



Content list available at www.urmia.ac.ir/ijltr

*Iranian Journal
of
Language Teaching Research*



Urmia University

The impact of pushed output on accuracy and fluency of Iranian EFL learners' speaking

Aram Reza Sadeghi Beniss^a, Vahid Edalati Bazzaz^{a,*}

^a *Semnan University, Iran*

ABSTRACT

The current study attempted to establish baseline quantitative data on the impacts of pushed output on two components of speaking (i.e., accuracy and fluency). To achieve this purpose, 30 female EFL learners were selected from a whole population pool of 50 based on the standard test of IELTS interview and were randomly assigned into an experimental group and a control group. The participants in the experimental group received pushed output treatment while the students in the control group received non-pushed output instruction. The data were collected through IELTS interview and then the interview of each participant was separately tape-recorded and later transcribed and coded to measure accuracy and fluency. Then, the independent samples t-test was employed to analyze the collected data. The results revealed that the experimental group outperformed the control group in accuracy. In contrast, findings substantiated that pushed output had no impact on fluency. The positive impact of pushed output demonstrated in this study is consistent with the hypothesized function of Swain's (1985) pushed output. The results can provide some useful insights into syllabus design and English language teaching.

Keywords: pushed output; accuracy; fluency; EFL speaking

© Urmia University Press

ARTICLE SUMMARY

Received: 28 Jan. 2013

Revised version received: 23 Apr. 2014

Accepted: 27 May 2014

Available online: 1 July 2014

* Corresponding author: Semnan University, Iran
Email address: edalati_vahid@yahoo.com

Introduction

To date, all researchers in the field of applied linguistics have been encouraged to investigate how second language learners can produce accurate and fluent target language. The demand for accuracy and fluency in English has been sharply increasing because of strong situation of English as a language for international communication. The ability to speak English fluently and accurately opens up wider opportunities to achieve success in life. So, scholars and researchers began to recognize the role of Swain's (1985) Pushed Output (PO) as an essential factor for the enhancement of accuracy and fluency.

The notion of pushed output is grounded in Swain's data collection from a Canadian French immersion program. Swain (1985, 1995) mentioned that immersion program in Canada proved that comprehensible input alone was insufficient to ensure that learners achieve accuracy. Her observation showed that immersion learners did not gain the ability for accurate production while they were fluent. Swain (1985) concluded that the lack of grammatical accuracy of immersion learners could be attributable to the restricted chances to produce output or for being pushed to produce output.

Regarding what contributing elements may help language learners enhance their accuracy and fluency, and due to scarcity of empirical studies that support or rebut Pushed Output Hypothesis (POH) (Swain, 1985), specifically, this study aims at examining the impact of PO on accuracy and fluency of Iranian EFL learners. The results of the study might propose positive impact of PO and lead us to conclude that English teachers can apply PO as an alternative to develop accuracy and fluency of Iranian EFL learners.

Literature review

Pushed Output Hypothesis

As explained earlier, Swain (1985) devalued input as an important role for language acquisition. Her observation on immersion program revealed that production was necessary for acquisition. Thus, Swain (1985) proposed the concept of Pushed Output. What is meant by the concept of PO is that learners are "pushed" or "stretched" in their production as a necessary part of making themselves understood. Ellis (2003) defines PO as "output that reflects what learners can produce when they are pushed to use target language accurately and concisely" (p.349). Swain (1985) asserts that when the students are pushed to engage in production, they have the chance to deliver the messages which are precise, coherent and appropriate. Moreover, Swain (1995, 2005) claimed that the production of output, notably PO, could enhance fluency and accuracy. Mackey (2012) defines output as "process of rephrasing or reformulating one's original utterance in response to feedback or self-monitoring" (p.16). She further explains that when learners are exposed to feedback, they try to reformulate their utterances and produce more accurate, appropriate, complex and comprehensible target language.

Swain (1998) argues that output forces learners to move from semantic analysis of the target language to a more syntactic analysis of it. In support of this position, Nation and Newton (2009) claim that “Comprehension processes involve semantic decoding and production involves syntactic processing” (p.115). Achard and Niemeier (2004) suggest that PO can develop learners’ consciousness to notice the gap between what they mean to say and what they are able to say. So “pushed output may assist the learners in acquiring” (Lee, 2002, p.276). There seems to be consensus among scholars concerning that a language learner can obtain comprehensible output by pushing his production in a language interaction. As a result, drawing on this evidence PO can occur in second language (L2) conversation or classroom interaction (Lynch, 1997; Pica, Lincoln-Porter, Paninos, & Linnell, 1996; Shehadeh, 2001; Swain, 1995).

After proposing Output Hypothesis, Swain refined output hypothesis and suggested different functions of POH (Swain 1993, 1995, 1998, 2005). Noticing function as the first function posits that L2 learners consciously understand their linguistic problems through PO activities. As the learners notice the gap between their own target output and L2, they may become more attuned to the related structure in the target language. Thus, specific grammatical forms may become more salient and create a context for L2 development (Mackey, 2012). The second function (i.e., hypothesis testing) claims that language acquisition is developed when L2 learners consciously use target form and reformulate it upon receiving feedback from interlocutors. Third, metalinguistic function highlights the role of PO and defines metalinguistic as “using language to reflect on language produced by others or the self, mediates second language learning” (Swain, 1985, p.478). Fourth, fluency function states that PO is important in developing learners’ fluency or automatization of language use. And the final function of PO posits that when language learners are pushed to produce output, they are forced to move from “semantic processing” predominant in comprehension to “syntactic processing” (Swain, 1985, p. 249).

