

Nonverbal Prompts: A Remedy for Speaking Grammatical Inaccuracy

Arshya Keyvanfar
Fatemeh Azimi

Abstract

This study investigated the impact of nonverbal prompts on the speaking grammatical accuracy of Iranian male and female EFL learners. To this end, 80 EFL learners who were students of the intermediate level of a language school in Tehran and were randomly assigned to two experimental and two control groups took part in this study. The four groups were statistically compared in terms of their general language proficiency level and also speaking skill, indicating that they were eligible for this study. Two control groups of male and female participants received recasts and two experimental groups of male and female participants received nonverbal prompts in the form of body language and facial expressions. The results showed that using nonverbal prompts did significantly improve the speaking grammatical accuracy of the two experimental groups of male and female participants. Furthermore, female learners receiving nonverbal prompts outperformed their male counterparts. To investigate the degree of the impact of nonverbal prompts on male and female learners, the researchers conducted two regression analyses which revealed that male learners benefited more from the experimental condition than the female learners. Finally, this study showed that the speaking improvement of female learners, unlike that of male learners, is not influenced by the type of feedback they receive during instruction.

Keywords: nonverbal prompts, corrective feedback, recasts, speaking grammatical accuracy

Introduction

Everyone makes mistakes, even speakers using their own language when they are lost for words or forced into inappropriate language by a difficult or unusual situation. It is generally agreed that correction is a part of the teaching/learning process, but that over-correction and poor correction techniques can be demotivating for the learner and may lead to reluctance to

try out the new language or even to speak at all. Teachers need to make informed decisions about what, when, and how to correct in order to help learners improve their speaking skills without damaging their confidence.

As a result, one of the questions that second language teachers most often address to second language researchers is what to do about error correction: how and when should they correct whom, if at all? Error correction is not only of practical importance, but is also a controversial issue in the second language acquisition literature where it is often subsumed under the more general term “negative evidence” (Panova & Lyster, 2002).

The explicitness of this negative evidence or “an interlocutor’s interactional move that indicates... any non-target-like feature in the learner’s speech” (Iwashita, 2003, p. 2) is thus the focus of many studies. It has been suggested that corrective feedback which contains explicit evaluations of learners’ errors and makes the students reformulate their output have been effective for learning difficult features (Carroll & Swain, 1993) or features with low perceptual salience (Ellis, Loewen, & Erlam, 2006). Other works have shown that more implicit approaches such as reformulations of students’ errors have positively influenced their learning of communicatively relevant forms (Han, 2002; Mackey & Philp, 1998).

The superiority of corrective feedback (CF) techniques that push learners to self correct was reported by Havranek (1999) in a classroom study. Results of this study revealed that prompts (referred to as elicitation) were more successful in both respects (i.e. they were recalled more often and were more facilitative of language development). The most successful format of correction, both for the learners receiving the feedback and for their peers, is the feedback which successfully elicits self-correction in practice situations. Of the learner characteristics taken into consideration, verbal intelligence, relative language proficiency, and the learners’ attitude towards correction proved to be most influential (Havranek & Cesnik, 2001).

Long, Inagaki, and Ortega (1998) and Philp (2003) show relationships between the effectiveness of feedback and the type of error that it addresses and the stage of learner language development at which the feedback is provided. Most of these studies have focused on the impact of CF on learners’ ability to correct their production and have done so primarily in face-to-face oral interaction.

The target feature and context are two additional variables that require consideration because previous research has revealed that the effects of a CF technique can be selective and can vary from one context to another

(Ammar & Spada, 2001). Only continued, systematic research designed to examine these variables will provide definitive information as to which CF techniques are more effective.

Kinds of Error Correction

Lyster and Ranta (1997) outline seven approaches to providing corrective feedback:

