the following methodological approaches: (1) longitudinal; (2) typical error; (3) advanced-learner; (4) corrective-feedback; and (5) length-of-residence. Only the three approaches of longitudinal, typical error and corrective-feedback are relevant to this research; therefore, the remaining part of this section briefly surveys some of the studies done in each of these approaches.

Han and Odlin (2006) state that the longitudinal approach is a posteriori, data-driven way of investigating fossilization which seeks to first establish fossilization and, if successful, proceeds to describe the phenomenon. Schumann's (1978) 10-month long case study of Alberto, a 33-year-old Costa Rican immigrant to the United States, exemplifies a study of fossilization within the longitudinal framework. Alberto showed little linguistic progress during the study. For example, he stayed in stage one of negation, using "the uniform negative *no* [No + verb] for most of his negative utterances" (Schumann, 1978, p. 367). He also rarely inverted in yes/no questions. Schumann explained Alberto's fossilization as a result of the failure to acculturate. Schumann's *Acculturation Model* (1986) predicts that learners like Alberto who do not acculturate will fail to acquire the L2; indeed, their interlanguage may even fossilize because lack of acculturation would lead to lack of exposure to the L2 input data, which, in turn, would result in fossilization.

Lardiere's (1998) longitudinal study of Patty, an adult native speaker of Chinese, had a far longer time span (more than eight years) which allowed Lardiere to examine her informant's performance cross-sectionally, i.e., at different points over time with significant intervals in between. Patty had lived in the United States for 18 years prior to the study. Out of these 18 years, she was totally immersed in the English speaking environment for 10 years. Data for Lardiere's study came from three audio-recorded conversations with Patty supplemented by two grammaticality judgment tasks, administered 18 months apart, Lardiere examined her informant's pronominal case marking and past tense inflectional morphology across the three recordings. A quantitative analysis of the informant's past tense marking in finite obligatory contexts showed that Patty had "remained unchanged over the eight years, despite massive exposure to target-language input by native

speakers in a virtually exclusively target-language environment" (1998, p. 17). In contrast, her mastery of pronominal marking was perfect, leading Lardiere to argue that "fossilization in one domain (inflectional morphology) does not preclude development in another (knowledge of syntactic features (use of pronouns) and word order)" (p. 41).

Jarvis and Pavlenko (2000) conducted a case study of a 33-year-old woman pseudo-named Aino, a native speaker of Finnish, who had resided in the United States for 10 years consecutively. The researchers, applying the two criteria of regularity and persistence of error appearance in the interlanguage of the learners, collected samples of Aino's oral and written production over five years at regular intervals. They reported that the five-year longitudinal database provided evidence of fossilization. They also argued the errors manifested influence from the learner's L1; and second, they alternated with corresponding target-like or correct forms, a phenomenon known as *backsliding* Selinker (1972).

In typical-error approach, errors that are pervasive in the interlanguage of learners with a homogeneous L1 – and usually across different proficiency levels – are used to demonstrate fossilization. Kellerman's (1989) study of the use of *would* in hypothetical conditionals by Dutch learners of English is an example of the application of this approach. Kellerman was driven by an interest to know how these learners' use of the linguistic feature gave them a "syntactic accent." He based his argument, among other things, on an earlier empirical study conducted by Wekker, Kellerman and Hermans (1982) of a typical error in Dutch-speaking learners of English that involves using *would* in the protasis of hypothetical conditionals such as "If I *would* be able to live all over again, I *would* be a gardener." Wekker *et al*.'s study examined the performance of advanced Dutch learners of English on non-past and past hypothetical conditional sentences in Dutch and English under experimental conditions using a pseudo-longitudinal design to come up with a diachronic view of the interlanguage structure under scrutiny. Results indicated that even advanced learner had at least some tendency to use and perceive such a construction as correct English structure, irrespective of their choice in Dutch which allows the use of both *were*/*was* and *would* in the protasis of hypothetical conditionals. The fact that even the most advanced learners persisted in the typical error was, for Kellerman, evidence of the tendency to fossilize. He cited psycho-typological evidence to explain the fossilized structure

and concluded that the fossilized structure was a function of the intersection of multiple tendencies with the native language.

L2 learners' reaction to corrective-feedback has also been employed as a means of determining whether the learner's erroneous usage of a particular linguistic form has fossilized or it is merely a temporary cessation of learning. For example, in Kellerman (1989) study mentioned above, it was not only a typical error in Dutch-English interlanguage community but also an error which seemed to have been immune to the pedagogic intervention that was made the linguistic focus for investigating fossilization. Thep-Ackrapong (1990, cited in Han, 2004), studied an ethnic Chinese Vietnamese refugee student, named Lin, at an American university for a year and a half. Thep-Ackrapong focused on the Lin's use of infinitival complements and the related structures, providing her with explicit rule explanation and corrective-feedback for one semester. Data collected at three times over a year and a half indicated that the tutoring had little effect on Lin's use of the linguistic structures.

Although there some studies on how to prevent features of L2 learners' interlanguage from becoming fossilized, the choice of the linguistic targets in many of them have been quite arbitrary, and we still do not have a clear idea on which aspects of interlanguage (lexical, morphosyntactic, sentential) tend to fossilize and the role implicit and explicit correction can play in *de- fossilizing* them. This study attempts to answer the following research questions:

1. What are the most frequent types of errors committed in the argumentative written production of Iranian EFL learners across three proficiency levels?
2. What is the impact of feedback (implicit vs. explicit) and proficiency level on reducing errors in the argumentative essays written by these learners?
3. Is there any error type(s) in the written output of these learners that shows resistance to feedback (either implicit or explicit) within and across the three proficiency levels?

The Study

This study was based on two main approaches to the study of fossilization, namely *typical-error analysis* (Kellerman, 1989; Schouten, 1996) and *corrective-feedback* (Kellerman, 1989; Thep-Ackrapong, 1990, cited in Han, 2004). In typical-error analysis approach, errors characteristic of learners with the same L1 background are identified and analyzed to spot instances of fossilization. As Han (2004) states,

this approach is underpinned by two assumptions. First, errors that are typical of a whole L2 population with homogeneous L1 background are possible candidates for fossilization; second, errors that are not only shared by that L2 population but also remain in the linguistic performance of its most advanced learners can be suspected of having a tendency to fossilize. Corrective-feedback examines L2 learners' reaction to feedback and assumes that errors that are immune to the pedagogic intervention should be made the linguistic focus for investigating fossilization. Furthermore, due to the time restraint, this study adopts a *pseudo-longitudinal* method of data collection in that learners of different proficiency levels are used as informants to provide a diachronic view of the interlanguage structures under scrutiny.