Pushed Output Tasks

In the light of POH, considerable research effort was expended on the study of role of PO tasks in L2 acquisition (e.g., Colina & Garcia Mayo, 2007; Izumi, 2002; Izumi & Bigelow, 2000; Izumi, Bigelow, Fujiwara, & Fearnow, 1999; Izumi & Izumi, 2004; Mennim, 2007; Nassaji & Tian, 2010; Qin, 2008; Reinders, 2009; Song & Suh, 2008; Swain & Lapkin, 1995; Yoshimura, 2006).

In designing pushed output tasks, Nation and Newton (2009) claim that PO tasks must consider a wide range of unfamiliar topics because talking about unfamiliar topics pushes the learner to produce target language. They also claim that learners are pushed when they are asked to produce the target language without opportunity for planning or preparation. In support of this position, Nation (2011) posits that “pushed output occurs when the learners have to produce spoken language in tasks that they are not completely familiar with” (p. 445). Therefore, it is the role of the teacher to design speaking tasks with unfamiliar topics that an opportunity is created for the learners to be pushed. To cite Nation (2011) once again, he maintains that pushed output task must consider various different text types. Applying different texts may help create not only a wide range of grammatical features and vocabularies but also informal and formal conditions for interaction.

Two types of tasks which are linked with many studies of PO are one-way and two-way tasks. Mackey (2012) defines one-way tasks as non-reciprocal tasks in which the learner does most of the talking and is responsible to transmit the information to successfully complete the task. In contrast, in two-way tasks, both participants exchange the information. Shehadeh (1999) maintains that evidence proves that it is one-way tasks which provide a situation for the students to generate PO rather than two-way tasks. In one way tasks “the burden of completing the task successfully is placed on the participants who hold the information although other participants can contribute by demonstrating when they comprehend and when they do not” (Ellis, 2003, p.88). Izumi and Izumi (2004) found that picture description task was used to elicit oral production and substantiated that picture description was the best task to provide a situation for students to be pushed in oral output. Also, Shehadeh (1999) simply found and revealed that picture description tasks gave more chances for using pushed output than opinion exchange. This result was generally corresponded to that in Iwashita (1999). She indicated that one-way tasks could create more opportunities to produce modified output than two-way tasks. Pica et al. (1996) concluded that the storytelling tasks utilized in their research provided non-native speakers with higher percentages of production of PO in both native speaker and non-native speaker interactions than the house sequence task.

The issue of feedback has received substantial attention in POH. A plethora of studies confirmed that learners were pushed to reformulate or modify their output when they were exposed to clarification requests or confirmation checks during negotiation (Iwashita, 1993; Lyster & Ranta, 1997; Mitchell, Myles, & Marsden, 2013; Pica, Holliday, Lewis, & Morgenthaler, 1989). These sorts of feedback can help the speakers focus on linguistic problems and will also lead them to notice the gaps between features of their interlanguage and the target language (Mackey, 2012). Gass and Mackey (2007) believe that this modification and reflection in response to clarification request can contribute to SLA. According to Takashima and Ellis (1999), whenever a learner generated an utterance that was not understood, the listener might answer with a clarification request, which caused the learner to subsequently modify the problematic utterance, as in this example:

Student: Cinderella change into the beautiful girl.

Teacher: Sorry?

Student: Cinderella changed into a beautiful girl (Takashima & Ellis, 1999, p.173).

Two Components of Speaking

As it was mentioned above, all scholars measured one or two features of speaking proficiency among many other aspects of speaking proficiency such as accuracy, fluency, vocabulary, etc. Since 1990s, these two concepts (i.e., accuracy and fluency) along with complexity have been used predominantly and prominently as dependent variables to assess variation with respect to independent variables such as acquisitional levels or task features (e.g., Hilton, 2008; Guillot, 1999; Larsen-Freeman, 2006; Lennon, 2000; Tonkyn, 2007; Towell & Dewaele, 2005). According to Ellis and Barkhuizen (2005), there are two approaches for assessing fluency and accuracy: 1) Obtaining ratings by providing general descriptors of different levels of accuracy and fluency. 2) Calculating various discourse based measures. Due to the practical way of measuring learners' language and high face validity, language testers employ the first approach. The first approach is based on how

language learners are able to communicate. In contrast, SLA researchers tend to use the second approach in order to obtain precise measures of the different aspects (Ellis & Barkhuizen, 2005).

Accuracy

Accuracy is a simple concept to be identified because all researchers agree with the main goal of accuracy. Skehan (1996) defines accuracy as the production of target language according to its rule systems. Housen and Kuikken (2009) regard accuracy as “error-free” speech. Researchers distinguished between two types of measures of grammatical accuracy: specific measures of accuracy and general measures of accuracy (Ellis & Barkhuizen, 2005). As specific measures of accuracy (Crookes, 1989; Kawauchi, 2005; Wigglesworth, 1997), researchers attempt to measure certain forms on learners’ proficiency levels and development. As an example, measuring target-like verbal morphology and target-like verbal use of plural are specific measures of grammatical accuracy. In contrast to specific measure of accuracy, general accuracy is a more realistic and sensitive measure (Skehan & Foster, 1999) and captures more general changes in accuracy (Skehan & Foster, 1997). In general measures of accuracy, accuracy can be measured through percentage of error-free speech (Foster & Skehan, 1996), error-free T-units (Ortega, 1999; Robinson, 1995), error-free AS-units (Lambert & Engler, 2007), the number of errors per 100 words (Kuiken & Vedder, 2007; Wolfe-Quintero, Inagaki, & Kim, 1998) number of errors per T-unit (Bygate, 2001), errors per one hundred words (Mehnert, 1998); percentage of target-like use of plurals (Crooks, 1989); and target-like use of vocabulary (Skehan & Foster, 1997).

