



The Effect of First Language Metalinguistic Comments within Iranian EFL Learners' Task-Based Instruction: Durable Gains of Third Person /-s/

Soheil Rahimi

Department of English Language and Literature, Faculty of Humanities, Arak University, Arak, Iran
E-mail: s-rahimi@phd.araku.ac.ir

*Moussa Ahmadian **

Department of English Language and Literature, Faculty of Humanities, Arak University, Arak, Iran
E-mail: M-ahmadian@araku.ac.ir

Majid Amerian

Department of English Language and Literature, Faculty of Humanities, Arak University, Arak, Iran
E-mail: m-amerian@araku.ac.ir

Hamid Reza Dowlatabadi

Department of English language and Literature, Faculty of Humanities, Arak University, Arak, Iran
E-mail: H-dowlatabadi@araku.ac.ir

Abstract

This study investigated the effect of first language (L1) metalinguistic comments on Iranian EFL learners' acquisition of third person /-s/ as measured by the learners' performance on a grammaticality judgment test. Accordingly, two intact university classes- randomly assigned to experimental and comparison groups- were homogenized based on their scores on language proficiency and grammatically judgment tests. During 6 hours of task-based instructions, the experimental group was treated by L1 metalinguistic comments and the comparison group through L2 metalinguistic comments. Using ANCOVA and t-test to compare the groups' scores, the results indicated that although the mean accuracy scores of both groups were improved on the immediate posttest, the experimental group outperformed the comparison group on the delayed posttest. It can be postulated that L1 metalinguistic comments lead to more durable gains than L2 metalinguistic comments in EFL classrooms. The findings suggest that L1 use may aid language learners to become more cognitively conscious of L2 grammar features.

Keywords: Accuracy, Durability, First language, Focus on forms, Metalinguistic comment, Task-based instruction

1. Introduction

The inception of form-focused instruction (FFI) as one of the latest approaches in grammar teaching (Trendak, 2015) has aroused great interest among second language acquisition (SLA) researchers and language methodologists to describe its nature and propose different options for its implementation in the classroom. For its nature, considering Ellis's (2005) discussion of direct and indirect interventions, one could argue that FFI belongs to the former because in the direct intervention what learners learn and when they learn are totally specified by the instruction. With regard to FFI implementation options, Brown (2014) adds that FFI options fall on a continuum and that the sides of this continuum represent two different FFI options. On the one side, comes focus on forms activities, that is, activities which are explicit in nature such as direct teachers' presentations of rules or explanation of errors. On the implicit side of this continuum are incidental focus on form activities, noticing devices, and grammar consciousness raising tasks. Ellis (2001) extended the classification of FFI to include another category. In other words, in the new categorization, the key focus of attention (i.e., form or meaning) and distribution of attention (i.e., intensive or extensive) are integrated. Ellis (2008) mentions that one reason why language learners cannot achieve high levels of communicative competence in a communicative classroom can be their failure to pay attention to L2 forms. He adds that this is related to the fact that in such classes learners are engaged with the meaning-focused activities, and as a result, they do not notice L2 forms such as past tense markings. This fact necessitates FFI which can help such students notice L2 forms during a communicative activity. To this, Nassaji and Fotos (2011) add that recent SLA researchers recognize the necessity of attention to grammar and acknowledge that FFI, as grammar instruction, is especially effective once it is integrated into a meaningful communicative event.

Therefore, it seems intriguing to investigate how L1 can be used as an explicit FFI option within the task-based instruction (TBI) of learners in order to help them notice and acquire L2 forms such as third person /-s/. VanPatten (2004) describes third person /-s/ as a feature which is redundant and for which learners may not form the necessary form-meaning connections on the input immediately. Ellis et al. (2009) describe a third person /-s/ as a feature which is late-acquired (based on the result of the previous research) or not acquired at all. They mention that part of the difficulty involved in learning third person /-

s/ lies in the fact that it does not have salience, and has no obvious functional value. Moreover, they attribute part of the difficulty to the feature's advanced processing procedure and to the intermittent exposure to input.

In addition, considering FFI options, one may find that there is a gap in FFI options about an explicit FFI choice using the first language (L1), especially within the context of TBI where English is taught as a foreign language, that is, an EFL context. In this study, thus, we tried to bridge the gap by introducing teacher-presented L1 metalinguistic comments as one of the options of explicit focus on forms whereby the learners' L1 is the medium of the instruction for presenting the rule of the form which was the focus of this study.

2. Literature Review

2.1. Form-Focused Instruction

The last era in grammar instruction, that is form-focused instruction (FFI), represents a paradigm shift in the way grammar is treated in second/foreign language classrooms. Trendak (2015) explicates this paradigm as the revival of interest in grammar during the late 1980s due to the failure of meaning-centred approaches (e.g., Trendak cites Krashen's, 1981, 'zero approach') to develop the grammatical competence of language learners. For Spada (1997, p. 73), the term FFI refers to "pedagogical events which occur within meaning-based approaches to instruction, but in which a focus on the language is provided in either a spontaneous or predetermined ways" and applies to a range of instructional options that coincide with theories on the role of noticing and awareness in L2 learning (Schmidt, 2001; Tomlin & Villa, 1994). Similarly, for Ellis (2001, p. 1), FFI "is used to refer to any planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic form". To this, one may add that the way learners are instructed to process language input may help them notice language forms (VanPatten, 2004). VanPatten (2004) discusses that input processing directs learners' attention to the L2 forms presented in the input.