Participants

Seventy six male and female EFL learners, selected through purposive sampling out of a pool of 95, participated in this investigation. Out of the 76 learners, 24 provided the data for Stage One and 52 served as subjects of the researcher's six-month de-fossilization attempt in Stage Two and Three. The participants were all undergraduate and graduate students of different (non-English) majors in several universities across Tehran and ranged from 18 to 38 in age (mean age 27.5). They had all started learning their L2 after the puberty age and had been learning English for a minimum of five and a maximum of fifteen years up to March 2011 (the start date for the study). Persian was their first language and their primary contact with English was in their language classrooms. The participants were enrolled in EFL classes of five English language institutes. Their classes were held twice a week with each session lasting approximately 105 minutes.

Instruments

The materails and the tasks used in this study were as follows:

English Language Proficiency Test: A validated teacher-made English proficiency test was developed to determine the subjects' level of proficiency. It comprised of 100 multiple-choice items which measured the knowledge of English grammar and structures, vocabulary and reading.

Writing Tasks: As part of their coursework, learners were assigned two take-home topics, representing *argumentative* rhetorical mode, and were asked to write a 300-word essay on each topic.

Rating Scale: L2 learners**'** essays were evaluated using a six-point holistic scoring rubric patterned after the Test of Written English (TWE) scoring guide. The scores were the basis for classifying the learners into three levels of writing proficiency.

Data Collection Procedure

The study - implemented in three stages - adopted the error analysis model offered by Gass and Selinker (1994). Stage One consisted of error identification and classification in a sample of written productions of 24 Iranian EFL learners from the same pool as the participants in an attempt to get a general picture of the errors common in their writing; Stage Two involved two phases: identification and classification of errors made by 52 Iranian EFL learners in a written argumentative task and then pedagogical focus on these errors, and finally like Stage Two, Stage Three comprised two phases: first, identification and classification of errors made by the same 52 learners in another written argumentative task, then pedagogical focus on these errors.

Stage One

In the first stage, the researcher reviewed essays written by 34 EFL learners. The instructors for this group of learners were the researcher's colleagues who had at least five years of teaching experience. The researcher discussed his need for samples of argumentative writing with these colleagues and provided them with a topic and a sample essay in rhetorical mode. He also asked instructors to follow these steps:

- Administer the English proficiency test one session before having them to write on the topic,
- Give them the copy of the essay written in argumentative rhetorical mode
- Discuss with learners the principles of argumentative writing
- Assign them the topic as their homework and ask them to write a 300-word essay on that
- Ask them to pay particular attention to the coherence, cohesion, grammar, vocabulary and organization of their written work
- Remind them that the assigned written work would count towards their course grade

The overall quality of each essay was first assessed by two independent raters using the six-point holistic scoring rubric. To ensure higher consistency in the

scoring, the raters had a trial rating session during which the scoring guide was discussed and sample essays were rated. The inter-rater reliability index of the scores turned out to be .82. The assigned scores were the basis for classifying the essays into different levels of writing proficiency. The agreement between their writing proficiency and general English proficiency levels was .88. After classifying the writings into different levels, 10 were discarded because they either stopped short of the 300-word limit or did not observer the principles of argumentative writing, and 24 compositions, eight from each level, were reviewed to identify and classify the errors within them.

To classify the errors, the researcher reviewed the existing taxonomies of errors in the literature (Chandler, 2003; Darus, Tg Mohd Maasum, Stapa, Omar & Ab Aziz, 2007; Lee, 1990; Richards & Sampson, 1974) but could not come up with an ideal model which could accommodate all the errors since each taxonomy seemed to have been designed for a specific purpose and population and were either too broad or narrow in their classification of linguistic errors. So the researcher developed his own error taxonomy, which was in fact a collection of selected categories from several different error taxonomies. The model divided the errors into three classes of morphosyntactic, lexical and sentential, with morphosyntactic errors being further divided into *tense, preposition, article* (indefinite articles "a" or "an" and definite article "the"), *pronoun, word order, negative, passive*, *verb "to be", word form* (gerund/infinitive, verb/noun, adjective/ adverb), *conjunction, bound morphemes* (plural, third person singular, possessive, comparatives and superlative, subject/verb agreement); lexical errors being *word choice* (use of inappropriate and wrong words and phrases) and *sentential errors* involving faulty sentence structures.

Finally, as Huang (2002) pointed out, error identification requires sensibility and sensitivity to language on the part of the analyst. To establish the sensibility and sensitivity, the two raters independently reviewed the essays in order to identify and categorize the errors in them. For those cases that the raters felt unsure as to which category a particular error belonged, they debated the error among themselves until an agreement about its make-up was reached. Moreover, a native speaker of North America marked 20% of randomly selected papers in order to calculate inter-rater agreement. The percentage of agreement between the native speaker's categorization of the errors with the researcher and the other EFL instructor was .78 and .82 percent respectively.

The number of errors that occurred in each category was counted, and the descriptive statistics were computed. Table 1 shows the order of error types from most frequent to least frequent for each proficiency level.

Table 1
The order of error types across proficiency levels in Stage One

Level	**Error types**								
pre-inter-mediate	word form 6.75	word choice 6.00	article "the" 4.75	pre-position 4.25	plural 3.75	conju-nction 3.00	pronoun 3.00	sentence structure 2.75	
	tense 2.00	S-V agreement 1.63	article "an" 1.50	passive .88	word order .63	"to be" verb .50	negative .38	poss-essive .25	comp-arative .13
inter-mediate	word choice 5.63	plural 3.75	article "the" 3.25	word form 3.13	pre - position 3.00	sentence structure 2.28	S-V agreement 2.25	article "an" 2.25	
	conju-nction 1.88	pronoun 1.50	"to be" verb .88	tense .88	word order .75	poss-essive .38	passive .25	negative .13	comp-arative .000
advanced	word choice 3.25	article "the" 2.50	pre-position 1.25	plural 1.13	conju - nction 1.13	word form .88	S-V agreement .88	sentence structure .75	
	pro-noun .75	article "an" .75	poss-essive .38	"to be" verb .38	word order .38	negative .25	tense .25	passive .25	comp-arative .000

To get a better picture of the most frequent errors and for the sake of comparison of errors across levels, the researcher divided the errors for each proficiency level into three upper, middle and lower boundaries, with the upper and middle boundary containing six error types and the lower boundary having five categories. As can be seen in Table 1, the top six most frequent types of errors for pre-intermediate learners in descending order were: *word form, word choice, article "the", preposition, plural* and *conjunction*. Similarly, the top six types of errors for intermediate learners in order of frequency included: *word choice, plural, article "the", word form, preposition* and *sentence structure*. For advanced learners this order turned out to be: *word choice, article "the", preposition, plural,*

conjunction and *word form.* Comparing the categories of errors in the upper boundary reveals that the three proficiency levels share four out of six categories of errors: *word choice, article "the", plural* and *preposition*. Given the findings of this stage, a set of intervention activities and remedy materials were prepared to teach the troublesome points.

