Classroom Metatalk: Uncovering the Role of Elaborate vs. Limited Engagement in Fostering Iranian L2 Learners' Writing Accuracy

Karim Shabani[1*], Mona Hosseinzadeh[2]

[1*] English Department, Allameh Mohaddes Nouri University, Nour, Mazandaran, Iran,
shabanikarim@gmail.com
[2] English Department, Allameh Mohaddes Nouri University, Nour, Mazandaran, Iran,
mteach.hoseinzade@gmail.com

Article Info	ABSTRACT
Article Type: **Research Article** **Received:** **12/06/2021** **Accepted:** **08/01/2022**	The underlying goal of this study was to assess the effects of *metatalk* at *elaborate* and *limited* levels of engagement on Iranian L2 learners' grammatical accuracy in writing. Thirty-four male and female students were recruited following the administration of the Quick Oxford Placement Test (QOPT) and then randomly divided into an experimental group with elaborate engagement (n=16) and a comparison one with limited engagement (n=18). Ten narrative tasks were used during the treatment sessions, and both groups were asked to write the stories. The initial drafts were reformulated by the teacher. They were then asked to compare the two versions. The experimental group was asked to discuss the reasons for the applied changes, while the comparison group only noted the differences. The learners' sheets were scored, and their accuracy was measured drawing on Ellis and Yuan's (2004) accuracy scales. The results revealed that both groups' grammatical accuracy in writing was enhanced. However, the experimental group outperformed the comparison one since their posttest scores were statistically different. The learners' language-related episodes were also analyzed, and the qualitative scrutiny brought to the surface four patterns of interaction, including collaborative, expert-novice, dominant-dominant and dominant-passive. Finally, the current study discusses implications for L2 instructional settings and the use of metatalk as a means to enhance noticing the target forms and expedite their grammatical accuracy and learning processes. ***Keywords:*** Accuracy, Elaborate Engagement, Limited Engagement, Metatalk

1. Introduction

A long-standing concern reverberating over the minds of second language researchers is how best to teach English grammar, an issue on which there has been little consensus to date. Accordingly, many methods, techniques, and approaches have been proposed, one being the language-focused instruction that provides an opportunity to attend to the linguistic forms. It is argued that letting the students notice the linguistic features in task-based activities raises awareness of language forms leading to the development of explicit knowledge that can be later on used in real communication (Nation, 2002). Metatalk, as a tool for language-focused instruction, can be used to direct the learners' deliberate attention to linguistic forms (Faerch, 1985). Metatalk is grounded on Vygotsky's Sociocultural theory (SCT) proposed in the 1990s and is one of the several functions of output stated by Swain (1993), which was later replaced by the term *languaging* defined as "tools of mind, mediating the cognition and recognition of experience and knowledge" (Swain, 2006, p. 106). Swain, further, believed that metatalk could occur in the form of writing as well as speaking. Thus, languaging was suggested as an alternative term (Swain, 2006).

Through metatalk, learners reflect on their language and verbalize metalinguistic knowledge (Myhill & Newman, 2016). For example, to fulfill a complex task, the learners may resort to explaining it to someone else, talking aloud or whispering to themselves and, in this way, facilitating its undertaking (Swain et al., 2009). Metatalk could develop L2 acquisition through enhancing awareness of linguistic rules and instigating noticing (Kuiken & Vedder, 2002). Metatalk could mediate the learners' cognitive process, especially when they engage in a more elaborate level where the linguistic features are discussed, and reasons are verbalized as well (Swain et al., 2009). As regards the Iranian foreign language context, Moradian et al. (2017) asserted that languaging combined with direct written corrective feedback could enhance the learners' writing accuracy. However, what remained intact in their study is whether the level of engagement (i.e., elaborate vs. limited) with the task results in differential gains. Moreover, what types of teacher-student interactional patterns led to the students' development remained inchoate.

Hence, to address the aforementioned lacuna, the present study was designed to examine whether metatalk at two levels of engagement, i.e., elaborate vs. limited, leads to Iranian L2 learners' grammatical accuracy. It also made an attempt to identify the interactional patterns exchanged between the teacher and learners that contributed to the learners' writing development at each level of engagement.

2. Literature Review

The interactionist perspective views language in more social terms and foregrounds the effect of environmental language on language development concerning the three concepts of input, output, and interaction among the interlocutors (Mitchell & Myles, 2004). The input hypothesis proposed by Krashen (1985) emphasized the critical role of comprehensible input and the learners' readiness to receive it. Following Krashen's Input hypothesis, Long (1981) stressed that interaction between conversational partners adds to the comprehensibility of the input. Swain (1993), on the other hand, asserted that the effective role of comprehensible input could not be ignored. However, learners do not have the chance to produce L2. She claimed that engaging in syntactic or grammatical processing, which could occur only through language production, is important as well as the semantic processing of the input. Thus, the Output hypothesis was proposed by Swain (1993).

2.1. L2 Output and its Functions

Producing output in the target language leads to becoming conscious of one's linguistic problems and what is needed to know about L2 through which the learners can get aware of the gaps in their interlanguage preventing them from conveying their intended meaning (Swain, 1993). Nobuyoshi and Ellis (1993) argued that pushed output not only enhances the learners' accuracy but also helps them to have more control over their existing grammatical knowledge.

According to Swain (1993), the potential merits of output in L2 acquisition are hypothesis-testing, noticing, metalinguistic function or a reflection, higher level of fluency, and shifting from semantic to syntactic processing. Noticing, proposed by Schmidt (1990), occurs when a form has been consciously attended to in short-term memory, due to which the input is converted into the intake.