Long, (as cited in Ellis, 2001 & Spada, 1997) divides FFI into focus on forms and focus on form respectively. According to Ellis (2008), focus on forms characterizes the traditional ways of grammar teaching and represents the kinds of instruction peculiar to the traditional approaches to grammar teaching which adhere to a structural syllabus where it

is assumed that language learning is a process of gathering definite and clear-cut entities. Whereas, according to Long, “focus-on-form overtly draws students’ attention to linguistic elements as they arise incidentally in lessons whose overriding focus is on the meaning or communication” (Ellis, 2001, p. 14).

The accuracy and the durability of the instructed forms have always been of great importance to SLA researchers. Ellis (2008) emphasizes that there exist great concerns for accuracy and durability in all three SLA theoretical paradigms (i.e., Generative linguistics, Interactionalism, and Emergentism) which have tried to measure learning in FFI studies. As Ellis explicates, in all these paradigms, FFI is considered to be effective if learners demonstrate statistically significant gains of the instructed features in terms of accuracy over time.

This study aimed to examine the impact of the L1 metalinguistic comments on the accuracy and durability of the learners’ acquisition of the third person /-s/. As such, an elucidation of the concepts metalinguistic and L1 is in order.

2.2. Metalinguistic Comments

Within SLA research, metalinguistic comments are given in the target/second language (i.e., L2 metalinguistic comments) and are classified as explicit incidental focus-on-form option in FFI methodology (Ellis, 2001). Myhill, Jones, and Wilson’s (2016) research on writing showed that the teachers’ management of the metalinguistic discussions plays a vital role in helping young writers to express meanings in their works and make metalinguistic decisions while writing. Borg (as cited in Trendak, 2015) mentioned that the employment of metalanguage “ought to be adjusted to the learners’ level of language advancement, their cognitive development but also their learning styles” (p. 13). As for the employment of L1 in the second language classrooms, Ellis (2008) argues that L1 is an issue which is both complex and controversial. According to him, complexity depends on the instructional context. Turnbull and Arnett (2002) surveyed teachers’ use of L1. They concluded that exigency exists for investigating the relationship between teachers’ use of L1 and L2 on the learners’ achievement and emphasized that except for some small-scale studies, this relationship has not been investigated. As for the use of L2 metalanguage in the classroom, Ellis (2008) argues that little research has been conducted to investigate the teachers’ use of L2 metalanguage and reports a study by

Basturkmen, Loewen, and Ellis (2002) that found no relationship between teachers' use of L2 metalanguage and L2 acquisition. Trendak (2015) mentions that the use of metalanguage in terms of cross-linguistic comparison is very useful: "comparing the target structure with learners' L1 might arouse interest and, hopefully, make it more memorable" (p. 13). As the literature review shows, in spite of few studies which have scrutinized the role of L2 metalanguage in the classroom, studies have yet to examine the impacts of L1 metalanguage on the accuracy and durability of learners' acquisition of linguistic forms in EFL classroom. Hence, the following research questions were set to be answered by this study:

1. Do Iranian EFL learners who are instructed through explicit teacher-delivered metalinguistic comments in Persian outperform those who receive the same instruction(s) in English in terms of accurate acquisition of English third person /-s/?
2. Do Iranian EFL learners who are instructed through explicit teacher-delivered metalinguistic comments in Persian outperform those who receive the same instruction(s) in English in terms of durable acquisition of English third person /-s/?

3. Methodology

The choice of methodology was made on the basis of Skehan's (1996) TBI framework. That is, each session of the instruction was organized into three phases: pre-task, during-task, and post-task phases. The targeted feature was the third person /-s/ which turned out to be problematic for the groups of the study as determined by the pretest results. Two intact classes, each consisting of 40 Persian speaking male adult EFL learners, were chosen as the experimental and the comparison groups on a random-assignment basis. Non-Persian L1 speakers were excluded and the rest of the participants were homogenized in respect of their language proficiency based on their scores on the Oxford Placement Test (the OPT) and grammar proficiency based on their scores on an Untimed Grammaticality Judgment Test (the UGJT) in order to ensure that they have the same level of language proficiency and grammar knowledge before the onset of the treatment.

After the first testing session (administering the OPT & the UGJT), the instruction was commenced in terms of four teaching sessions. The teaching sessions were divided into three periods: a pre-task activity (10 min), a task phase (50 min), and a post-task phase

(30 min). All classes received the same instruction. During the pre-task phase, the learners were prepared to perform the task (Van den Branden, 2006). Moreover, during this phase, the third person /-s/ and its rule were instructed in Persian to the experimental group, that is L1 metalinguistic comments were utilized as an explicit focus on the form options. For the comparison group, this feature was explained and presented in English through L2 metalinguistic comments. In this way, metalinguistic comments were utilized as an explicit focus on forms option in the pre-task phase and were integrated within the adopted TBI framework. Moreover, the pre-task phase was used to guide the learners about the task phase by making them interested and practice the essential language that was felt to be necessary for task completion. In other words, as Van den Branden (2006) explicates, during this period, learners were prepared to perform the task. The task phase consisted of written enrichment tasks. Four enrichment tasks in the form of input flood texts were developed by the researchers and were instructed to both groups in four sessions (each session around 1 hr 30 min). These enrichment tasks constituted the teaching materials in this study. During this period, the learners were required to read the tasks and answer the comprehension questions based on their texts on a pair-work basis. The instructor referred to the pairs and provided more explanations about how to carry out the task and clarified new vocabulary items if needed. The post-task phase included Presentation-Practice-Production (PPP) procedure to present the third person /-s/ (Ur, 2011) and asking the learners to evaluate and reflect on the task. The PPP was uniformly presented to both groups as a whole-class activity and was aimed to practice the targeted structure which was found to be problematic for the participants based on the results of their pretest scores. Moreover, as a reflection activity, the participants were asked to take into consideration the rule which was presented to them and once again were required to refer to the text of the tasks they had produced in the production stage of the PPP and to find cases where they had applied the rule incorrectly.