Fluency

Even though the concept fluency is constantly applied within the field of applied linguistics, there is no global agreement about what is perceived as fluency (Chambers, 1997). Ellis and Barkhuizen (2005) defined fluency as “the production of language in real time without undue pausing or hesitation” (p. 139). The oral fluency is frequently assessed through proficiency tests (e.g. IELTS, TOEFL iBT, etc.). However, investigators have tried to unravel the distinctive elements that develop fluency rating.

Research into fluency within the field of SLA has followed two general categories. First, studies which focused on the measurement of quantifiable aspects. Second, studies which were expended on elements that impact on evaluation of learners’ fluency by the listeners. Lennon (1990) proposes that fluency can be measured based on two main features: First, temporal measures which consider speed of delivery and second, hesitation markers that relate to disfluencies like repetitions and false starts.

Research Questions and Hypotheses

The current study attempted to investigate the following research questions:

1. Does pushed output have any significant impact on Iranian EFL learners' accuracy?
2. Does pushed output have any significant impact on Iranian EFL learners' fluency?

To answer the above questions, the following null hypotheses were formulated:

H01: Pushed output does not have any significant impact on Iranian EFL learners' accuracy.

H02: Pushed output does not have any significant impact on Iranian EFL learners' fluency.

Methodology

Participants

This study was conducted with 30 upper-intermediate female English language learners (aged 18-22) studying at Jihad Institute in Mashhad. The participants happened to have already passed 14 courses in general English which lent credence to their being at an upper-intermediate level. They were all Iranian nationals sharing the same language background. None of the participants had lived or stayed in an L1 English environment and they had virtually no opportunity to use English for communicative purposes outside the classroom context. The only contact they had with English outside the classroom was at school or university.

The selection of participants was done using IELTS speaking administered by an IELTS examiner before the treatment. Those who obtained a score of below 4.5 in speaking were eliminated from the participant pool, while those who scored 4.5 to 6 were selected. A total of 30 out of 50 students met the requirement and were hence remained as the participants in the experiment. In order to ensure appropriate sample selection, the remaining participants were randomly assigned to two groups (15 participants in each group). Deviations from normality were identified in this study using Kolmogorov-Smirnov test. The mean rank of experimental group was 15.30, whereas that of the control group was 15.70. Since the data in participants' IELTS score was not normal, the result of Mann-Whitney test, the non-parametric equivalent to the independent samples t-test, with $z = -0.130$, $p = .896$, indicated no statistically significant difference among IELTS speaking test scores of the two classes. Thus, the test confirmed that the two groups were at the same level of speaking proficiency.

Materials

For the purpose of the current study, a description of the treatment and testing materials is provided in order.

Treatment Materials

For PO group, the participants had to take part in activities which pushed them to generate the target language. Four types of communication tasks based on students' level were selected and designed for the purpose of PO: picture description, retelling, ask and answer task and storytelling. The selection of these four types of tasks was motivated by previous investigations (Izumi & Izumi, 2004; Nation & Newton, 2009; Pica et al, 1989; Shehadeh, 2001). In the picture description, the learner was given a set of pictures and was asked to apply his creativity to create a story from those pictures and describe it to her partner in order to enable the interlocutor to reproduce the pictures (see, e.g., Mackey, 1999; Mackey & Philip 1998; Swain & Lapkin 2000). In retelling task, the learner spent one minute reading the passage silently from the book "Steps to Understanding" written by Hill (1988) and then closed the storybook to start retelling it confidently to her partner or the teacher. Or learners listened to a recording from "Steps to Understanding" and noted down key words to retell the content of the recording in pairs. Nation and Newton (2009) believe that one of the most effective ways to bring receptive language knowledge into productive use is through retelling. A retelling task pushes the learner to make sense of items from the input and to use them productively. In ask and answer, the students practiced in pairs. One student had a text to read and another student had some questions based on the passage. The one who had the questions started asking questions from another student. This activity took a few minutes and eventually performance was done in front of the class without looking at the text. Finally, Lynch (1997) in his paper considered story telling as a PO task. In this study learners were asked to retell the story after watching a short silent movie. One variable manipulated in the treatment condition was PO in the form of clarification requests (e.g., pardon? what? huh?). In contrast, for the control group, the participants had to be involved in activities which did not push them to produce target language. Unlike PO, there are wide varieties of activities in the field of L2 – such as picture sequencing, listening to news and so on – that do not push the learner to speak (see, e.g., Izumi & Izumi, 2004; Mackey, 2012).

Testing material

Currently, the best known instrument for measuring speaking skill is speaking proficiency interview based on IELTS speaking assessment descriptors (public version). This scale ranges from 1 to 9 with 1 representing no proficiency and 9 representing the speaking proficiency of an educated native speaker. Since the main focus of this study is accuracy and fluency, the transcription of speaking module of IELTS test was employed to measure learners' accuracy and fluency before and after the implementation of the instructional program to find the impact of PO on participants' accuracy and fluency. The interviewers rated the participants based on the IELTS speaking assessment descriptors (public version). Basically, the IELTS speaking test takes between 11 and 14 minutes (Jakeman & McDowell, 2008). The test for both pretest and posttest which was used in this study was a sample IELTS speaking test from the books "Cambridge IELTS 5, 6,7,8,9 self-pack" written by Cambridge ESOL (2006, 2007, 2009, 2011, 2013). So, participants were not exposed to the same topics in two interviews so as to minimize any potential practice effect.

Procedure

The study was carried out in the spring semester of 2013 at Jihad Language Institute in Mashhad, Iran for twelve 30-minute sessions through 4-week semester. As noted earlier, before the researcher performed the treatment all candidates were given the IELTS interview as a pretest in order to select participants for the study and assess their current level of knowledge before treatment.