Stage Two

The data for this phase were collected from 61 university level EFL learners over two terms at 5 English language centers in Tehran. These learners were taught by the researcher himself. The first session of these classes started off by administering a 100-item test to categorize them into homogenous groups based on their English proficiency and correlate this with their writing proficiency level. Based on these scores, the students were divided into three proficiency levels. Eliminating the outliers, which numbered four, and excluding five more learners who had been learning English for less than two years, left the researcher with 52 subjects. To see whether there were any significant differences across the levels, a one-way ANOVA was conducted. The results revealed a significant difference across the three levels ($F_{(2, 49)} = 251.211$ $p < .05$). Post-hoc comparisons showed that each level was significantly different from the other two groups.

In the following session, the students were assigned one take-home topic in the argumentative rhetorical mode and were asked to write 300 words on that. Prior to the writing, however, the researcher provided the students with a sample essay written in argumentative mode. After collecting the first draft of writings, the researcher and an EFL teacher, independently assessed the quality of each composition using the six-point holistic scoring rubric and grouped them into level One, Two or Three writing proficiency. The inter-rater reliability turned out to be .78. Moreover, correlation between the English proficiency and writing proficiency was .90. The raters then reviewed the essays to identify and classify the errors in the essays.

Next, the researcher provided implicit feedback on the erroneous forms in the essays and tried to attract learner's attention to faulty language usages using minimal marking (Carduner, 2007) devices such as various colors and symbols to attract their attention to the faulty language usages. The students were previously informed of what these colors and symbols signified: red for grammatical errors, yellow for lexical errors, ▼ for omission of a linguistic element, × for when a linguistic element was extra and needed to be deleted and

_____(underlining) to indicate that a sentence was structurally flawed and needed to be revised. After marking the errors, the writings were handed back to the students who had to revise and return them to the teacher in the following session. The purpose of providing the learners with implicit feedback first and allowing them to self-correct was two-fold. First, engaging the learners in identifying the errors in their own production would count as a consciousness-raising activity (Ellis, 1993) and second, it would help the researcher determine whether the erroneous use was indeed an error or merely a mistake.

When the learners handed in the second draft of writings, the researcher read them to see if they corrected their erroneous uses, and for those cases that they failed to recognize the errors, he provided them with explicit feedback. These writings were again given back to the learners who had to make further changes based on the explicit corrective-feedback. Their corrections were further checked by the teacher and once he made sure they had gotten everything right, he asked them to rewrite their essay, incorporating the teacher's feedback and corrections. They gave the teacher the final version. It should be pointed out that the researcher tracked students' reaction to implicit and explicit feedback and recorded it.

Stage Three

The third stage of the study, which started three weeks after the completion of Stage Two and lasted for three months, involved the same 52 learners and replicated Stage Two procedures in almost every detail. The learners were first assigned a writing topic in the argumentative mode and then classified by the two raters into three levels of writing proficiency (inter-rater reliability = .79). The raters then reviewed the essays to identify and categorize the errors in them. The students went through receiving first implicit feedback and then explicit feedback in the subsequent sessions and handed the final revised version to the teacher. The purpose behind having another stage similar to Stage Two was twofold. The first aim was to examine the effectiveness of corrective-feedback in reducing errors. The second aim was to find out whether errors corrected in the previous stage would appear in this stage too, and if they did, whether learners could correct them by implicit feedback alone. The assumption was that if previously corrected errors lingered into this stage, they would qualify as likely candidates of fossilization.

Results

To answer the first question, i.e., the most frequent types of errors in the argumentative writings of Iranian EFL learners, the first drafts of essays written by 52 Iranian EFL learners in Stage Two were analyzed and errors within them identified and then categorized. Table 2 demonstrates the order of errors from the most common (left) to the least common (right) for each proficiency level. A comparison the top error categories for the pre-intermediate and intermediate learners shows that they share all but one *(preposition)* upper boundary error categories. It should also be pointed out that *word choice* and *plural* topped the list of errors for both of these levels. A comparison of error types for pre-intermediate and advanced learners also indicates that the two proficiency levels have the four errors of *word choice, plural, word form* and *article "the"* in common. Similarly, intermediate learners share five error types (*word choice, plural, preposition, word form,* and *article "the"*) in their upper boundary with advanced learners. It can be seen that four of the top six error categories (*word choice, plural, word form* and *article "the"*) are the same for all the three proficiency levels, with *word choice* topping the list.

Table 2
The order of error types across proficiency levels in Stage Two

Level	**Error types**								
pre-inter-mediate	word choice 7.263	plural 5.722	word form 4.666	article "the" 3.947	conju-nction 3.888	sentence structure 3.263	pre-position 3.222	pronoun 2.625	article "an" 2.071
	tense 2.071	poss-essive 2.00	S-V agreement 1.83	word order 1.714	comp-arative 1.50	negative 1.33	passive 1.18	"to be" verb 1.00	
inter-mediate	word choice 4.214	plural 3.384	pre-position 3.153	article "the" 3.00	word form 2.857	sentence structure 2.60	conju-nction 2.384	pronoun 1.888	tense 1.875
	article "an" 1.625	S-V agreement 1.50	poss-essive 1.50	word order 1.500	passive 1.25	comp-arative 1.00	"to be" verb 1.00	negative 1.00	
advanced	word choice 3.125	article "the" 2.941	word form 2.235	pre-position 1.944	plural 1.800	pronoun 1.70	article "an" 1.666	S-V agreement 1.615	poss-essive 1.50
	tense 1.50	sentence structure 1.35	word order 1.16	conju-nction 1.125	comp-arative 1.00	passive 1.00	negative 1.00	"to be" verb .000	