After the last teaching session, the UGJT was administered as the posttest to measure the effects of instruction on the immediate acquisition of the third person /-s/. Two weeks later, the UGJT was again administered. This time, the purpose was to determine how durable the instruction had been in terms of acquisition of the instructed feature, that is, whether the participants had acquired the feature or not.

In order to be able to determine the effects of the metalinguistic comments, attempts were made to exert control as much as possible. For this purpose, the same teacher was asked to instruct both groups. It was also attempted to control the learners' language proficiency level, grammar proficiency, L1, major field of study, instructor's methodology, and the instruction material. The only difference in terms of instruction between the two groups was the L1 metalinguistic comments which were utilized to explain third person /-s/ rule to the experimental group of the study. For the comparison group, the metalinguistic comments were presented in English. The following examples are presented in order to elucidate the difference in the treatment between the two groups:

Teacher's L1 metalinguistic comments in the experimental group:

The teacher writes the following English sentence containing the third person /-s/ on the board and explains in Persian the rule of the third person /-s/ to the students as:

He works hard.

T: darse emrouze ma rajebe es sevom shaks ast. Yek ghaede koli dar zaban englisi in ast ke agar jomle daraye fael sevom shaks mofrad manmade He, She, va It bashad, fele jomle bayad motabeghat konad ba fael jomle va -s be fel ezafe mishavad. [English literal translation: Our lesson today is on third person singular in English. As a basic rule in English, we need to know that if the subject of the sentence is He, She, and It, the verb of the sentence should agree with the subject of the verb and -s will be added to the end of the verb].

Teacher's L2 metalinguistic comments in the comparison group:

The teacher writes the following English sentence containing third person /-s/ on the board and elaborates the third person /-s/ rule in English to the students:

She works hard.

T: Today, we are going to learn third person /s/. As a basic rule in English we need to know that if the subject of the sentence is He, She, and It (that is a third person singular subject), the verb of the sentence must agree with the subject; thus, it takes -s ending.

3.1. Design and Context of the Study

This study utilized a quasi-experimental design. The design included a pretest, a posttest, a delayed posttest, one experimental group, and one comparison group. The study was carried out at Shahid Sattari University, Tehran, Iran.

3.2. Participants

Two intact university classrooms were chosen as the experimental and the comparison groups on a random assignment basis. Table 1 describes the participants' background information.

Table 1.

The participants' demographic background

No. of Participants	63
Gender	Male
Native Language	Persian
Major	Accounting
University	Shahid Sattari
Academic Years	2017-2018
Age Range	19-22
Language Proficiency Level	Lower-intermediate

3.3. Instrument(s)

The test battery used in this study included the OPT and the UGJT (used as the pretest, the immediate, & the delayed posttests).

The UGJT was developed and validated by Ellis et al. (2009) and was used to measure the accuracy and durability of the learners' acquisition of the targeted feature. The participants were required to read the recognition items of the test and indicate whether the sentences were grammatically correct or incorrect. The test includes 68 recognition items. As an example, four items of the test are presented below. The participants had to read the recognition items and decide whether they were correct or incorrect:

1. He plays soccer very well.
2. I think that he is nicer and more intelligent than all the other students.
3. The bird that my brother caught it has died.
4. Hiroshi found some keys on the ground.

Four items in the UGJT test for the acquisition of third person /-s/, the focus of the instruction in this study, were referred to as the intended items. They were the items 20, 25, 45 and 59. Items 20 and 45 were grammatically correct and items 25 and 59 were grammatically incorrect. The rest of the items in the test (i.e., the non-intended items) measured other grammatical features on which the groups received no instruction.

Some other studies have also used grammaticality judgment tests for similar purposes. White (as cited in Ellis, 2008) utilized a grammaticality judgment test as the measure of accuracy.

3.4. Data Collection Procedure

The UGJT was used in all testing sessions (the pretest, the immediate posttest, and the delayed posttest). That is to say, different formats of the test were developed by randomizing the items in the test (Loewen, 2009). In other words, the data was collected at 3 time intervals: the pretest, the immediate, and the delayed posttest testing sessions.

3.5. Data Analysis Procedure

Data analyses were run for the participants' scores on the sum of the mean scores of the intended items and the non-intended items of the UGJT. Analyses of the participants' scores on the grammatical and the ungrammatical items were done separately because according to earlier research (Ellis, 2009a), the ungrammatical sentences on the UGJT are a measure of explicit knowledge whereas the grammatical sentences are a measure of implicit knowledge. Following Ellis (2009b), our major focus was on four items of the UGJT, that is, the intended items, as the measures for the acquisition of the third person /-s/. Ellis (2009b) considers learners' accurate judgment of two ungrammatical structures in the UGJT as the measure for the acquisition of that structure.