The interview was conducted by the head of IELTS department at Jahan-Elm higher education, Iran. The interviewer participated in the IELTS examination on September 4, 2009 and obtained the overall score of 8 in academic module with the speaking band score of 9. To have inter-rater reliability, two other IELTS interviewers listened to the audio-taped interviews and agreed with the scores given by the first rater.

A week later, after the researcher assigned the participants into experimental and control groups, the treatment started. Two groups received two different sorts of tasks. Participants in PO group received the practices in focus of pushed output activities. They were engaged in picture description tasks, retelling, ask and answer tasks and storytelling on the first, second and the third session. The other treatment sessions were exactly the same as the first three sessions except the topic of the materials. In contrast, the control group received non-pushed output activities in twelve sessions. During PO task peers or the instructor were instructed to ask for clarification requests when a learner produced an inaccurate utterance.

One week after the twelfth treatment session, the posttest was administered in order to determine any variation in participants' accuracy and fluency after treatment. The procedure for posttest was exactly the same as pretest. Both pretest and posttest audio-recordings of all 60 interviews in this study were transcribed by independent transcribers and the transcripts were reviewed by the researcher. The transcriptions were coded for scoring and statistical analysis of fluency and accuracy to find out the impact of treatment. In the process of coding, the researcher coded all the data. To assess inter-rater reliability for the coding, a second rater coded approximately 30 % of the interview data achieving inter-rater agreement at 95 %.

It must be noted that a Samsung mp3 recorder/player and a Samsung mobile phone (S3) were applied for recording the learners' interviews. Both gadgets were used at the same time to avoid the probable loss of data.

Bear in mind that both groups were conducted by the same instructor who was also the researcher. All efforts were made to maintain both experimental and control groups comparable with regard to time on task and amount of exercises.

Measure

AS-unit for Measuring Spoken Language

The analysis of the learners' language in terms of accuracy and fluency requires a principled way of segmenting text into units (Ellis & Barkhuizen, 2005). Foster, Tonkyn and Wigglesworth (2000)

did a broad and critical survey on many units of analysis used in the field of applied linguistics and SLA over the past few decades for segmenting written or oral language and finally proposed AS-unit as an improvement to other units of analysis. They defined AS-unit as "... a single speaker's utterance consisting of an *independent clause*, or *sub-clausal unit*, together with any *subordinate clause(s)* associated with either" (p. 365). A plethora of studies have maintained that the AS-unit is particularly suitable for analyzing oral data (Foster et al., 2000; Norris & Ortega, 2009; Plough, Briggs, & Van Bonn, 2010).

Accuracy Measure

In this study, the second focus was on the learner's ability to produce accurate target language. As such, accuracy was measured as the percentage of error-free AS-unit (Foster et al., 2000; Lambert & Engler, 2007). The number of error-free AS-units produced by each participant during the interview was divided by the total number of AS-units the participant produced to arrive at the accuracy scores. To identify errors in participants' transcripts, the modified version of Ferris' (2002) written error types was used in this study (Swain, Brooks, Yoon, & Al-Alawi, 2012). Ellis and Barhuizen (2005) suggest that it is better to consider self-corrected unit as error free. Thus, whenever a learner made an accurate self-correction, that unit was coded as error-free.

Fluency Measure

In this study, fluency was achieved by calculating the ratio of dysfluency markers (pauses and repetitions) to AS-units. The number of AS-units produced by each learner during the interview was divided by the number of dysfluencies that the learner generated to create an inverse relationship in the fluency ratios. In this way, the ratio increased as number of dysfluencies decreased and vice versa (Lambert & Engler, 2007). Dysfluency pauses were those in which the learner seemed to be looking for language or which otherwise seemed due to a deficiency in target language skill (Lambert & Engler, 2007). As will be recalled, pausing at AS boundaries, pausing for rhetorical effects or pausing for setting off planned chunks were not included as dysfluency markers. Dysfluency repetitions were those repetitions in which the learners repeated previously produced speech. Repetitions which were used for rhetorical effects within spoken discourse, such as when the speaker repeated for emphasis, comprehensibility, in response to a clarification request, or interruption were not calculated as dysfluency.

Data Analysis

In order to obtain quantitative data needed for analysis, the interviews of each participant both in pretest and posttest were transcribed and coded for statistical analysis. Statistical Package for Social Sciences (SPSS, version 20) was used to analyze the data. As the data were normal for accuracy, first, an independent samples t-test was employed to show whether two groups were homogenous in terms of accuracy before treatment. Then, an independent samples t-test was run again for posttest of control and experimental groups to see any variation after treatment. Since the data for fluency were normal too, the same statistical tests were applied for fluency to answer the second research question.

Results

The Result of Pretest and Posttest on Accuracy

In order to establish the homogeneity of both control and experimental groups in terms of accuracy, an independent samples t-test was performed to compare pretest gain scores of the experimental and control groups in terms of accuracy.

Table 1

Result of Descriptive Statistics of Accuracy: Pretest

| Mean | Group | N | Mean | Std. Deviation | Std. Error Mean |
|----------|--------------|----|-------|----------------|-----------------|
| Accuracy | Experimental | 15 | 58.42 | 10.23 | 2.64 |
| | Control | 15 | 58.63 | 9.63 | 2.48 |

Table 1 displays that experimental group and the control group were found to be at similar levels of English accuracy before treatment because their means were similar (control group=58.63, experimental group=58.42). To establish whether these two means were statistically significant or not, an independent samples t-test was run (see Table 2).