1. **Explicit error correction** occurs when the teacher directly indicates that what the student had said was incorrect. Such explicit negative feedback is sometimes introduced by phrases such as “Oh, you mean X” or “You should say Y”. Explicit error correction therefore is characterized by an intentional and clear indication of an error and providing the target-like reformulation. It provides both negative and positive evidence, helping learners notice the gap between their interlanguage and the target-like form. However, explicit error correction reduces the need for the learner to produce a modified response.
2. **Metalinguistic feedback** distracts the focus of conversation towards rules or features of the target language and falls at the explicit end of the corrective feedback spectrum. Comments, information, or questions related to the well-formedness of the student’s utterance, without explicitly providing the correct form, can be considered metalinguistic feedback. Despite its name, however, metalinguistic feedback does not need to contain metalanguage; in other words, for metalinguistic feedback the inclusion of metalanguage is not the defining characteristic.
3. **Elicitation** is a correction technique that a teacher or interlocutor uses to prompt the learner to self-correct. It is usually accomplished in one of three ways during face-to-face interaction: (a) through requests for reformulation of an ill-formed utterance (e.g. Say that again?; Did you say that right?), (b) through the use of open questions (e.g. How do we say X in French?), and (c) through the use of strategic pauses to allow a learner to complete an utterance.
4. **Repetition**, as the name suggests, is a teacher’s or interlocutor’s repetition of the ill-formed part of the student’s utterance usually with a change in intonation. Through dividing the non-target-like part of the learner’s earlier utterance and usually including intonational or visual improvements to determine the precise location of the error, repetition

can be useful in helping learners to recognize the corrective intent of the repetition and to attempt a more target-like reformulation.

5. **Recasts** are defined as reformulations of all or part of a learner's non-target-like utterance while retaining the learner's intended meaning. Recasts resemble explicit error correction because they also include the full target-like reformulation of the initial error; however, recasts are also distinctive in that they are not introduced by phrases signaling their corrective intent and hence tend to fall on the implicit end of the corrective feedback spectrum.
6. **Translations** are corrective feedback that are provided in response to a student's unsolicited use of their first language. That is to say, while recasts are provided in response to a learner's ill-formed utterance in the target language, translations are generated in response to learner's well-formed utterance in a language other than the target language. Like recasts, the lack of clear indicators of an error places translation toward the implicit end of the corrective feedback spectrum, though the degree to which translations are communicatively obtrusive can also vary. Compared to recasts and other feedback types, translations are relatively rare in the face-to-face classroom and have thus not received the same critical attention.
7. **Clarification requests** are a feedback type that can refer to problems in either comprehension, accuracy, or both. They indicate that a misunderstanding in meaning has occurred. The ordinary function of clarification requests as a discourse move in conversation makes this kind of corrective feedback the least communicatively obtrusive and therefore perhaps the most implicit. This means that clarification requests, unlike other feedback approaches, seek clarification of meaning, as well as form. Typical requests for clarification may take the form of "I'm sorry", "Pardon?", or "I don't understand" in spoken interaction.

Only four of the seven feedback approaches described above (explicit error correction, metalinguistic feedback, recasts, and translations) actually provide learners with information related to the reformulation of the target form. Of these four, metalinguistic feedback is the only one which does not supply the learner with the actual target-like reformulation; instead, metalinguistic feedback provides learners with clues or questions on how to approach the reformulation (Sauro, 2007).

Moreover, Ellis, Loewen, and Erlam (2006) argue that it is the explicitness of metalinguistic feedback which causes the promotion of “the cognitive comparison that aids learning” (p. 34). Thus, one would expect that learners who received corrective feedback that eliminated the requirement to produce pushed output (e.g. recasts) would not show improvements in the production of the target form comparable to those seen in learners who received feedback that required them to modify their own output. However, it is not clear whether the lower level of learning found among learners who were given recasts resulted from a lack of pushed output or from some other feature of the feedback types.

In face-to-face spoken interaction, the ambiguity of the corrective intent of recasts and limitations in working memory capacity may prevent the learners’ ability to use recasts in ways that enable them to make effective cognitive comparisons. It is these effective properties of recasts expressed during face-to-face interaction that put physical gestures at an advantage for encoding them in ways that facilitate cognitive comparison (Nicolas, Lightbown, & Spada, 2001). Thus, while the effectiveness of spoken recasts may be prevented by these limitations, properties of physical gestures, such as enhanced perceptual salience, increased processing, and enduring trace, may compensate for the disadvantages of recasts and may therefore facilitate a deeper level of processing.

Recasts versus Prompts

Prompts as an alternative type of feedback have been usually compared with recasts in classroom settings (Lyster, 2007). Prompts provide signals that stimulate learners to self-repair rather than providing them with a correct reformulation of their non-target utterance, as do recasts. Prompts include clarification requests, repetition of learner error, metalinguistic cues, and elicitation moves (as discussed above). Several classroom studies have shown prompts to be more effective than recasts. For example, Havranek and Cesnik (2001) found repair following prompts to be the most effective feedback combination in a range of EFL classrooms.