To ascertain that the findings are reliable, the researcher examined the pattern of errors on the first draft of the essays written by the same group of students on another argumentative topic in Stage Three. As Table 3 shows, the pattern of

errors, although in a different order, seems to be almost the same in Stage Three, i.e., the top six frequent types of errors for the pre-intermediate learners were: *word choice, word form, sentence structure, plural, preposition* and *article "the"*(sharing the five categories of *word choice, word form, sentence structure, plural,* and *article "the"* with the pre-intermediate learners in Stage Two); for the intermediate students the order was: *word choice, article "the", word form, preposition, plural* and *conjunction* (sharing the five categories of *word choice, article "the", word form, preposition* and *plural* with the intermediate learners in Stage Two); for the advanced level: *word choice, article "the", word form, preposition, conjunction* and *plural* (sharing the five categories of *word choice, article "the", word form, preposition,* and *plural* with the advanced learners in Stage Two). The order of top six categories of common errors for the three proficiency levels in Stage Three is thus almost the same as those obtained for the three proficiency levels in Stage Two, corroborating the earlier insight into the error distribution for the population under study, i.e., the six categories of *word choice, plural*, a*rticle "the", word form*, albeit in different orders, were common across the proficiency levels.

Table 3

The order of error types across proficiency levels in Stage Three

Level	**Error types**								
pre-inter-mediate	word choice 6.75	word form 6.12	sentence structure 5.37	plural 4.43	pre-position 4.40	article "the" 4.40	conju-nction 3.80	pronoun 3.23	S-V agreement 2.92
	tense 2.08	article "an" 1.90	word order 1.42	passive 1.37	poss-essive 1.14	comp-arative 1.00	"to be" verb 1.00	negative 1.00	
inter-mediate	word choice 4.87	article "the" 4.23	word form 3.266	pre-position 3.25	plural 2.66	conju-nction 2.66	sentence structure 2.53	article "an" 2.0	pronoun .80
	tense 1.75	S-V agreement 1.70	poss-essive 1.6	comp-arative 1.50	negative 1.20	"to be" verb 1.00	passive 1.00	word order 1.0 0	
advanced	word choice 3.05	article "the" 2.45	word form 2.29	pre-position 2.12	conju-nction 1.87	plural 1.78	word order 1.66	article "an" 1.42	sentence structure 1.36
	"to be" verb 1.33	tense 1.33	S-V agreement 1.28	passive 1.16	poss-essive 1.14	comp-arative 1.00	negative 1.00	pronoun 1.00	

There are two differences between the two sets of top error categories for the three proficiency levels in Stage Two and Three tables, however. First, the order of the top six categories, except for *word choice* and *preposition*, are quite different;

the order of top six categories for Stage Two were: *word choice, plural, article "the", word form, preposition,* and *conjunction* but for Stage Three this order was *word choice, word form, article "the", conjunction, preposition* and *plural*. It is also interesting to note that these categories were also the six top error types obtained for learners in Stage One. Second, a quick survey of the means of the various errors in Stage Three gives the initial impression that there was not a significant reduction in the errors, even after corrective-feedback was provided in Stage Two. In some cases, there was even a slight increase.

To answer the second research question, i.e., whether feedback type had differential impact on error reduction, a repeated measures ANOVA with one within group factor, corrective-feedback type with three levels (no feedback, implicit feedback and explicit feedback) and one between group factor, writing proficiency with three levels was conducted. The results are presented in Table 4. Since there were a few learners making *negative, "to be" verb* and *comparative* errors, no analyses could be run for these categories.

Table 4
Descriptive statistics plus repeated measures ANOVA and post-hoc scheffe test results for three proficiency levels in Stage Two

Error type	No. of learners making the error per level	Mean prior to any feedback	Mean after implicit feedback	Mean after explicit feedback	RM ANOVA		Post-hoc Scheffe test	
					Main effects & interaction	P.	Feedback results	Level results
tense	Level 1 = 14 Level 2 = 5 Level 3 = 4	2.07 2.00 1.50	1.78 1.80 1.25	.71 .40 .25	Feedback Level Feed*level	.000 .59 .96	1 = 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
preposition	Level 1 = 17 Level 2 = 12 Level 3 = 14	3.23 3.25 2.21	2.29 2.16 1.50	.64 .33 .07	Feedback Level Feed*level	.000 .062 .78	1 > 2 1 > 3 2 > 3	1 = 2 1 > 3 2 ≥ 3
article "the"	Level 1 = 18 Level 2 = 11 Level 3 = 15	4.11 3.27 3.06	2.38 1.90 1.60	.44 .00 .06	Feedback Level Feed*level	.000 .059 .81	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 = 3

article "an"	Level 1 = 11 Level 2 = 8 Level 3 = 6	2.09 1.62 1.83	1.72 1.37 1.16	.36 .37 .000	Feedback Level Feed*level	.000 .59 .74	1 > 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
pronoun	Level 1 = 15 Level 2 = 8 Level 3 = 9	2.73 2.00 1.77	2.33 1.87 1.55	.53 .12 .33	Feedback Level Feed*level	.000 .18 .49	1 ≥ 2 1 > 3 2 > 3	1 = 2 1 ≥ 3 2 = 3
word order	Level 1 = 11 Level 2 = 7 Level 3 = 5	1.81 1.57 1.20	1.63 1.57 1.2	.18 .000 .200	Feedback Level Feed*level	000 .18 .49	1 = 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
passive	Level 1 = 10 Level 2 = 4 Level 3 = 4	1.20 1.20 1.00	1.1 1.00 0.75	.3 .000 .000	Feedback Level Feed*level	.000 .649 .71	1 = 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
word form	Level 1 = 17 Level 2 = 14 Level 3 = 15	4.47 2.85 2.40	3.05 2.14 1.93	.88 .07 .000	Feedback Level Feed*level	.000 .015 .37	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 = 3
conjunction	Level 1 = 18 Level 2 = 12 Level 3 = 14	3.88 2.50 1.14	3.27 2.16 1.07	1.66 0.58 0.07	Feedback Level Feed*level	.000 .000 .003	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 > 3
plural	Level 1 = 18 Level 2 = 11 Level 3 = 10	5.72 3.45 1.90	3.05 2.36 1.30	.50 .09 .000	Feedback Level Feed*level	.000 .000 .000	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 > 3
possessive	Level 1 = 5 Level 2 = 5 Level 3 = 3	2.00 1.60 1.66	1.20 1.00 1.33	.000 .000 .000	Feedback Level Feed*level	.000 .69 .90	1 = 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
sub-verb agreement	Level 1 = 12	2.00 1.60	1.08 1.00	.000 .000	Feedback Level	.000 .50	1 > 2 1 > 3	1 = 2 1 = 3