Table 2

Result of Independent Sample T-Test in Accuracy: Pretest

| | | Levene's Test for Equality of Variances | | | | 95% Confidence Interval of the Difference | | | | | |
|-----------------------------|--|---|------|-------|--------|---|------------|-------|------------|--------|-------|
| | | F | Sig. | t | df | Sig (2-tailed) | Difference | Mean | Std. Error | Lower | Upper |
| Equal variances assumed | | .236 | .631 | -.057 | 28 | .955 | -.206 | 3.630 | | -7.642 | 7.229 |
| Equal variances not assumed | | | | -.057 | 27.898 | .955 | -.206 | 3.630 | | -7.643 | 7.230 |

The result indicated that there was not a significant difference between the mean scores of the participants in the control group and the experimental group, $t(28) = -.057, p = .95$. In simple words, the experimental group and the control group were equal in terms of accuracy before the treatment.

To definitely answer research question one: "Does PO have any significant impact on Iranian EFL learners' accuracy?" an independent samples t-test was run again for posttest phase to compare the group means for the experimental and control groups (see Table 3).

Table 3

Result of Descriptive Statistics of Accuracy: posttest

| | Group | N | Mean | Std. Deviation | Std. Error Mean |
|----------|--------------|----|--------|----------------|-----------------|
| Accuracy | Experimental | 15 | 70.092 | 9.92 | 2.562 |
| | Control | 15 | 59.176 | 10.97 | 2.834 |

Table 4

Result of Independent Sample T-Test in Accuracy: posttest

| | | Levene's Test for Equality of Variances | | | | 95% Confidence Interval of the Difference | | | | |
|----------|-----------------------------|---|------|-------|-------|---|-----------------|-----------------------|-------|--------|
| | | F | Sig | t | df | Sig (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Accuracy | Equal variances assumed | .189 | .667 | 2.857 | 28 | .008 | 10.916 | 3.82 | 3.089 | 18.742 |
| | Equal variances not assumed | | | 2.857 | 27.72 | .008 | 10.916 | 3.82 | 3.085 | 18.746 |

Table 4 illustrates the result of posttest for the experimental and control groups. As shown in this table, the difference between the experimental group and the control group, $t(28) = 2.857$, $p = .008$, was considered to be statistically significant. That is, the participants in experimental group outperformed the ones in control group in terms of accuracy. This confirmed that PO could be beneficial for improving accuracy in speaking proficiency. Therefore, the first previously-developed null hypothesis of the study "PO does not have any significant impact on Iranian EFL learners' accuracy" was rejected.

The Result of Pretest and Posttest on Fluency

First, in order to establish the homogeneity of both control and experimental groups in terms of fluency, an independent samples t-test was carried out to examine the differences between the performances of the two groups on fluency before the treatment.

Table 5

Result of Descriptive Statistics of Fluency: Pretest

| | Group | N | Mean | Std. Deviation | Std. Error Mean |
|---------|--------------|----|--------|----------------|-----------------|
| Fluency | Experimental | 15 | 1.2427 | .27228 | .07030 |
| | Control | 15 | 1.1440 | .26907 | .06947 |

As it can be clearly seen in Table 5, the mean score of the control group ($M=1.144$) was lower than the experimental group's ($M=1.242$).

Table 6

Result of Independent Sample T-Test in Fluency: Pretest

| | | Levene's Test for Equality of Variances | | | | 95% Confidence Interval of the Difference | | | | |
|-----------------------------|--|--|------|--------|------------|--|------------|------------|-------|-------|
| | | | | Mean | Std. Error | | | | | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Difference | Difference | Lower | Upper |
| Equal variances assumed | | .184 | .671 | .998 | 28 | .327 | .0986 | .0988 | -.103 | .301 |
| Equal variances not assumed | | | .998 | 27.996 | .327 | .0986 | .0988 | -.103 | .301 | |

However, the result of the independent samples t-test did not show any significant difference in the mean scores of the control and experimental groups on the fluency, $t(28) = .998$, $p = .327$ (see Table 6). Thus, two groups were homogeneous in terms of fluency at the beginning of the training.

To determine whether PO provided for experimental group was effective in enhancing participants' fluency and consequently whether they outperformed the fluency of the control group, an independent samples t-test was performed on the posttest results of both groups (see Table 7).

Table 7

Result of Descriptive Statistics of Fluency: Posttest

| | Group | N | Mean | Std. Deviation | Std. Error Mean |
|---------|--------------|----|--------|----------------|-----------------|
| Fluency | Experimental | 15 | 1.2873 | .25423 | .06564 |
| | Control | 15 | 1.3053 | .27885 | .07200 |

Table 8
Result of Independent Sample T-Test in Fluency: Posttest

| | | Levene's Test for Equality of Variances | | | | 95% Confidence Interval of the Difference | | | |
|-----------------------------|------|---|-------|--------|-----------------|---|------------|---------|--------|
| | | | | | | Mean | Std. Error | | |
| | F | Sig. | t | df | Sig. (2-tailed) | Difference | Difference | Lower | Upper |
| Equal variances assumed | .493 | .488 | -.185 | 28 | .855 | -.01800 | .09743 | -.21758 | .18158 |
| Equal variances not assumed | | | -.185 | 27.764 | .855 | -.01800 | .09743 | -.21765 | .18165 |

Table 8 illustrates the result of posttest for the experimental and control groups. As shown in this table, the difference between the experimental and control groups, $t(28) = -.185$, $p = .855$, was not considered to be statistically significant. So the null hypothesis, predicting that "PO does not have any significant impact on Iranian EFL learners' fluency" was not rejected. That is, the participants in the experimental group did not outshine the ones in control group in terms of fluency. This confirms that PO had no significant effect on Iranian EFL learners' fluency.