Furthermore, when comparing the effectiveness of explicit and implicit negativeness, Lyster (2004) noted the advantages of prompts over their more implicit opponents – recasts – and found that learners receiving prompts showed greater acquisition than those provided with recasts on both written and oral tasks. He interpreted this as the result of prompts’ capability of enabling students to self-repair by using accurate forms. Recasts, despite

allowing learners to hear target forms repeatedly in input, rarely make them notice and correct their own mistakes.

Another important finding that emerges from the previously mentioned empirical research is that prompts can more positively affect learners' accuracy than recasts (Havranek & Cesnik, 2001; Lyster, 2004). For example, research shows that CF techniques that push learners to self-correct can be effective particularly with low proficiency learners. As Ammar (2003) and Lyster (2004) found in their studies, prompts helped lower proficiency learners more than higher proficiency ones who benefited similarly from both recasts and prompts.

The role of gender too is among the factors that require further investigation. Some researchers believe that females are better language performers in almost all the areas of EFL learning (e.g. Ehrlich, 1997). However, no simple answer has been formulated as to which feedback technique is more effective for males and females. In this regard, the following questions can be raised: Is there any difference between the ways male and female foreign language learners react to nonverbal prompts and recasts? If there is, which technique is more effective and contributes to the superiority of the males and females?

In the present study, the objective was to investigate the effect of feedback in the form of nonverbal prompts on male and female EFL learners based on the assumption that female and male learners may respond differently to the teacher's corrective feedback. Thus, the following research hypotheses were formed to explore the impact of prompts in isolation and in interaction with gender on the speaking grammatical accuracy of Iranian EFL learners:

1. There is no significant difference between the speaking grammatical accuracy of the group of EFL learners exposed to nonverbal prompts and that of the group exposed to recasts.
2. There is no significant difference between the speaking grammatical accuracy of male and female EFL learners who are exposed to nonverbal prompts.

Method

In order to verify the two aforementioned hypotheses of this study, a series of measures were taken which are described below.

Participants

The participants of this study were 80 Iranian intermediate EFL learners studying at a language school in Tehran. To ensure the homogeneity of the participants, a norm-referenced placement test of the New Interchange series, which included reading, listening, language use, essay writing, and oral interview (speaking) was administered to the participants. Since the results of the mentioned test indicated that the participants were homogeneous with respect to their general proficiency, they were randomly assigned to two control groups, one consisting of 18 females and the other 22 males, and two experimental groups, one including 19 females and the other 21 males. Their ages ranged from 19 to 27, and 90% of them were university students or graduates. Most of them said they were studying English to gain better job opportunities, emigrate to an English-speaking country, or follow their educational goals. This information was obtained via a questionnaire that all those who register in the language school have to routinely fill out as they enroll. It is also worth mentioning that 70 other intermediate students who had enrolled at the intermediate level of the same language school the previous semester took the proficiency test mentioned above in a pilot study.

Instrumentation

Two tests were used in this study: To make sure the participants in the four groups belonged to the same population in terms of proficiency level, the researchers utilized the norm-referenced placement test of Richards' (2005) *New Interchange* series. In its multiple-choice section, the test contained 20 items of listening, 20 reading, and 30 language use. In the other section, there was an essay writing task of eight points and an interview of 12 points (taken as the speaking grammatical accuracy test in this study and administered both inside the battery for subject selection at the outset of the study and as the posttest). The listening, reading, and language use sections of this test had already been piloted with 70 students studying at the

intermediate level of the same language school the semester before the study was conducted and as a result one listening item and one language use item had been deleted, giving the whole test a maximum score of 88.

In order to determine the level of the speaking grammatical accuracy of the participants both in the subject selection and in the posttest, two professional interviewers of the language school were requested to conduct the interviews for 10 minutes using the New Interchange Passages Placement Testing Program which includes the conversation placement test as well as its assessment scale. Each participant received three tasks during the interview and the selection of subsequent tasks was adaptive; that is, it depended on how successful the interviewee would perform on the previous task.