	Level 2 = 5 Level 3 = 6	2.33	1.16	.000	Feed*level	.72	2 > 3	2 = 3
word choice	Level 1 = 19 Level 2 = 14 Level 3 = 14	7.26 4.21 3.21	6.00 3.07 2.28	2.36 .50 .14	Feedback Level Feed*level	.000 .000 .015	1 > 2 1 > 3 2 > 3	1 = 2 1 = 3 2 > 3
sentence structure	Level 1 = 19 Level 2 = 10 Level 3 = 13	3.26 2.60 1.38	3.05 2.20 1.23	2.05 1.30 .07	Feedback Level Feed*level	.000 .002 .79	1 > 2 1 > 3 2 > 3	1 = 2 1 > 3 2 > 3

As Table 4 illustrates, there was a reduction in the means of all error categories when learners were provided with either explicit or explicit feedback as opposed to when there was no feedback. The main effect of feedback was significant for all error types. Post-hoc comparisons suggested that implicit feedback was successful in reducing all error types except for *tense, pronoun, word order, passive,* and *possessive* errors, indicating that the learners did not have the linguistic proficiency to remedy these errors even after being pointed out to them. It was also revealed that explicit feedback brought about a significant decrease in all error types in comparison with both implicit and no-feedback.

ANOVA results also indicated that proficiency level played a significant role for only five error categories (*word form, conjunction, plural, word choice,* and *sentence structure*), suggesting that for the rest of the errors, learners benefited from corrective-feedback to the same degree. Post-hoc comparisons further showed that advanced learners benefitted more from the other two levels for *preposition, conjunction, plural* and *sentence structure*; for *article "the"* and *word form*, intermediate and advanced learners performed better than pre-intermediate level learners but did not differ from each other. Intermediate learners benefitted more than pre-intermediate and advanced learners for *conjunction* and *plural* errors. Additionally, the interaction between feedback and proficiency levels turned out to be insignificant, except for the categories of *conjunction, plural,* and *word choice* ($p < .05$), which means that except for these three error types, the three proficiency levels benefited equally from the different feedback types.

Repeated measures ANOVA was also employed to analyze the data for Stage Three. Table 5 presents the results for this stage (again, since there was few number of learners making *negative, "to be" verb* and *comparative* errors, no analyses could be conducted for these categories). As this table illustrates, the variable feedback had a significant effect for all error types. Post-hoc comparisons showed that for all error types where feedback was effective, both implicit and explicit feedback were effective in reducing the errors except for errors of *pronoun, word order, passive* and *possessive* where implicit feedback had played no role.

As for level, this variable was effective only for *pronoun, word form, conjunction, plural, possessive, word choice* and *sentence structure* categories. Post-hoc comparisons showed that for *conjunction* and *word choice,* the three groups were benefiting from the feedbacks differently and in each case the higher groups made the best of the feedbacks. For three other types of errors, i.e., *word form, plural,* and *sentence structure*, the feedback was effective for intermediate and advanced levels to the same extent. For the error category of *pronoun*, only the third group could benefit more than the first group and for the other levels the effect was the same. As for *possessives*, intermediate learners made the least benefit from feedback. Additionally, the interaction between feedback and proficiency levels turned out to be insignificant, except for the categories of *pronoun, word choice, word form* and *plural*, meaning that feedback did not have the same impact for all levels.

Table 5

Descriptive statistics plus repeated measures ANOVA and post-hoc scheffe test results for three proficiency levels in Stage Three

Error type	No. of learners making the error per level	Mean prior to any feedback	Mean after implicit feedback	Mean after explicit feedback	Repeated measures ANOVA		Post-hoc Scheffe test	
					Main effects & interaction	P.	Feedback results	Level results
tense	Level 1 = 12 Level 2 = 8 Level 3 = 3	2.08 1.75 1.33	1.75 1.62 1.00	.50 .37 .000	Feedback Level Feed*level	.00 .33 .97	1 > 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
preposition	Level 1 = 15 Level 2 = 15	4.40 3.40 2.88	2.66 2.33 1.88	.46 .2 .1	Feedback Level Feed*level	.00 .12 .37	1 > 2 1 > 3 2 > 3	1 = 2 1 ≥ 3 2 = 3

	Level 3 = 9							
article "the"	Level 1 = 15 Level 2 = 13 Level 3 = 12	4.00 4.23 2.83	2.60 2.15 1.33	.40 .000 .000	Feedback Level Feed*level	.00 .1 .34	1 > 2 1 > 3 2 > 3	1 = 2 1 > 3 2 = 3
article "an"	Level 1 = 10 Level 2 = 9 Level 3 = 4	2.00 2.11 1.75	1.5 1.66 1.75	.20 .22 .25	Feedback Level Feed*level	.00 .94 .76	1 > 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
pronoun	Level 1 = 13 Level 2 = 9 Level 3 = 7	3.23 1.88 1.00	2.69 1.88 1.00	.69 .44 .0000	Feedback Level Feed*level	.00 .01 .009	1 = 2 1 > 3 2 > 3	1 = 2 1 > 3 2 = 3
word order	Level 1 = 7 Level 2 = 5 Level 3 = 6	1.42 1.00 1.66	1.28 1.00 1.66	.28 .000 .66	Feedback Level Feed*level	.00 .21 .20	1 = 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
passive	Level 1 = 7 Level 2 = 3 Level 3 = 5	1.42 1.00 1.20	1.42 1.00 1.00	.42 .000 .000	Feedback Level Feed*level	.00 .25 .95	1 = 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
word form	Level 1 = 16 Level 2 = 15 Level 3 = 14	6.12 3.26 2.57	5.06 2.40 1.35	1.25 .26 .07	Feedback Level Feed*level	.00 .00 .00	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 = 3
conjunction	Level 1 = 15 Level 2 = 12 Level 3 = 16	3.80 2.66 1.87	3.53 2.50 1.56	1.66 .83 .12	Feedback Level Feed*level	.00 .00 .60	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 > 3
plural	Level 1 = 16 Level 2 = 7 Level 3 = 8	4.43 2.71 2.12	2.68 1.71 1.12	.18 .000 .000	Feedback Level Feed*level	.00 .009 .013	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 = 3