Discussion

In this section, the researcher is going to begin by discussing the impact of PO on accuracy. Later on, the impact of PO on fluency is going to be discussed.

The Impact of Pushed Output on Accuracy

The first research question addressed the impact of PO on accuracy of Iranian EFL learners. The independent samples t-test showed that the students in the experimental group outperformed the students in the control group after treatment. In other words, findings of the present research indicated that PO was shown to have the potential to boost accuracy of EFL learners' speech.

In terms of accurate production, it was revealed that our study and Nobuyoshi and Ellis (1993) had the same results. In other words, in both studies, PO opportunities promoted significantly greater accuracy of target language than did non-pushed output condition. Furthermore, picture description was recognized as an effective tool in drawing learners' attention to grammatical accuracy of target language. The significant impact of PO task was also claimed by Mennim (2007) who detected the development in participants' accurate production of target language nine months after the beginning of the treatment.

As mentioned above, learners were exposed to clarification request during PO treatment. There was one good theoretical reason that clarification request followed by learners' PO led to higher scores in the experimental group. Clarification requests were chosen due to the fact that previous

investigations showed that these types of feedback provided more opportunities for language learners to produce output (Anton, 1999; de Bot, 1996; Linnell, 1995; Lyster & Ranta, 1997; Pica, 1988; Swain, 1985). In a study by Nobuyoshi and Ellis (1993), clarification requests helped learners in direction of greater accuracy in their production. Therefore, in the current study, one reason why participants in the experimental group were more successful in developing their accuracy probably lied in the nature of clarification requests in PO treatment. As the learners reproduced and reformulated what they said to make sure that their utterances were accurate and understood, they simultaneously paid heed to grammatical forms in their production which pushed them to be more accurate in their target language.

Shehadeh (2002) claimed that after well over a decade of studies on Swain's (1985) POH, there was shortage of results revealing that learners' output or output modifications had any impact on L2 learning. As mentioned by Krashen (1998), "output and especially comprehensible output was too scarce to make a real contribution to linguistic competence" (p.180). However, the findings of the present research were contrary to Shehadeh's (2002) expectations and in contrast with Krashen's hypothesis (1998) and lent support to Swain's (1985) hypothesis that modified output could be related with L2 learning in developing linguistic competence.

The Impact of Pushed Output on Fluency

The second null hypothesis of the research that stated PO did not have any significant impact on Iranian EFL learners' fluency was confirmed based on the results obtained from this study. This meant that the experimental group did not seem to benefit from being 'pushed' in terms of fluency.

One underlying explanation why in this study PO did not succeed to be more influential in developing fluency might be the nature of PO tasks or the way PO was put into practice. The findings of the second research question were in line with Krashen's Input Hypothesis and Long's Interaction Hypothesis suggesting that input modification and interaction could promote the process of L2 learning not the production.

Another possible interpretation for the result of the second research question in which the students in the experimental group did not show any significant gains in fluency might be due to feedback they received. According to Brown (2001), in teaching speaking skills, it is really important that the instructor be very sensitive to students' need to practice freely and openly without fear of being corrected at every minor flaw. Also, Harmer (2007) noted that feedback which students received during speaking would have significant impact not only on how they performed at the time but also how they behaved in fluency activity in the future. Hence, there was a consensus among all scholars in this regard as they all stressed the importance of not impeding or distracting students' attempts to speak during fluency activity (Brown, 2001; Harmer, 2007; Hedge 2000; Scrivener, 2011). However, in PO activities when a language learner produced an utterance that was not comprehended, the listener responded with a clarification request, which caused the learner to subsequently reformulate the problematic utterance. This stopping and reformulation might hinder fluency of learners in their performance in speaking test.

The results of the study were also consistent with previous experiments which were concerned with accuracy, and fluency as measures of language learners' speech production. Skehan and Foster (2001) claimed that attentional resources were restricted and that to notice one feature of oral performance might well mean that other features were neglected. They asserted that for developing language, a balance needed to be determined between these oral speech features. So, in this study, the experimental group improved in accuracy instead of fluency.

With respect to planning and preparation in tasks, the findings of this study mirrored those of Foster and Skehan (1996), Skehan and Foster (1997), and Yuan and Ellis (2003). The results of all these researchers revealed that when learners were given some time to organize what they were to talk about, there was a statistically significant improvement in terms of fluency. On the other hand, Nation and Newton (2009) claimed that language learners were pushed in their production when they were asked to produce target language without opportunity for planning or preparation. Therefore, it can be argued that due to lack of preparation in PO practices, the participants in experimental group did not improve significantly in oral fluency.

Last but not least, the major explanation for the non-significant changes can be that the duration of the investigation was short. Considering that it was difficult to expect immediate improvement in learners' fluency, it is debatable that a four-week treatment time was plainly not long enough to calculate significant changes in fluency. Although there were some individual learners who improved significantly in the experimental group, total results did not show the same significant improvement.

Conclusion

The motivation for the current investigation derived from two hypotheses which were formulated regarding the effect of PO on accuracy and fluency. Up to the present, most evidence advocating Swain's (1985) POH has been qualitative in nature. This study has added quantitative investigation to our understanding of how PO contributed to accuracy and fluency; therefore, this study tried to provide empirical support for Swain's POH. The results of this study proposed that PO might assist learners to develop accuracy.

The findings relating to the accuracy were of interest. While the method applied in measuring accuracy was not infallible in this study, the evidence presented here certainly suggested that accuracy as one aspect of speaking proficiency was affected by PO. Since it was observed that accuracy in fact improved to a statistically significant degree when learners engaged in PO, we could hypothesize that POH, certainly in the Iranian context, contributed significantly to accuracy. Interestingly, the result displayed that simply exposing students with PO might not be a sufficient condition for inducing developments in fluency.