Every interview was recorded and subsequently rated by two different raters. Each sentence would receive a point varying from zero, for a very ungrammatical overall structure, to two, for a very well structured compound or complex sentence. These points were added up by each rater at the end of listening to the recording of the interview and a final score of each participant's grammatical accuracy in speaking was thus achieved. The raters in this study were one male and one female teacher with more than 10 years of experience in teaching/testing in the Iranian EFL setting. They were fully familiar with the current study and its purpose. The inter-rater reliability index was computed prior to the main study.

Procedure

Since the researchers needed to select and homogenize the participants of the study, they first embarked on piloting the listening, reading, and language use sections of the norm-referenced placement test of Richards' (2005) *New Interchange* series described above with 70 students at intermediate level of the same language school where the study was conducted. Subsequently, the test was administered to the 80 target participants who were intermediate students of a language school in Tehran. The results of the test indicated that all the participants at this level were homogeneous. Thus, they were randomly assigned to four different groups: two control groups, one including 18 females and the other 22 males, and two experimental groups, one consisting of 19 females and the other 21 males.

The speaking test used both for the subject selection process at the outset and the posttest in this study was the model presented by Richards in

the New Interchange series. Each interview lasted 10 minutes starting with the interviewer introducing himself/herself. S/he would then ask the participant to introduce himself/herself, give a little background of what s/he did, and describe how long s/he had been studying English. Based on their initial impression of the participants' speaking ability, the interviewers would then select a task from among 12 different tasks in New Interchange at a level judged to be appropriate to the proficiency level of the participant. The 12 tasks had already been ranked in terms of their difficulty level in the original test battery. Examples of these tasks were:

- Giving advice to a tourist visiting one's country/city
- Describing one's leisure activities
- Talking about travel
- Describing hopes and plans for the future

Each participant received three tasks. Throughout the conversation, the selection of tasks depended on the participant's performance on prior tasks. For example, if an interviewee provided an excellent response to one or several questions within a level, the interviewer would move on to a task at a higher level, and vice versa.

Every interview was recorded and subsequently rated by two different raters. As noted before, each sentence received a point varying from zero for a very ungrammatical overall structure to two for a very well structured compound or complex sentence. These points were added up by each rater at the end of listening to the recording of the interview and a final score of each participant's grammatical accuracy in speaking was thus achieved and converted to a scale of zero to 12 based on the instructions of the test battery.

The instructional intervention consisted of 18 sessions of 90 minutes each, which spread out over three weeks. The course was intensive with two sessions a day held three days a week. The two control groups (one with male and one with female participants) received feedback mainly in the form of recasts while the two experimental groups of male and female participants received nonverbal prompts as another form of feedback. These nonverbal prompts appeared in the form of a certain set of arbitrarily designated body gestures that were demonstrated immediately after the grammatical inaccuracy occurred as the learners were speaking. For instance, if a learner was failing to use the simple past tense where s/he had to, a gesture to show

behind was used which arbitrarily in that class meant that the past tense had to be used. Or moving both hands smoothly across the chest would mean that a progressive tense should have been used.

The grammatical features targeted in this study were chosen based on the ones introduced in the first half of *Interchange Three*. In order to make all students speak and produce language and to give equal opportunity to all subjects to have output in English during the treatment, interactions were in the person-to-person form; that is to say the teacher had interaction with all the students in each session. As a result, the participants were given equal chance to have output and receive feedback on their production accordingly.

At the end of the instruction period, the subjects were interviewed once again in a posttest (exactly with the same procedure explained above) to track any possible improvement in their speaking grammatical accuracy with respect to both the kind of correction they received throughout the treatment period and the way male and female participants reacted to the two types of correction they received.

Results

Before and after the treatment, certain pertinent statistical analyses were conducted to both guarantee maximal accuracy of the procedure and also check the value of the hypotheses.

Piloting the Placement Test

The first step was to pilot the test which was to be used to make sure that all groups were equal and belonged to the same population in terms of their general English proficiency. Therefore, the objective sections of the test comprising listening and reading each with 20 items and language use with 30 items were administered to 70 subjects with the same qualities as those of the main study, and then NRT item analysis including item facility and item discrimination was conducted for each item. After omitting the malfunctioning items, the reliability of the test was estimated using the KR-21 formula and it came out to be satisfactory with an index of 0.89 (Table 1).

Table 1 – The reliability of the objective sections of the placement test

Cronbach's Alpha	N of Items
0.89	68

Next, the inter-rater reliability of the speaking section was computed for 25 out of the 70 subjects who took part in the piloting procedure. Table 2 bears the results.