4. Results

To compare the performance of the two groups, it was necessary to ensure that no significant differences existed between the groups before the beginning of the treatment. Accordingly, to determine whether the difference between the mean proficiency scores of the groups based on the OPT, used at this time as the pretest, was significant or not, an independent samples t-test was run. The t-test results indicated that the difference in the mean proficiency scores was not significant between the experimental group ($M = 17.47$,

$SD = 6.36$) and the comparison group ($M = 17.42$, $SD = 4.08$), $t(62) = -0.037$, $p = .97$ (two-tailed). It should be noted that one of the assumptions of t-test as a parametric test is normality, which was examined by checking skewness and kurtosis ratios from the groups' OPT descriptive data which showed that the ratios were within ± 1.96 ; therefore, the data were considered meeting the normality assumption for running the t-test.

After ensuring the homogeneity of the comparison and the experimental groups, the participants of the two groups were further checked concerning their grammatical proficiency level. In order to determine whether the difference between the grammar mean scores of the groups was statistically significant or not, an independent samples t-test was run. The t-test results showed that there was not any significant difference in the grammar mean scores for the experimental group ($M = 40.44$, $SD = 7.86$) and the comparison group ($M = 40.26$, $SD = 7.62$), $t(62) = -0.092$, $p = .92$ (two-tailed). Normality of distribution was checked for the grammar pretest scores by checking skewness and kurtosis ratios using the same procedure as explained above.

Next, the reliability of the UGJT was examined. Cronbach Alpha showed that the alpha was above .70, hence showing that the test was reliable.

As mentioned earlier, all the analyses of the research questions were done in terms of three groups of test items; that is, the sum of the mean scores of items 20 and 45, the sum of the mean scores of items 25 and 59, and the sum of the mean scores of items 20, 45, 25, and 59 (i.e., the sum of all intended items mean scores). Hence, it was necessary to compare the sum of the pretest mean scores of items 20 and 45, the sum of the pretest mean scores of items 25 and 59 and the sum of the mean scores of the items 20, 25, 45, and 59 with the sum of the mean scores of the said items in the posttest to see whether any improvement had happened from the pretest to the posttest in the comparison and the experimental groups. Moreover, the pretest and the immediate posttest mean scores of those items on the grammar test which were not focused on the treatment (i.e., the non-intended items) were also computed to see whether any improvement had happened from the pretest to the posttest in the comparison and the experimental groups' knowledge of the instructed feature and to see whether the instruction could help them to improve their knowledge of the other grammatical features which were not focused in this study.

The descriptives of the pretests and the posttests means of intended items, as well as the non-intended items for the comparison and the experimental groups, showed that there

were higher grammar mean scores on the posttests of only intended items. To ascertain whether these differences were significant or not, paired samples t-test was run. Before running the t-test, the normality of the data was checked by examining skewness and kurtosis ratios which all turned out to be within ± 1.96 , hence meeting the normality assumption.

Tables 2 and 3 present the results of the paired samples t-tests, which revealed that there was indeed a significant improvement from the pretests to the posttests in all scores of the intended items in both the comparison and the experimental groups ($p < .001$). It needs to be mentioned that the effect sizes for the comparison group were between 0.5 to 0.7, which are medium to large, and the ones for the experimental group were above .8 showing very large effect sizes as per Cohen’s (1988) guidelines for effect size.

Table 2.

Tests of Difference Between Pretest and Posttest Means of Intended Items

Groups		Paired differences			t	df	Sig.(2-tailed)	
		Mean	SD	Std. error				
				Mean				
				Cohen’s d				
Comparison	Pair Pre.sum.Q20&45	-.38710	.55842	.10029	.52	-3.860	30	.001
	1 Im.post.sum.Q20&45							
	Pair Pre.sum.Q25&59	-.48387	.50800	.09124	.51	-5.303	30	.000
	2 Im.post. sum.Q25&59							
	Pair Pre. sum-Im.post. sum	-.87097	.76341	.13711	.7	-6.352	30	.000
	3							
Experimental	Pair Pre.sum.Q20&45	-.84375	.67725	.11972	1.3	-7.048	31	.000
	1 Im.post. sum.Q20&45							
	Pair Pre.sum.Q25&59	-	.62217	.10999	1.8	-9.092	31	.000
	2 Im.post.sum.Q25&59	1.00000						
	Pair Pre. sum-Im.post. sum	-	.88388	.15625	1.8	-11.800	31	.000
	3	1.84375						

Note. Pre = pretest; Im = immediate; Post = posttest.

On the contrary, Table 3 indicates that regarding the non-intended items (as mentioned in the instrument section), no significant change has happened from the pretest to the immediate posttest mean scores of the experimental group ($p = .104$) and the comparison group ($p = .326$).

Table 3.

Tests of Difference Between Pretest and Posttest Means of Non-intended Items

Groups	Paired differences		Mean	SD	Mean	t	Df	Sig.(2-tailed)
	Pretest.sum.non	Im.posttest.sum.non						
Comparison	1	1	-	1.79437	.32228	-.998	30	.326
Experimental	1	1	-	1.60111	.28304	-1.677	31	.104

Note. Non = non-intended items; Im = immediate; Post = posttest.