possessive	Level 1 = 7 Level 2 = 3 Level 3 = 3	1.14 1.66 1.00	1.14 1.66 1.00	.000 .33 .000	Feedback Level Feed*level	.00 .017 .91	1 = 2 1 > 3 2 > 3	1 > 2 1 = 3 2 > 3
S-V agreement	Level 1 = 11 Level 2 = 6 Level 3 = 3	3.27 2.16 1.33	1.9 1.5 1.00	.000 .000 .000	Feedback Level Feed*level	.00 .80 .28	1 > 2 1 > 3 2 > 3	1 = 2 1 = 3 2 = 3
word choice	Level 1 = 16 Level 2 = 16 Level 3 = 16	6.75 4.87 3.18	5.93 4.31 2.12	2.43 1.06 .18	Feedback Level Feed*level	.00 .00 .023	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 > 3
sentence structure	Level 1 = 16 Level 2 = 15 Level 3 = 10	5.37 2.53 1.4	4.87 2.46 1.20	3.56 1.00 .10	Feedback Level Feed*level	.00 .001 .48	1 > 2 1 > 3 2 > 3	1 > 2 1 > 3 2 = 3

Although the effect of feedback turned out to be significant for both stages in this study, it cannot be automatically concluded that error correction had a positive and durable effect on the students' written performance. In a bid to investigate the effectiveness of corrective-feedback (De Graaff & Housen, 2009), a paired samples t-test was used to compare the means of error types in the first drafts of Stage Two (before learners received any feedback on their written output), and Three (after learners received feedback on their written output) to see if feedback could assist the learners to reduce their errors in a new task. T-test results, as shown in Table 6, indicate that there was not a significant reduction in means of errors for any of the error categories, hinting that the feedback did not have any long term effect. It should be pointed out that there was a marginally significant effect of feedback on the *subject-verb agreement* error category: $t_{(14)} = 2.125$ $p \leq .05$, and the learners made fewer errors in this category in the third stage of the study. There was also a significant effect of feedback on *preposition* errors: $t_{(16)} = -3.395$ $p < .05$; however, comparing the means of this category in Stage Two and Three demonstrated that learners in fact made more preposition errors in the third stage of the study. As mentioned earlier, there seems to be an increase in the means of all error categories.

Table 6
T-test results comparing first drafts of stage two and three

Error categories		t-test results		
		t	df	sig
Pair 1	Tense	-.197	8	.849
Pair 2	Preposition	-3.395	16	.004
Pair 3	Article "the"	.243	17	.811
Pair 4	Article "an"	.936	8	.377
Pair 5	Pronoun	-.714	12	.489
Pair 6	Word order	1.581	5	.175
Pair 7	Passive	-.397	3	.718
Pair 8	Word form	-.867	17	.398
Pair 9	Conjunction	.677	16	.508
Pair 10	Plural	1.783	16	.094
Pair 11	Possessive	-1.512	2	.270
Pair 12	S-V Agreement	2.125	14	.052
Pair 13	Word choice	.929	18	.365
Pair 14	Sentence structure	1.909	18	.072

To answer the third question, i.e., if there is any error type(s) that shows resistance to feedback across the proficiency levels and hence a tendency towards fossilization, the researcher sought to find error categories that received both implicit and explicit feedback in Stage Two which, nonetheless, kept reappearing in Stage Three and defied being eliminated by implicit feedback only and still needed explicit feedback. The assumption was that if there are some error categories which, despite being treated by corrective-feedback in the previous stage, continue to show themselves in Stage Three and resist disappearing by less intrusive feedback type (implicit feedback), then we have cases where fossilization might be setting in. the results showed that the four error categories of *pronoun, word order, passive* and *possessive* are likely candidates of fossilization on the ground that, although for all the other error types implicit feedback could be beneficial, for these four categories, only explicit feedback could lead to a reduction of errors, suggesting the toughness or resistance of these error categories. Furthermore, except for pronoun errors in Stage Two and possessive errors in Stage Three, proficiency level did not play a determining role in how the learners benefited from corrective-feedback, indicating that learners across the three language proficiency levels needed explicit feedback to eliminate these error types from their essays.

Discussion

The first aim of this study was to uncover the most frequent types of errors in the argumentative written production of Iranian EFL learners across three proficiency levels. The results of analysis of the errors in the first drafts of essays written in Stage Two and Three showed that the six categories of *word choice, plural*, a*rticle "the", word form, preposition*, and *conjunction*, albeit in different orders, were common across the proficiency levels. This was also the case for the 24 learners involved in Stage One, which further strengthened the researcher's belief in the typicality of these errors. It was further discovered that, as noted by Martin (1984), there was not always a linear relationship between proficiency and number of errors. Sometimes the advanced learners committed more errors in their productive use.

Another noteworthy finding of this study was that the error category that topped the list of errors of all proficiency levels in each stage was *word choice*, indicating the prevalence of that lexical error among even advanced learners. In fact, empirical evidence suggests that lexical errors are the most frequently occurring category of errors in written English (Grauberg, 1971; Lennon, 1991; Meara, 1984). Moreover, Ferris (1999) makes a distinction between ''treatable'' and ''untreatable'' errors, suggesting that the former are rule-based, and so learners can correct them using resources such as a grammar book, while the latter are idiosyncratic and require learners to utilize acquired knowledge of the language to resolve them. In two empirical studies, Ferris, Chaney, Komura, Roberts and McKee (2000) and Ferris and Roberts (2001) examined this distinction and found that learners made substantial progress over a semester in reducing errors in verb tense and form (treatable) but made only slight progress in reducing lexical (untreatable) errors.

The second research question addressed the issue of the impact of feedback type on error reduction. The results indicated that providing learners with corrective-feedback led to a significant drop in all errors as opposed to when there was no feedback. That the corrective-feedback can assist learners improve their performance has been supported by many empirical studies (e.g., Adams, 2003; Bitchener & Knoch, 2010; Ferris, 2004, 2010). However, a deeper analysis of the data revealed that implicit and explicit feedback had differential effect on errors, in that implicit feedback was successful in bringing about a significant decrease in all error categories but *tense, pronouns, word order, passive,* and *possessive*, a finding

that, except for the error type of *tense*, was repeated in Stage Three. This may point to the fact that while implicit feedback can be effective for L2 learning, more explicit types of feedback tend to have an even greater effect (Rosa & Leow, 2004; Sauro, 2009; among others). Dabaghi (2008) suggests a number of reasons for the better impact of explicit feedback: (1) explicit correction created more attention, (2) the fact that learners were explicitly corrected on their errors created a contrast with the form in their interlanguage, (3) the provision of the correct form in implicit correction may not have been effective because it was less clear to learners what was wrong with their erroneous utterances and without such understanding, hypothesis revision was not possible, and (4) learners most likely perceived the explicit corrections as corrective-feedback requiring them to correct their errors whereas this was not the case with the implicit feedback.