Table 2 – Inter-rater reliability for the speaking test

		S rater A	S rater B
S rater A	Pearson Correlation	1	.874**
	Sig. (2-tailed)		.000
	N	25	25
S rater B	Pearson Correlation	.874**	1
	Sig. (2-tailed)	.000	
	N	25	25
**. Correlation is significant at the 0.01 level (2-tailed).			

Evidently, the correlation between the two raters was 0.874 meaning that the average score of their marking could be safely used as the speaking score of every individual for the speaking pretest and posttest. The same two raters were also checked for their inter-rater reliability on writing. Table 3 shows that the correlation coefficient of 0.88 was significant with less than 1% error (ρ value of 0.000 being far less than 0.01) allowing the researchers to use them for marking the writing papers as well.

Table 3 – Inter-rater reliability for the writing test

		W rater A	W rater B
W rater A	Pearson Correlation	1	.880**
	Sig. (2-tailed)		.000
	N	25	25
W rater B	Pearson Correlation	.880**	1
	Sig. (2-tailed)	.000	
	N	25	25
**. Correlation is significant at the 0.01 level (2-tailed).			

Administering the Placement Test

Following the piloting phase and conducting inter-rater reliability measures, the placement test was administered to the four target groups of the study, the descriptive statistics of which are presented in Table 4. It is worth mentioning that the score of the test in its original form (being the sum of the writing, speaking, and multiple-choice sections) adds up to a maximum of 90, with the assumption that a learner with a score between 37 and 42 can be considered to be at the intermediate level.

Table 4 – Descriptive statistics of the placement test

	N	Range	Min	Max	Mean	SD	Var	Skewness		
	Stat	Stat	Stat	Stat	Stat	Std. Error	Stat	Stat	Std. Error	
GP female Cont	18	20	38	58	45.78	1.19	5.06	25.59	.65	.54
GP female Exp	19	20	37	57	46.26	1.11	4.83	23.31	.4	.52
GP male Cont	22	19	37	56	44.86	1.06	4.99	24.88	.68	.49
GP male Exp	21	17	38	55	46.67	1.10	5.04	25.43	-.09	.50
Valid N listwise	80									

The results were subjected to an ANOVA to ascertain the equality of the four groups in terms of their general proficiency. Table 5 presents the results.

Table 5 – One-way ANOVA on the results of the placement test for the four groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	35.578	3	11.859	.971	.411
Within Groups	928.372	76	12.215		
Total	963.950	79			

The insignificant value of 0.411 being greater than 0.05 in the table above shows that the four groups were at the same level of language proficiency at

the beginning of the study and belonged to the same population in this respect.

Speaking Pretest

The next step was to analyze the speaking section of the above mentioned test in isolation in order to make sure that the participants of the four groups were at same level in terms of their speaking skill as well. Table 6 demonstrates the descriptive statistics of the speaking section of the test.

Table 6 – Descriptive statistics of the speaking pretest

	N	Range	Min	Max	Mean		SD	Var	Skewness	
	Stat	Stat	Stat	Stat	Stat	Std. Error	Stat	Stat	Stat	Std. Error
S Female cont.	18	6	6	12	8.22	.40	1.70	2.89	.74	.54
S Female Exp.	19	6	6	12	7.89	.37	1.59	2.54	1.48	.52
S Male Cont.	21	6	6	12	7.62	.32	1.46	2.15	1.79	.50
S Male Exp.	21	6	6	12	8.24	.38	1.73	2.99	.88	.50
Valid N (listwise)	18									

As it can be seen in Table 7, the p value of 0.577 being greater than 0.05 indicates that the four groups did not exhibit any significant differences in their speaking and hence were eligible to participate in the study as samples of the same population.

Table 7 – One-way ANOVA of the results of the speaking pretest for the four groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.186	3	1.729	.664	.577
Within Groups	197.801	76	2.603		
Total	202.987	79			

Speaking Posttest

Following the instruction, all groups were interviewed again and their average scores given by the two raters were analyzed. Table 8 summarizes the descriptive statistics of the speaking posttest of the four groups separately.