The results of Tables 2 and 3 showed that when the comparison and the experimental groups were considered separately, EFL learners who were instructed through explicit teacher-delivered L1 metalinguistic comments (i.e., the experimental group) and those EFL learners who received the same instructions in English (i.e., the comparison group) in terms of accurate acquisition of the targeted feature in the sum of the mean scores of items 20 and 45, the sum of the mean scores of items 25 and 59, and the sum of the mean scores of items 20, 45, 25, and 59 (i.e., the sum of the all intended items mean scores), improved significantly from the pretest to the posttest regarding only the intended items.

Because Tables 2 and 3 showed that the two groups did not have equal pretest scores on the sum of the mean scores of items 20 and 45, the sum of the mean scores of items 25 and 59, and the sum of the all intended items mean scores, the pretest initial differences were all considered as covariate, and hence multivariate analysis of covariance (MANCOVA) was run to take into account the initial differences.

Table 4 indicates the results of the multivariate test on the initial differences on the pretests and the posttests taking all the dependent variables together. Accordingly, the results were significant ($p < .001$) with partial eta squared effect sizes larger than .25 showing large effect sizes.

Table 4.

Multivariate Tests Between the Pretest and the Posttest Means

Effect	Value	F	Hypothesis		Error df	Sig.	Partial eta squared
			df				
Group	Pillai's Trace	.389	18.499 ^b	2.000	58.000	.000	.389
	Wilks' Lambda	.611	18.499 ^b	2.000	58.000	.000	.389
	Hotelling's Trace	.638	18.499 ^b	2.000	58.000	.000	.389
	Roy's Largest Root	.638	18.499 ^b	2.000	58.000	.000	.389

In order to check each set of the intended items separately (i.e., the sum of the mean scores items of 20 & 45, the sum of the mean scores of items 25 & 59, and the sum of the all intended items mean scores) a univariate test (i.e., a separate ANCOVA) was run for each dependent variable. The results of the Levene's test on the quality of groups' variances in terms of each dependent variable were not significant ($p > .05$), hence meeting the assumption of homogeneity of variances.

The main ANCOVA univariate results demonstrated that the two groups were significantly different on the posttest in terms of the sum of the mean scores of items 20 and 45, $F(1, 60) = 12.322, p < .001$, partial eta squared = .175, the sum of the mean scores of items 25 and 59, $F(1, 60) = 18.201, p < .001$, partial eta squared = .239, and the sum of all the intended items mean scores, $F(1, 60) = 37.369, p < .001$, partial eta squared = .39, hence supporting the MANCOVA results above.

On the other hand, the two groups did not differ significantly from each other in terms of the non-intended items means $F(1, 60) = .245, p = .622$, partial eta squared = .004. Taken together, all these results pointed out that the experimental group

outperformed the comparison group regarding the immediate acquisition of the targeted feature.

To answer the second research question, it was required to compare the sum of the immediate posttest mean scores of items 20 and 45, the sum of the immediate posttest mean scores of items 25 and 59, and the sum of the all intended items immediate mean scores with the scores of the said items in the delayed posttest to see whether any durable acquisition had happened from the immediate posttest to the delayed posttest in the comparison and the experimental groups. It should be mentioned that the same comparison was also done in terms of the non-intended items to make the necessary comparisons with the immediate and delayed posttest means of the intended items.

The descriptive statistics of the sum of the immediate posttests and the delayed posttests means of items 20 and 45, the immediate posttests and the delayed posttests means of items 25 and 59, and the sum of the all intended items immediate and delayed mean scores for the comparison and the experimental groups as well as their non-intended items means showed that there were some lower grammar mean scores on the delayed posttests. In order to ascertain whether the observed differences were significant or not, the paired samples t-test was run (Tables 5 & 6).

Table 5.

Tests of Difference Between the Posttest and the Delayed Posttests Means of Non-intended Items

Groups	Pair	Im.post.sum.Q20&45	Paired differences				Sig. (2-tailed)	
			Mean	SD	Mean	T		
Comparison	1	Del.post.sum.Q20&45	-.22581	.49730	.08932	2.528	30	.017
	2	Del.post. sum.Q25&59	-.29032	.46141	.08287	3.503	30	.001
Comparison	3	Im.post. sum	.51613	.67680	.12156	4.246	30	.000
	3	Del.post. sum						

Experimental	Pair	Im.post.sum.Qs20&45	-.12500	.65991	.11666	1.072	31	.292
	1	Del.post.sum.Qs20&45						
	Pair	Im.post.sum.Q25&59	-.06250	.50402	.08910	.701	31	.488
	2	Del.post.sum.Q25&59						
	Pair	Im.post. sum	.18750	.85901	.15185	1.235	31	.226
	3	Del.post. sum						

Note. Im = immediate; Post = posttest; Del = delayed.

Table 6.

Tests of Difference Between the Posttest and Delayed Posttests Means of Intended Items

Groups	Comparison	Pair	Paired differences			T	df	Sig. (2-tailed)	
			Mean	SD	Std. error				
		Pair	Im.posttest.sum.non	-.11957	1.15244	.20698	.578	30	.568
		1	Delayed.post. sum.non						
Experimental		Pair	Im.posttest.sum.non	-.08547	.56043	.09907	.863	31	.395
		1	Delayed.post.sum.non						

Note. Im = immediate; Post = posttest.