The question is, then, raised as to whether the short-term impact of corrective-feedback could be sustained over time. Statistical analysis of the errors students committed in their first draft of the two essays in Stages Two and Three showed that there was an increase in the means of almost all the error categories. In general, these findings are congruent with those of Truscott (2007) who suggested that correction had little or no effect on student writing and should be abandoned. From an analysis of studies investigating the effect of corrective-feedback, including those by Kepner (1991), Semke (1984) and Sheppard (1992), he concluded that there was no convincing evidence that error correction ever helped student writers improve the accuracy of their writing. Truscott goes even further and makes the controversial claim that error correction is not only ineffective but even harmful as it creates more problems. Much to the chagrin of the researcher, this study appears to support this contentious claim because, despite positive findings on the role of feedback in Stage Two and Three separately, error correction created more problems for learners than it yielded benefits. Truscott cites two reasons for this finding: first, error correction treats different linguistic categories (lexical, syntactic, and morphological) as being equivalent, when in fact they are acquired through different stages and processes, and second, negative evidence only alters language performance but does not change learners' underlying grammar, which develops only through exposure to the language in natural interaction.

Finally, the third and main objective of this research was seeking instances of erroneous linguistics features that were impermeable to any external influences.

Results showed that the four error categories of *pronoun, word order, passive* and *possessive* are likely candidates of fossilization because results for Stage Two showed that, although for all the other errors types implicit feedback could be beneficial, for these four categories, only explicit feedback could lead to a reduction of errors. This pattern turned out to be the same in Stage Three, suggesting the toughness of these error categories. Furthermore, except for pronoun errors in Stage Two and possessive errors in Stage Three, proficiency level did not play a role in how the learners benefited from corrective-feedback, indicating that learners across different proficiency levels needed explicit feedback to improve the accuracy of their essays.

This finding may not be surprising given SLA findings on learning these features. Ellis (1994) states that the general view is that learners of different languages, English, French, German, Dutch and Spanish, "experienced similar problems with pronouns" (p. 96). As far as English pronouns are concerned, Scheffler (2009) believes it is L2 learners' perception of a linguistic feature that determines its difficulty and learners benefit most from L2 instruction when it addresses errors they perceive as difficult. It can be argued that pronoun errors persisted in the written performance of these learners because from their perspective, pronouns did not constitute a difficult linguistic domain.

Word order rules are often troublesome for students learning English (Borer, 1983; White, 1988, 1989). Jabbari and Niroomizadeh (2008) investigated parametric difference concerning modifier placement between Persian and English and whether providing feedback on modifier placement would be effective in helping learners of English to master the contrast. The results showed that there was a significant difference between the performance of the experimental group and the control group in the immediate posttest but this effect was not long-lived, as demonstrated by the delayed post-test 12 weeks later, proving the resistance of word order errors to instructional feedback and Truscott (1996)'s contention that negative evidence cannot alter learners' linguistic competence.

The English passive has also been notoriously difficult for EFL/ESL learners. Birjandi, Maftoon and Rahemi (2011) examined the effects of two types of instruction - processing instruction, an input-based approach to grammar instruction introduced by VanPatten (1996), and an output-oriented type of instruction - on the acquisition of the English passive by Persian-speaking learners

of English. Using a pretest-treatment-posttest (immediate and delayed) quasi-experimental design, they reported a short-term benefit of instruction on improving production and comprehension of passives, but the effect disappeared after a one-month interval, which in a way indicated the resistance of these syntactic errors to corrective-feedback.

The final error category on the list of potentially fossilizable errors is possessive construction. Despite its apparent simplicity, possessive morpheme appears last in the order of acquisition for major grammatical morphemes in L2 (Dulay & Burt, 1973) and L2 morphological development (Larsen-Freeman, 1976) for both natural and classroom setting. Celece-Murica and Larsen-Freeman (1983) claims the difficulty arises because they are sometimes interchangeable, for example, "the man's name" can be changed to "the name of the man."

Conclusion

Conclusions drawn from this study indicate that the effect of corrective-feedback did not seem to be durable, thus corroborating Truscott's (1996, 1999, 2007) contention that correcting learners' errors in a written composition may enable them to eliminate the errors in a subsequent draft but has no effect on grammatical accuracy in a new piece of writing, i.e., it does not result in acquisition and learners need sustained natural exposure to and interaction in the target language to be able to develop their underlying linguistic competence.

It was also observed that the most frequent errors were not the ones most at risk of fossilization and errors in the use of *pronouns, word order, passives* and *possessives*, which occurred with lower frequency in the learners' performance, turned out to be more resistant to corrective-feedback, revealing a shortcoming of typical error approach to investigating fossilization. Due to the persistence of these four error types, EFL practitioners teaching learners with Persian as their L1 need to give serious pedagogical considerations to the teaching of these features to help avoid fossilization.

Notes on Contributors:

Fahimeh Marefat *is an associate professor at Allameh Tabataba'i University. Her area of interest is writing, CALL, and creating effective teaching / learning environment.*

Musa Nushi *is a Ph.D. candidate in TEFL at the College of Foreign Languages and Persian Literature in Allameh Tabataba'i University. He holds an M.A. from the College of Foreign Languages of Tehran University. He is currently an EFL instructor and interested in different areas of second language learning, psycholinguistics, and language teaching methodology.*

References

Adams, R. (2003). L2 output, reformulation and noticing: implications for IL development. *Language Teaching Research, 7*, 347-376.

Birdsong, D. (2004). Second language acquisition and ultimate attainment. In A. Davies & C. Elder (Eds.), *Handbook of applied linguistics* (pp. 82-105). London: Blackwell.

Birjandi, P., Maftoon, P., & Rahemi, J. (2011). VanPatten's processing instruction: Links to the acquisition of the English passive structure by Iranian EFL learners. *European Journal of Scientific Research*, *64*(4), 598-609.

Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, *31*, 193-214.

Borer, H. (1983). *Parametric syntax: Case studies in Semitic and Romance languages*. Foris Publications, Dordrecht.