Table 8 – Descriptive statistics of the speaking posttest of the four groups

	N	Range	Min	Max	Mean		SD	Var
	Stat	Stat	Stat	Stat	Stat	Std. Error	Stat	Stat
S Female Cont.	18	3	9	12	10.72	.240	1.018	1.036
S Female Exp	19	4	8	12	11.32	.230	1.003	1.006
S Male Cont.	22	4	7	11	9.09	.271	1.269	1.610
S Male Exp.	21	3	9	12	10.14	.232	1.062	1.129
Valid N listwise	80							

Table 9 below displays the descriptive statistics of the experimental and control group without taking the gender variable into consideration. Based on the information depicted in the table, the experimental group gained a mean of 10.70 on the whole while the control group's mean was 9.82.

Table 9 – Descriptive statistics of the speaking posttest of the experimental and control groups

	G	N	Mean	SD	Std. Error Mean
Pt	Exp.	40	10.70	1.181	.187
	Cont.	40	9.82	1.412	.223

As it is revealed in Table 10, female participants in the experimental group achieved a mean of 11.32 with the male experimental group achieving a mean of 10.14.

Table 10 – Descriptive statistics of the speaking posttest of the female and male experimental groups

	Experimental Groups	N	Mean	Std. Deviation	Std. Error Mean
Pmfp	Female Group	19	11.32	1.003	.230
	Male Group	21	10.14	1.062	.232

To simultaneously address hypothesis 1, which stated there was no significant difference between the speaking grammatical accuracy of EFL learners exposed to nonverbal prompts and that of those exposed to recasts., and hypothesis 2, which stated that there was no significant difference between the speaking grammatical accuracy of female and male participants who received nonverbal prompts, the researchers needed to run a two-way ANOVA to test the impact of the independent and moderator variables, i.e. nonverbal prompts and gender, on the dependent variable, i.e. the speaking grammatical accuracy of EFL learners.

To legitimize running a two-way ANOVA, the Levene’s test of homogeneity of variance was carried out. Table 11 demonstrates the results.

Table 11 - Levene’s test of homogeneity of variance on speaking posttest

	F	Sig.
Experimental and control groups	.935	.337
Female and male experimental groups	.462	.501

The results of testing the homogeneity of variances for the experimental and control groups ($\rho = 0.337$) and for the male and female experimental groups ($\rho = 0.501$) revealed that the distribution of scores enjoyed homogeneity of variance since all ρ values were higher than 0.05. Therefore, running the two-way ANOVA was legitimate. The ANOVA was 2*2 and was carried out on the speaking posttest scores of the four groups to simultaneously address the two hypotheses. This type of ANOVA could enable the researchers to investigate the impact of the independent and the moderator variables on the dependent variable as well as their possible interactions. Table 12 presents the results:

Table 12 – Two-way ANOVA of the speaking posttest of the four groups

Tests of Between-Subjects Effects					
Dependent Variable: SP1234					
Source	Type 3 Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	52.583 ^a	3	17.528	14.036	.000
Intercept	8450.601	1	8450.601	6.767	.000
GroupG	35.601	1	35.601	28.510	.000
GroupC	14.170	1	14.170	11.347	.001
GroupG * GroupC	1.670	1	1.670	1.337	.251
Error	94.904	76	1.249		
Total	8573.000	80			
Corrected Total	147.487	79			
a. R Squared = .357 (Adjusted R Squared = .331)					

To interpret the above findings, one has to pay attention to the last column where significant levels of the variable gender (i.e. Group G), the variable correction type (i.e. Group C), and finally the interaction of the two variables (i.e. Group G * Group C) can be found. The mean difference between the experimental and control groups reported in Table 9 above (mean of 10.70 for the experimental group and mean of 9.82 for the control group) came out to be significant according to Table 12 ($F_{(1,76)} = 11.347, \rho = 0.001 < 0.05$) indicating that the experimental group significantly outperformed the control group. Therefore, nonverbal prompts were more significantly effective than recasts with respect to speaking grammatical accuracy.

Moreover, the mean difference between females (11.32) and males (10.14) as demonstrated in Table 10 before, came out to be significant based on the results depicted in Table 12 ($F_{(1,76)} = 28.510, \rho = 0.000 < 0.05$) showing that females significantly outperformed males in the experimental group. Yet, the results of the two-way ANOVA demonstrated that the interaction of the two variables (gender and correction type) was ineffective on how learners improved their speaking grammatical accuracy ($F_{(1,76)} = 1.337, \rho = 0.251 > 0.05$).