The above analyses (Tables 5 & 6) only showed the comparison of the immediate posttest and the delayed posttest means in each group separately. To compare the long-term acquisition of both groups in terms of the intended and the non-intended items and to see which group showed more durable acquisition of the targeted feature and because the two groups did not have equal immediate posttests means, multivariate analysis of covariance (MANCOVA) was run to take into account the initial differences (Table 7).

Table 7.

Multivariate Tests between the Immediate and Delayed Posttest Means

Effect	Value	F	Hypothesis			Partial eta squared	
			df	Error df	Sig.		
Group	Pillai's Trace	.229	8.615 ^b	2.000	58.000	.001	.229
	Wilks' Lambda	.771	8.615 ^b	2.000	58.000	.001	.229
	Hotelling's Trace	.297	8.615 ^b	2.000	58.000	.001	.229
	Roy's Largest Root	.297	8.615 ^b	2.000	58.000	.001	.229

In order to check each set of the intended items separately (i.e., the sum of the mean scores of items 20 & 45, the sum of the mean scores of items 25 & 59, and the sum of the all intended items mean scores) as well as the non-intended items mean scores a univariate test (i.e., a separate ANCOVA) was run for each dependent variable. The results of Levene's test of the quality of groups' variances in terms of each dependent variable were not significant ($p > .05$), hence meeting the assumption of homogeneity of variances.

The main ANCOVA univariate results demonstrated that the two groups differed significantly on the posttest in terms of the sum of the mean scores of items 20 and 45, $F(1, 60) = 6.454$, $P = .014$, partial eta squared = .100, the sum of the mean scores of items 25 and 59, $F(1, 60) = 11.302$, $p < .001$, partial eta squared = .163, and all intended items mean scores, $F(1, 60) = 1.301$, $p < .001$, partial eta squared = .172, hence supporting the MANCOVA results above. On the other hand, the two groups did not differ significantly from each other in terms of the non-intended items means, $F(1, 60) = .747$, $p = .391$, partial eta squared = .013.

In sum, the results indicated that the comparison group showed a significant decline in their mean grammar scores on the intended items from the immediate to the delayed posttests; however, the experimental group showed no significant decline in their mean grammar scores on these items from the immediate to the delayed posttest. These results indicate that the instruction had been more effective for the experimental group. Regarding the other items of the test which do not test the third person /-s/ (i.e., the non-intended items) both groups showed no decline from the immediate posttest to the delayed posttest

and as was already mentioned, the results of the main ANCOVA showed that there was no significant difference between the two groups' mean scores on the non-intended items.

5. Discussion

Considering FFI methodology, one may notice that there is a dearth in terms of an explicit focus on the form option which draws on the learners' L1, especially within the context of TBI. Teachers may be in doubt whether to utilize the learners' L1 within TBI frameworks. To this end, the L1 metalinguistic comment was proposed, as an explicit focus on the forms option, and its effects on the accuracy and the durability of Iranian EFL learners' acquisition of the third person /-s/ were investigated. The reason for focusing on a specific FFI option, as Reinders and Ellis (2009) argue, is that although studies that have investigated FFI involving a combination of a number of options may enjoy ecological validity, as far as SLA theory testing is concerned, these studies are problematic. Reinders and Ellis (2009), on the basis of Norris and Ortega's (2000) meta-analyses, explicate that the key feature which contrasts the different types of FFI are operationalized in an inconsistent manner and that such a situation can only be resolved if researchers test the effects of distinct FFI options. According to Reinders and Ellis, if research is to determine what effects various types of input have on learning a second language, one is required to separate particular instructional strategies and test their impacts on L2 acquisition.

Ellis (2009c) adds that one difficulty in FFI research is that it seeks to conform both SLA theory and language pedagogy. He, following Reinders and Ellis's (2009) line of argument, mentions that if the purpose of FFI research is a testing theory, it is imperative to conduct experiments that test the effects of a narrowly defined instructional option (as it is the case in our study). However, if the goal is to determine the role FFI plays in language pedagogy, studies should be designed in a way that they combine instructional options that mirror current best practice. Thus, according to Ellis (2009c), what is useful for the testing theory may not be good for informing pedagogy and vice-versa. Ellis's discussion provides support for this study because it tested the effects of a narrowly defined instructional option, L1 metalinguistic comments. Moreover, the instructional option was used in the pre-task phase as an explicit focus on the form option within TBI framework which reflects the current best practice considering the proposals made in favor of TBI. For instance, Long (2015) mentions that as an approach to language teaching (LT), TBI satisfies the

minimum seven criteria that a new approach to LT should have. These criteria are consistent with SLA theory and research findings, having a basis in the philosophy of education, accountability, relevance, avoidance of known problems with existing approaches in LT, the learner centeredness, and functionality.