Generally, COH is one way of teaching which is based upon too many years of research and practical application by hundreds of thousands of scholars and applied linguists, and now exists for

virtually every imaginable instructional purpose. Moreover, we now recognize a great deal about the impacts of POH work on students and the conditions necessary for accuracy and fluency. The positive findings in this study revealed that POH was an effective technique for triggering learners to notice the target form and had a positive effect on accuracy. This study is significant in that it was conducted in an EFL environment where exposure to the target language was limited and the interlocutors were most likely peers who shared the same L1.

In view of the results obtained in the current study, some pedagogical implications are suggested. One pedagogical implication of this study relates to the use of PO as a means for increasing accuracy. So, curriculum designers and material developers are recommended to include POH in the English textbooks. Also, it seems that English teachers and practitioners can play a significant role in helping their learners overcome their problems through providing them with opportunities to make use of their linguistics resources. It is therefore recommended that the instructors use PO tasks as follow-up activities for teaching grammatical forms which are specifically difficult for students to use correctly.

All research studies have limitations, and the present study is no exception. This study had inevitable limitations in examining the impact of PO on the development of accuracy and fluency. One important limitation of the research was the length of the treatment. So, various long term investigations, consequently, are required to be studied. The second limitation of this study concerning measures was that the present study focused on quantitative differences in participants' accuracy and fluency and did not include qualitative assessment. Hence, Future investigation is also essential to study the effects of PO on learners' fluency and accuracy improvement from a more qualitative point of view. As this study was done with a small number of participants (thirty), further investigations are expected to gather data from a greater number of participants to increase the external validity or generalizability. Finally, since the current study evaluated the impact of PO on females, these results may not be generalized to males. So, further studies with males or both males and females are essential to be able to generalize the findings of further studies.

References

- Achard, M., & Niemeier, S. (Eds.). (2004). *Cognitive linguistics, second language acquisition, and foreign language teaching*. Berlin: Mout de Gruyter.
- Anton, M. (1999). The discourse of a learner-centered classroom: Sociocultural perspectives on teacher-learner interaction in the second-language classroom. *Modern Language Journal*, 83, 303–318.
- Brown, D. (2001). *Teaching by principle: An interactive approach to language pedagogy*. Addison Wesley Longman: New York.

- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 23–48). Harlow: Longman.
- Cambridge Examinations Publishing. (2006). *IELTS 5*. Cambridge: Cambridge University Press.
- Cambridge Examinations Publishing. (2007). *IELTS 6*. Cambridge: Cambridge University Press.
- Cambridge Examinations Publishing. (2009). *IELTS 7*. Cambridge: Cambridge University Press.
- Cambridge Examinations Publishing. (2011). *IELTS 8*. Cambridge: Cambridge University Press.
- Cambridge Examinations Publishing. (2013). *IELTS 9*. Cambridge: Cambridge University Press.
- Chambers, F. (1997). What do we mean by fluency? *System*, 25, 535-544.
- Colina, A.A., & Garcia-Mayo, M.P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91–117). Clevedon: Multilingual Matters.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367–383.
- De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning*, 46, 529-555.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Ferris, D. (2002). *Treatment of error in second language student writing*. Ann Arbor, MI: University of Michigan Press.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3), 299–323.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354-375.
- Gass, S.M., & Mackey, A. (2007). Input, interaction and output: An overview. *AILA Review*, 19, 3-17.
- Guillot, M. (1999). *Fluency and its teaching*. Clevedon: Multilingual Matters.
- Harmer, J. (2007). *The Practice of English language teaching*. Essex: Pearson Longman.

- Hedge, T. (2000). *Teaching and Learning in the Language Classroom*. Oxford: OUP.
- Hill, L. A. (1988). *Steps to understanding* (24th ed). New York: Oxford University Press.
- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 Fluency. *Language Learning Journal*, 36(2), 153–66.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30 (4), 461-473.
- Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, 24, 541-577.
- Izumi, S., & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly*, 34, 239-278.
- Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the output hypothesis: Effects of output on noticing and second language acquisition. *Studies in Second Language Acquisition*, 21, 421-452.
- Izumi, Y., & Izumi, S. (2004). Investigating the effects of oral output on the learning of relative clauses in English: Issues in the psycholinguistic requirements for effective output tasks. *Canadian Modern Language Review*, 60 (5), 587-609.
- Iwashita, N. (1993). *Comprehensible output in NNS-NNS interaction in Japanese as a foreign language* (Unpublished master's thesis). University of Melbourne, Australia.
- Iwashita, N. (1999). Tasks and learners' output in nonnative-nonnative interaction. In K. Kanno (Ed.), *The acquisition of Japanese as a second language* (pp. 31-53). Amsterdam: John Benjamins' Publishing Company.
- Jakeman, V. & McDowell, C. (2008). *New Insights into IELTS*. Cambridge. Cambridge University Press.
- Kawauchi, C. (2005). *Pre-task planning in L2 oral performance: Quantitative and qualitative approaches*. Tokyo: Kinseido.
- Krashen, S. (1998). Comprehensible output? *System*, 26, 175-182.
- Kuiken, F., & Vedder, I. (2007). Cognitive task complexity and linguistic performance in French L2 writing. In M.P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp.117–135). Clevedon: Multilingual Matters.