As the next step in the data analysis, a regression analysis was carried out. In the regression analysis, the magnitude of Beta coefficient indicates the degree of the impact of the independent variable on the dependent. In this study as mentioned earlier, the main independent variable was the type of correction feedback and the dependent variable was the speaking grammatical accuracy of the participants. Hence by looking at the Beta coefficient in the Tables 13 and 14, one can compare the degree of improvement of male and female participants.

Table 13 – Regression between the speaking posttest scores of the female experimental and control groups

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10.722	.238		45.032	.000
	Pos Fec group	.594	.332	.289	1.786	.083
a. Dependent Variable: F postec						

Table 14 – Regression between the speaking posttest scores of the male experimental and control groups

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.091	.250		36.359	.000
	Pos Mec group	1.052	.358	.417	2.940	.005
a. Dependent Variable: mpostec						

As demonstrated in Tables 13 and 14, since the Beta coefficient of female groups (0.289) is less than that of the male groups (0.417), it can be concluded that the males benefited more than the females from the employment of nonverbal prompts when considering the participants' degree of achievement. This means that although the female experimental group outperformed the male experimental group on the posttest prompts were more efficient in reducing the speaking grammatical inaccuracy of the male participants. Hence, males on the whole benefited more from nonverbal prompts than females.

Conclusion

In this study, nonverbal prompts did significantly improve the speaking grammatical accuracy of the two experimental groups of male and female subjects. Among the studies that suggest a superior benefit for prompts that generates pushed output, the above finding confirms those of Lyster's (2004) study of eight fifth grade French immersion classes in which teachers supplied recasts, prompts (i.e. CF which withholds the correct target language form but prompts learners to attempt self-repair), or no feedback in response to errors of grammatical gender. Again in another study, Lyster and Izquierdo (2009) investigated the differential effects of prompts and recasts, in the context of dyadic interaction, on the acquisition of grammatical gender by adult second language learners of French. They concluded that learners receiving recasts benefited from the repeated exposure to positive exemplars, as well as from opportunities to infer negative evidence, whereas learners receiving prompts benefited from the repeated exposure to negative evidence as well as from opportunities to produce modified output.

However, beyond facilitating the noticing of difficult features, it has also been suggested that certain types of CF may also function to promote different degrees of L2 processing (Panova & Lyster, 2002). For instance, CF which contains positive evidence about the target language (e.g. recasts) can be useful in the internalization of new forms (Panova & Lyster, 2002) and can allow learners to *notice the gap* (Schmidt, 2001) through comparison of mismatches between target language norms and current interlanguage knowledge. In contrast, CF which does not contain a full reformulation but instead contains either negative evidence (e.g. clarification requests, elicitation) or hints and strategies for approaching the target-like reformulation (e.g. metalinguistic cues) requires that learners attempt self-repair or output modification. This opportunity to modify output can be useful in promoting the learners' awareness of what they are as yet unable to say in the target language or *noticing the whole* (Swain, 1995). Furthermore, this opportunity to attempt self-repair may enhance control over internalized forms by requiring deeper processing of already acquired L2 knowledge during output modification (Panova & Lyster, 2002). As a result, this deeper processing may "establish a more durable memory trace" (Izumi, 2002, p. 570), which can be called upon more readily in subsequent L2 production and lead to rule strengthening or more automatic processing.

To investigate the degree of the impact of prompts on male and female learners, the researchers ran two regression analyses which surprisingly revealed that male learners benefited more from the experimental condition than the female learners. The findings of the regression analyses were contrary to the rejection of the second hypothesis through which it was concluded that the female experimental group had a significantly higher speaking mean than the male experimental group. The literature also supports the notion that women do better than men in language learning. A comprehensive list of such studies can be found in Ehrlich (1997), who alongside providing evidence for female superiority also critiques a number of the articles by pointing out a methodological problem in them: many studies done in support of the female superiority hypotheses involved L1 learners. Berk (2007) also mentions that females are generally better language learners than males while males generally perform better in mathematics.

As for those studies that looked specifically at L2 learning, Ehrlich (1997) argues that female superiority may be attributable to a better use of learning strategies than to greater giftedness in second language learning. This argument is also proposed by Oxford (1990) who bases her conclusion on a number of different studies she conducted on L2 learning strategies. Females seem to consciously use strategies more often than males. However, it may be the qualitative differences in their strategy use that favor females who

often show better classroom performance in a second or foreign language than males (Oxford, 1990). All of these works show that female students are more concerned than their male counterparts with rule following, adherence to guidelines, and pleasing the teacher. Research suggests (Gordon, 1997) females are benefiting more than their male classmates from school attempts to boost their performance.