To find the answers to the research questions raised in this study, as mentioned before (i.e., the methodology section), a quasi-experimental design was adopted involving one experimental group and one comparison group. During 6 hours of TBI, the experimental group was instructed through L1 metalinguistic comments and the comparison group through L2 metalinguistic comments,

The analyses of the data revealed that the subjects in the experimental group demonstrated better performance than subjects in the comparison group based on their scores on the intended items of the immediate posttest. Consequently, the answer to the first research question is that L1 metalinguistic comments positively affect the learners' acquisition of the targeted feature. The mean of the accuracy scores of the comparison group was also increased on the immediate posttest; however, it was dropped on the delayed posttest. The fact that the comparison group's mean scores were improved on the immediate posttest is explainable in terms of the effect of the treatments they were exposed to. These treatments included L2 explicit metalinguistic comments, enrichment tasks, and form-focused activities in the post-task phase. Considering the fact that both groups' mean accuracy scores were raised on the immediate posttests, one may argue that the metalinguistic comments, whether presented in L1 or L2, had been effective in promoting the learners' gains on the immediate posttests. Earlier research (Myhill et al., 2016) has also confirmed the beneficial value of metalinguistic discussions fostered by language teachers in improving learners' writing scores. In addition, the groups under this study were exposed to the different FFI options during the task phases. Hence, the findings are in agreement with the beneficial effects of FFI reported empirically (Norris & Ortega, 2000). However, the failure of the comparison group to retain their performance on the second posttest (i.e., the delayed posttest) stands in opposition to the Norris and Ortega's (2000) findings which point to the durability of FFI over time. This may suggest that not all types of FFI are durable over time, especially if they are given in L2 to those learners whose language proficiency level is lower-intermediate. For example, Loewen, Erlam, and Ellis's (2009) study showed that the English third person /-s/ cannot be acquired incidentally

despite the fact that subjects received substantial exposure to this feature. Hence, there was no increase in the subjects' accuracy scores. Ellis et al. (2009, p. 239) interpret these findings as "the special difficulty of third person /-s/".

Moreover, the results of data analyses pointed that the participants instructed through the L1 metalinguistic comments performed better than the participants who received such comments in L2 comparing their delayed posttests mean scores on all the dependent variables of this study. Thus, the answer to the second question is that although L2 metalinguistic comments had been effective in promoting and maintaining the comparison group's mean scores on the delayed posttest, the L1 metalinguistic comments led to more durable gains than their L2 counterparts.

There are three possible explanations for the superior performance of the experimental group. One is the role of orientation as expressed in Tomlin and Villa's (1994) theory of attention. They (1994, p. 199) put forward that "the learner may bias attentional resources to linguistic form, increasing the likelihood of detecting formal distinctions, but perhaps at the cost of failing to detect other components of input utterances". In this way, L1 has oriented the learners' attention mechanism toward the L2 form and consequently has led to a higher degree of acquisition of the third person /-s/ on the posttests. In the same vein, Cummins (2007) emphasizes that learning can be enhanced if teachers make their learners notice resemblances and dissimilarities between their L1 and L2 and if they fortify productive learning strategies in a balanced way between L1 and L2.

The next is Schmidt's (2001) concept of attention and noticing in driving the second language acquisition process. Schmidt considers noticing a key preliminary step in language learning and uses the meaning of noticing in a restricted sense to refer to becoming aware of the "surface structure of utterances in the input-instances of language, rather than any abstract rule or principles of which such instances may be exemplars" (Schmidt, 2001, p. 5). On the basis of Schmidt's reasoning, we may suggest that L1 has caused the learners to more noticing of the third person /-s/ and this has caused "the initial registration of this feature into conscious awareness" (Mitchell, Myles, & Marsden, 2013, p. 147). Thus we can argue that although noticing might have taken place in both groups, as the comparison of pre/post tests showed, as far as the results of this study are concerned, L1 metalinguistic comments seem to have been more graspable for the experimental group

comparing their performance with the performance of the comparison group taking into consideration the language proficiency level of both groups—lower intermediate. Thus, L1 has been more effective than L2 in paving the way for noticing process for the experimental group. In other words, to quote Izumi (2002), L1 has been more effective than L2 in making subjects notice the mismatch between their developing L2 system and the L2 considering the participants' proficiency levels in this study. Nassaji and Fotos (2011), referring to Fotos' earlier works in 1993 and 1994, mention that high levels of noticing are related to learners' language proficiency. Based on this argument, L1 may help those learners whose language proficiency level is below lower-intermediate to reach high levels of noticing of L2 features and in this way it can be argued that L1 plays a complementary role to L2.

The third is VanPatten's (2004) theory of language processing. For VanPatten (2004) processing is to make a linkage between form and meaning by learners through noticing a form and concurrently determining its meaning or function. According to language processing instruction, as Ellis (2008) makes it explicit, meaning and form are in a competition to attract learners' attention to themselves and only once learners can perceive the intended meaning of the input, they are able to pay attention to form.

Moreover, as Ellis (2008, p. 846) under the weak interface hypothesis discusses, instruction is facilitative because it helps learners to pay selective attention to form and form-meaning connections in the L2. On the basis of VanPatten's model and the above-mentioned Ellis's argument on the effects of instruction, it can be argued that L1 metalinguistic comments have been more effective than L2 metalinguistic comments in order to help the experimental group to make stronger mappings between the form and the meaning of the third person /-s/ and has paved the way to understand its meaning more easily for the experimental group. In other words, L1 has rendered form-meaning connections more explicit to the learners. In this way, L1 metalinguistic comments led to more durable effects on the delayed posttest.

Although we discussed the positive role of L1 in drawing learners' attention to L2 forms, the other side of the coin should be heeded. In the intermediate stages of L2 learning, when the learners' interlanguage is not completely developed, presenting grammar rules in L2 could be difficult to grasp for the learners, and this may have faced the comparison group with difficulty in learning the instructed feature. Consequently, this

has resulted in their poor performance on the measures of accuracy and durability. In this respect, Loewen et al. (2009) emphasized that learners (especially low-proficiency ones) face difficulty attending to two aspects (form & meaning) of the input simultaneously due to limited processing capacity. VanPatten (2004) also adds that both comprehension and processing are effortful for beginning and even intermediate learners. As far as the third person /-s/ is concerned, VanPatten explains that this feature performs a semantic function, describing a person which is neither the speaker nor the person addressed by the speaker, and due to this semantic notion, this feature becomes redundant. Hence, the learners may not make the specific form-meaning connection in the input immediately. This redundancy makes third person /-s/ a feature difficult to be learned for the learners right away.