- Lambert, C.P., & Engler, S. (2007). Information distribution and goal orientation in second language task design. In M.P.G. Mayo (Ed.), *Investigating Tasks in formal language learning*, (pp. 25-43). Clevedon: Multilingual Matters.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- Lee, L. (2002). Synchronous online exchanges: A study of modification devices on non-native discourse. *System*, 30, 275-288.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.
- Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25-42). Michigan: The University of Michigan Press.
- Linnell, J. (1995). Can negotiation provide a context for learning syntax in a second language? *Working Papers in Educational Linguistics*, 11, 83-103.
- Lynch, T. (1997). Nudge, nudge: Teacher interventions in task-based learner talk. *ELT Journal*, 51(4), 317-325.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37-66.
- Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21(4), 557-587.
- Mackey, A. (2012). *Input, interaction and corrective feedback in L2 classrooms*. Oxford: Oxford University Press.
- Mackey, A., & J. Philip (1998). Conversational interaction and second language development: Recasts, responses and red herrings? *Modern Language Journal*, 82, 338-356.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition* 20, 83-108.
- Mennim, P. (2007). Long-term effects of noticing on oral output. *Language Teaching Research* 11(3), 265 - 280.
- Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. London and New York: Routledge.

- Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research*, 14(4), 397-419.
- Nation, I.S.P. (2011) Second language speaking. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp.444-454). New York: Routledge.
- Nation, I.S.P., & Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.
- Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *ELT Journal*, 47, 203-210.
- Norris, J.M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555-578.
- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21(1), 109-148.
- Pica, T. (1988). Interlanguage adjustments as an outcome of NS-NNS negotiated interaction. *Language Learning*, 38, 45-73.
- Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11, 63- 90.
- Pica, T., Lincoln-Porter, F., Paninos, D., & Linnell, J. (1996). Language learners' interaction: How does it address the input, output and feedback needs of L2 learners? *TESOL Quarterly*, 30, 59-84.
- Plough, I.C., Briggs, S.L., & Van Bonn, S. (2010). A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. *Language Testing*, 27, 235-260.
- Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language Teaching Research*, 12, 1, 61- 82.
- Reinders, H. (2009). Learner uptake and acquisition in three grammar-oriented production activities. *Language Teaching Research*, 13(2), 201-222.
- Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45, 99-140.
- Scrivener, J. (2011). *Learning Teaching: The essential guide to English language teaching*. Oxford, UK: Macmillan.
- Shehadeh, A. (1999). Insights into learner output. *English Teaching Forum*, 37(4), 2-6.

- Shehadeh, A. (2001). Self- and other-initiated modified output during task-based interaction. *TESOL Quarterly*, 35, 433-457.
- Shehadeh, A. (2002). Comprehensible output, from occurrence to acquisition: An agenda for acquisitional research. *Language Learning*, 52, 597-647.
- Skehan, P. (1996). Second-language acquisition research and task-based instruction. In J. Willis, & D. Willis (Eds.), *Challenge and change in language teaching* (pp.17-30). Oxford: Heinemann.
- Skehan, P., & Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task based learning. *Language Teaching Research*, 1, 185-211.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93-120.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183-205). Cambridge: Cambridge University Press.
- Song, M., & Suh, B. (2008). The effects of output task types on noticing and learning of the English past counterfactual conditional. *System*, 36, 295-312.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, Mass: Newbury House.
- Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, 50, 158-164.
- 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 honour of Henry G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
- Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). New York: Cambridge University Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In: Hinkel, E. (Ed.), *Handbook on research in second language teaching and learning* (PP. 471-484). Mahwah, NJ: Lawrence Erlbaum.
- Swain, M., Brooks, L., Yoon, C., & Al-Alawi, M. (2012). *Contextualizing performances: Comparing perceptions of, and performances during, TOEFL iBT and real-life academicspeaking activities*. Final Report to TOEFL Research, Educational Testing Service.

- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processing they generate: A step towards second language learning. *Applied Linguistics*, 16, 371-391.
- Swain, M., & Lapkin, S. (2000). Task-based second language learning: The use of the first language. *Language Teaching Research* 4, 251-75.
- Takashima, H., & Ellis, R. (1999). Output enhancement and the acquisition of the past tense. In R. Ellis (Ed.), *Learning a second language through interaction* (pp. 173-188). Amsterdam: John Benjamins.
- Tonkyn, A. (2007). Short-term changes in complexity, accuracy and fluency: Developing progress-sensitive proficiency tests. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.), *Complexity, Accuracy and Fluency in Second Language Use, Learning and Teaching*. Contactforum.
- Towell, R., & Dewaele, J. M. (2005). The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. In J.M. Dewaele (Ed.), *Focus on French as a Foreign language: Multidisciplinary approaches*. Clevedon: Multilingual Matters.
- Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14, 85-106.
- Wolfe-Quintero, K., S. Inagaki, & H. Y. Kim. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Hawaii: University of Hawaii, Second Language Teaching and Curriculum Center.
- Yoshimura, F. (2006). Does manipulating foreknowledge of output tasks lead to differences in reading behaviour, text comprehension and noticing of language form? *Language Teaching Research*, 10(4), 419-434.
- Yuan, F., & Ellis, R. (2003). The effect of pre-task planning and on-line planning on fluency, complexity, accuracy in L2 monologic oral production. *Applied Linguistics*, 23(1), 1-27.

Aram Reza Sadeghi is an assistant professor in English Language and Literature Department at Semnan University for 17 years. He has given lectures at national and international conferences. His research interests, which resulted in publications as well, include culture in EFL, educational technology, and English as an International language.

Vahid Edalati Bazzaz received his MA in TEFL from Semnan University, Iran in 2014. He has been teaching at Jihad Daneshgahi for 10 years. He has a publication in MJLTM. He has also given a lecture in the 1st national conference on ELT: Challenges and Innovations 2012 (Torbat-e-Heydarieh). His research interests include English language skills and second language acquisition.