Finally, through the univariate two-way ANOVA, the researchers could compare the effects of prompts and recasts with and without the possible influence of gender on the speaking accuracy of the participants. The findings uncovered that the variables of correction type and gender each had an impact on the speaking accuracy of the learners, yet the interaction of the two variables was ineffective on the way learners improved in their speaking. If each gender type benefited more from one type of correction, then the interaction of the two variables would have become significant. The females unlike males improved equally under the two types of feedback, hence the reason behind the unproductive combination of gender and correction type.

Both learners and teachers agree that corrective feedback is an inevitable part of instruction; nonetheless, the most efficient type of feedback remains to be a matter of controversy. Prompts as an alternative form of feedback can occur in communicatively meaningful interactions and thus have greater pedagogical impact on learning. It is argued that in interaction, prompts are more likely perceived as correction and thus likely to be effective as those provided through other types of feedback. As Lyster (2004) and McDonough (2005) similarly suggest, the opportunity to produce pushed output in response to feedback benefit learners more. In fact, the opportunity to self-repair induced by prompts is the key factor in enhancing noticing and learning linguistic forms. Using the non-pedantic and problem-solving nature of prompts, teachers can completely alternate the embarrassing feeling of being corrected to an experience which is puzzle-like, interesting, and fun for learners. This requires language material developers to include sections in the teacher's guide which introduce different types of prompts and give them a brief training as to when and how they should be used. Though difficult, this job, if done, can hugely contribute to the grammatical accuracy of EFL learners.

The Authors

Arshya Keyvanfar is Assistant Professor of TEFL at Islamic Azad University, North Tehran Branch and has been teaching at undergraduate and graduate levels in the areas of language teaching methodology, testing, and research. She has published

several papers in academic journals and has also co-authored a book on teaching methodology. Dr. Keyvanfar is particularly interested in task-based language teaching and its application in EFL settings.

Fatemeh Azimi holds an MA degree in TEFL from Islamic Azad University, North Tehran Branch. She is currently an English teacher in a number of language schools in Gillan.

References

- Ammar, A. (2003). *Corrective feedback and L2 learning: Elicitation and recasts*. Unpublished doctoral thesis, McGill University, Montreal.
- Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts and L2 learning. *Studies in Second Language Acquisition*, 28, 543-57.
- Berk, L. (2007). *Development through the lifespan*. Massachusetts: Allyn & Bacon.
- Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, 15, 357-386.
- Ehrlich, S. (1997). Gender as social practice. *Studies in Second Language Acquisition*, 19, 421-46.
- Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28(2), 339-368.
- Gordon, E. (1997). Sex, speech, and stereotypes: Why women use prestige speech forms more than men. *Language in Society*, 26, 47-63.
- Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36(4), 543-572.
- Havranek, G. (1999). The effectiveness of corrective feedback: Preliminary results of empirical study. *Acquisition et Interaction en Langue Etrangere*, 2, 189-206.
- Havranek, G., & Cesnik, H. (2001). Factors affecting the success of corrective feedback. In S. H. Foster Cohen & A. Nizegorodcew (Eds.), *EUROSLA Yearbook* (pp. 99-122). Amsterdam, Philadelphia: John Benjamins.
- Iwashita, N. (2003). Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition*, 25, 1-36.
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, 24, 541-577.
- Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, 82(3), 357-371.
- Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, 26, 399-426.

- Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam, Philadelphia: John Benjamins.
- Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, 59(2), 888-905.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 137-66.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, 82(3), 338-356.
- McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses to ESL question development. *Studies in Second Language Acquisition*, 27, 79-103.
- Nicolas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language Learners. *Language Learning*, 51(4), 719-758.
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. New York: Newbury House Publishers.
- Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, 36(4), 573-595.
- Philp, J. (2003). Constraints on "noticing the gap": Nonnative speakers' noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, 25, 991-1026.
- Richards, J. C. (2005). *New interchange, passages placement*. New York: Cambridge University Press.
- Sauro, S. (2007). *A comparative study of recasts and metalinguistic feedback through computer mediated communication on the development of L2 knowledge and production accuracy*. Unpublished doctoral dissertation, University of Pennsylvania.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