6. Conclusion

Considering the lacuna in terms of an explicit FFI option which draws on learners' L1 in TBI, the L1 metalinguistic comment was introduced as an FFI option and its impact on promoting the accuracy and the durability of Iranian EFL learners' acquisition of English third person /-s/ was explored. This study pointed to the value of L1 in helping learners to become more cognitively aware of the English language forms. The findings demonstrated that teacher's use of L1 may play a facilitative function in the foreign language learning process (Cummins, 2007; Izumi, 2002).

Hence, a reasonable conclusion which can be drawn is that an explicit FFI component (the L1 metalinguistic comment) along with a task-based framework (enrichment tasks) could lead to higher achievements in EFL classrooms. This is in line with Ur's (2011) conclusion on explicit language teaching and its effects on the learners' achievement. Further, the results of our study suggest the pedagogical value of L1 in EFL contexts, especially when dealing with lower-intermediate learners. Language teachers may occasionally embark on learners' L1 to clarify and explain the complicated L2 grammatical forms. Thus, the L1 metalinguistic comments may be suitable for students with limited language proficiency and language teachers who prefer such manners of instruction (Loewen, 2011). These findings also have implications for processing instructions. VanPatten (2004) mentions that processing and comprehension require a lot of effort for beginners and even intermediate students. Considering the findings of this study, methodologists and researchers may develop intervention processing techniques

using learners' L1 to see how these techniques help students make form and meaning connections during comprehension.

The fact that L1 appears in almost any discussions on SLA indicates that L1 cannot be easily dismissed as a non-pivotal variable in any foreign or second language acquisition research. In this regard, as Savignon (2005, p. 641) mentions: "teacher rejection of research findings, renewed reliance in the classroom on the learners' native or first language, where possible, to be sure they 'get the meaning', have been in some cases reactions to the frustrations of teaching for communication".

The findings could be more fruitful if the performance of the learners in a writing test be measured. They could also be more appealing if the study had included learners at different proficiency levels. It would be interesting to see whether L1 has the same effects for learners with different language abilities. These limitations point to the need for future research.

References

- Brown, H. D. (2014). *Principles of language learning and teaching: A course in second language acquisition* (6th ed.). New York, NY: Pearson.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics*, 10(2), 221-241.
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, 51, 1-46.
- Ellis, R. (2005). Instructed language learning and task-based teaching. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 713-728). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, UK: Oxford University Press.
- Ellis, R. (2009a). Investigating learning difficulty in terms of implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 143-166). Bristol, UK: Multilingual Matters.

- Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 31-64). Bristol, UK: Multilingual Matters.
- Ellis, R. (2009c). Retrospect and prospect. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 335-353). Bristol, UK: Multilingual Matters.
- Ellis, R., Loewen, S., Elder, C., Erlam, R., Philip, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning and teaching*. Bristol, UK: Multilingual Matters.
- Izumi, S. (2002). Output, input enhancement and the noticing hypothesis. *Studies in Second Language Acquisition*, 24(4), 541-577.
- Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 94-112). Bristol, UK: Multilingual Matters.
- Loewen, S. (2011). Focus on form. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 576-592). New York, NY: Routledge.
- Loewen, S., Erlam, R., & Ellis, R. (2009). The incidental acquisition of third person -s as implicit and explicit Knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 262-281). Bristol, UK: Multilingual Matters.
- Long, M. (201^o). *Second language acquisition and task-based language teaching*. West Sussex, UK: John Wiley & Sons.
- Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. New York, NY: Routledge.
- Myhill, D., Jones, S., & Wilson, A. (2016). Writing conversations: Fostering metalinguistic discussion about writing. *Research Papers in Education*, 31(1), 23-44.
- Nassaji, H., & Fotos, S. (2011). *Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative contexts*. New York, NY: Routledge.

- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
- Reinders, H., & Ellis, R. (2009). The effects of two types of input on intake and the acquisition of implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 281-302). Bristol, UK: Multilingual Matters.
- Savignon, S. J. (2005). Communicative language teaching: Strategies and goals. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 635-652). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Schmidt, R. (2001). Attention. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge Applied Linguistics.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17(1), 38-62.
- Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30, 73-87.
- Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16(2), 183-203.
- Trendak, O. (Ed.). (2015). *Exploring the role of strategic intervention in form-focused instruction*. London, UK: Springer.
- Turnball, M., & Arnett, K. (2002). Teachers' use of the target and first languages in second and foreign language classrooms. *Annual Review of Applied Linguistics*, 22, 204-218.
- Ur, P. (2011). Grammar teaching: research, theory, and practice. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 507-522). New York, NY: Routledge.
- Van den Branden, K. (Ed.). (2006). *Task-based language education: From theory to practice*. Cambridge, UK: Cambridge University Press.
- VanPatten, B. (Ed.). (2004). *Processing instruction: Theory, research and commentary*. London, UK: Lawrence Erlbaum Associates.