Carduner, J. (2007). Teaching proofreading skills as a means of reducing composition errors. *Language Learning Journal, 35*(2), 283-295.

Celece-Murica, M., & Larsen-Freeman, D. (1983). *The grammar book*: *An ESL/EFL teacher's course*. Boston: Heinle & Heinle.

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.

Dabaghi, A. (2008). A comparison of the effects of implicit and explicit corrective feedback on learners` performance in tailor-made tests. *Journal of Applied Sciences, 8,* 1-13.

Darus, S., Tg Mohd Maasum, T. N. R., Stapa, S. H., Omar, N., & Ab Aziz, M. J. (2007). Developing an error analysis marking tool for ESL learners. *Proceedings of the 7th WSEAS International Conference on Applied Computer Science (ACS'07),* Venice, Italy, 21-23.

De Graaff, R., & Housen, A. (2009). Investigating the effects and effectiveness of L2 instruction. In M. Long, & C. Doughty (Eds.), *The handbook of language teaching* (pp. 726-755). Oxford: Blackwell.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*(2), 245-258.

Ellis, R. (1993). *Second language acquisition and the structural syllabus. TESOL Quarterly*, *27*, 91-113.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

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. (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. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical application. *Studies in Second Language Acquisition, 32*, 181-201.

Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*, 161-184.

Ferris, D. R., Chaney, S. J., Komura, K., Roberts, B. J., & McKee, S. (2000). Perspectives, problems, and practices in treating written error. *Colloquium presented at International TESOL Convention, March 14-18, 2000, Vancouver, British Columbia, Canada.*

Gass, S. M., & Selinker, L. (2004). *Second language acquisition: An introductory course*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Grauberg, W. (1971). An error analysis in German of first year university students. In G. Perren & J. Trim (Eds.), *Applications of linguistics*. Cambridge: Cambridge University Press.

Han, Z.-H. (2003). Fossilization: From simplicity to complexity. *International Journal of Bilingual Education and Bilingualism, 6*, 95-128.

Han, Z.-H. (2004). *Fossilization in adult second language acquisition.* Clevedon, UK: Multilingual Matters.

Han, Z.-H. (2009). Interlanguage and fossilization: Towards an analytic model. In V. Cook & L. Wei (Eds.), *Contemporary applied linguistics* (Vol. I: Language Teaching and Learning. London: Continuum, 137-162.

Han, Z.-H., & Odlin T. (Eds.) (2006). *Studies of fossilization in second language acquisition.* Clevedon: Multilingual Matters.

Huang, J. (2002). Error analysis in English teaching: A review of studies. *Journal of Chung-San Girls' Senior High School*, *2,* 19-34.

Jabbari, A. A., & Niroomizadeh, A. K. (2008). The effect of negative and positive evidence on learning English as a foreign language by Persian speakers. *Pazhuhesh-e Zaban-haye Khareji (Research on Foreign Languages)*, *51,* 45-61.

Kellerman, E. (1989). The imperfect conditional: Fossilization, cross-linguistic influence and natural tendencies in a foreign language setting. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism across life Span* (pp. 87-115). Cambridge, UK: Cambridge University Press.

Kellerman, E. (1995). Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics, 15*, 125-150.

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.

Jarvis, S., & Pavlenko, A. (2000). Conceptual restructuring in language learning: Is there an end state? *Paper presented at the SLRF 2000, Madison, WI.*

Lardiere, D. (1998). Case and tense in the "fossilized" steady state. *Second Language Research, 14*, 1-26.

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning, 25,* 125-135.

Lee, E. (2009). Issues in fossilization and stabilization. *Linguistic Research, 26*(2), 151-166.

Lee, N. (1990). Notions of error and appropriate corrective treatment. *Hong Kong Papers in Linguistics and Language Teaching, 13*, 55-69.

Lennon, P. (1991). Error and the very advanced learner. *International Review of Applied Linguistics*, *29*, 31-44.

Long, M. (2003). Stabilization and fossilization in interlanguage development. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 487-536). Oxford: Blackwell.

Martin, M. (1984). Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal*, *68*, 130-37.

Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper & A. Howatt (Eds.), *Interlanguage*. Edinburgh: Edinburgh University Press.

Richards, J. C., & Sampson, G. P. (1974). The study of Learner English. In J. C. Richards (Ed.), *Error Analysis:Perspectives on Second Language Acquisition.* London: Longman.

Rosa, E., & Leow, R. (2004). Awareness, different learning conditions, and L2 development. *Applied Psycholinguistics, 25*, 269-292.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning and Technology*, *13*(1), 96-120.

Schachter, J. (1996). Maturation and the issue of universal grammar in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 159-94). San Diego: Academic Press.

Scheffler, P. (2009). Rule difficulty and the usefulness of instruction. *ELT Journal, 63*, 5-12.

Schouten, E. (1996). Crosslinguistic influence and the expression of hypothetical meaning. In E. Kellerman, B. Weltens & T. Bongaerts (Eds.), *EUROSLA 6: A Selection of Papers*. *Toegepaste Taalwetenschap in Artikelen* (*Applied Linguistics in Article form*) (pp. 161-74). Amsterdam: VU Uitgeverij, 55.

Schumann, J. H. (1978). Social and psychological factors in second language acquisition. In J. C. Richards (Ed.), *Understanding second & foreign language learning* (pp. 163-178). Rowley, MA: Newbury House Publishers.

Schumann, J. H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual and Multicultural Development, 7*(5), 379-92.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*(3), 209-31.

Selinker, L. (1992). *Rediscovering interlanguage*. New York: Longman.

Semke, H. (1984). The effects of the red pen. *Foreign Language Annals, 17*, 195-202.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal*, *23*, 103-110.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, *46*, 327-369.

Truscott, J. (1999). The case for 'the case for grammar correction in L2 writing classes': A response to Ferris. *Journal of Second Language Writing*, *8*, 111-122.

Truscott, J. (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.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.

Wekker, H., Kellerman, E., & Hermans, R. (1982). Trying to see the 'would' for the trees. *Interlanguage Studies Bulletin, 6*, 22-55.

White, L. (1988). Island effects in second language acquisition. In S. Flynn & W. O'Neil (Eds.), *Linguistic theory in second language acquisition*. Kluwer Academic Publishers.

White, L. (1989). *Universal grammar and second language acquisition*. Amsterdam: John Benjamins.

White, L. (2003). *Second language Acquisition and universal grammar*. Cambridge: Cambridge University Press.