Sharwood Smith (1993) discussed two types of noticing required for acquisition to take place. The first type occurs when learners notice the linguistic features of the input without which the input would not be converted into the intake and is triggered through input enhancement activities. The second type occurs when learners notice the gaps or the differences between their own interlanguage output and the target language input, i.e., the "cognitive comparison" (Ellis, 1995, p. 90), which is triggered after the reception of corrective feedback. Metatalk, the third function of output, is theoretically grounded in Vygotsky's SCT-based epistemology proposed in the 1990s, which viewed the language learners as active agents

learning in social contexts and the language as a process being socially mediated. He believed that language functions as the most imperative mediating tool that plays a determining role in learning (Mitchell & Myles, 2004).

2.2. Metatalk

Metatalk, the term proposed by Faerch (1985), is taken as the explicit form of talk about language (Faerch, 1985). Swain (1998) defined it as "a surfacing of language used in problem solving; that is, language used for cognitive purposes" (p. 69), known as a consciousness-raising activity. It requires attention at a deeper level through which meaning, form, and function interact with each other (Storch, 2008). It is considered as a pedagogical tool as well, in that through verbalizing metalinguistic knowledge; teachers can discover what learners are learning (Myhill & Newman, 2016).

Swain (2006) believed that metalinguistic output is not limited to speaking and can be used in writing as well. Thus, languaging was proposed as an alternative term that could facilitate language learning. Lado (1979) made use of *languaging* as a concept referring to many uses of language that was given a new meaning in L2 by Swain (2006). Swain (2006) referred to *languaging* as "tools of mind, mediating the cognition and recognition of experience and knowledge" (p. 106).

Languaging could develop L2 learners' Zone of Proximal Development (ZPD), encourage the learners to notice their weaknesses (Yilmaz, 2016), enhance the learners' independence and L2 learning, control the metacognitive processes, improve self-regulation and lead to collaborative talk (Ishikawa, 2013). As a result, the learners have a conscious use of what has been used unconsciously before (Vygotsky, 1978) through which declarative and procedural knowledge are improved (Myhill & Newman, 2016). Metatalk is interpreted by Swain and Lapkin (1995) as

> Any segment of the protocol in which a learner either spoke about a language problem he/she encountered while writing and solved it either correctly or incorrectly, or simply solved it (again, either correctly or incorrectly) without having explicitly identified it as a problem. (p. 378)

Metatalk can be analyzed based on the language-related episodes or LREs the length of which shows the level of attention (Storch, 2008). Examining the role of languaging or more specifically metanote, about corrective feedback in developing L2 writing accuracy, Suzuki (2009) argued that, the learners' errors decreased to a great extent from their drafts to their revisions, and their writing accuracy was enhanced. Moreover, they had the

opportunity to self-regulate their L2 learning. In another study, he argued that languaging helped the learners to correct their errors more easily and have more accurate use of language through externalizing their knowledge (Suzuki, 2012). Moradian et al. (2017) also believed that provision of direct written corrective feedback combined with the learners' languaging results in noticing the gaps and weaknesses in their interlanguage. As they argued, this type of noticing turns awareness to a deeper level of attention called understanding and changes the input into intake, which could bring about more accuracy.

2.3. Elaborate vs. Limited Engagement

The term 'engagement' is defined by Storch (2008) as the extent to which the learner attends to the task throughout the learning process. Given the difficulties existing in getting access to the learner-internal processes, Storch suggests 'engagement with language' as a means to "describe the quality of the learner's metatalk" (Storch, 2008, p. 98). She distinguishes between two levels of engagement, *limited* and *elaborate*. Through limited engagement, the learners only note the linguistic items without talking about the reasons for selecting or rejecting them. However, at a more elaborate level of engagement, the learners discuss the items, and provide confirmation, explanations and alternatives as well (Leow, 1997; Storch, 2008). Elaborate engagement leads to a more profound layer of understanding that fosters the consolidation of language knowledge (Schmidt, 1990).

Following Vygotsky's theory of human learning, Storch (2002) proposed four interactional patterns namely, collaborative, dominant-dominant, dominant-passive, and expert-novice which, according to Damon and Phelps (1989), represented the two indices of 'equality' and 'mutuality'. *Equality* is defined as "authority over the task or activity" (p. 127) and *mutuality* refers to "the level of engagement with each other's contribution" (p. 127).

2.4. Research Questions

The current study delved into whether learners' metatalk, either with elaborate or limited engagement, could have any effects on developing their grammatical accuracy in the classroom context. It also seeks to identify the interactional patterns existing in the learners' language-related episodes (LREs). Hence, the following queries were formulated:

1. Does elaborate engagement through metatalk lead to higher grammatical accuracy in writing among Iranian L2 learners than limited engagement?

2. What types of interactional patterns could be found in Iranian L2 learners' language-related episodes with elaborate and limited engagement in the course of metatalk?

3. Method

The present study was carried out in line with a mixed-methods research (MMR) approach following a QUAN+qual model (Dörnyei, 2007) and a pretest-posttest quasi-experimental design. Mackey and Gass (2005) contend that experimental or quasi-experimental studies can be undertaken by including a comparison group. In the comparison group design, two or more experimental groups are included each receiving a certain treatment and then the groups are compared to find out the effects of different treatments. Accordingly, the comparison-group design was adopted in the current study. Elaborate and limited engagement were the two independent variables, and grammatical accuracy was the dependent variable.

At first, a pretest was administered to both groups, followed by an immediate posttest after the ten instructional sessions and a delayed posttest after two weeks. The produced LREs, lasting for 7 hours and 30 minutes in total, were audio-recorded and transcribed to be qualitatively examined through conversation analysis of the data. The motivation behind incorporating a QUAN+qual format of MMR was to confirm the quantitative findings and add further insights into the topic under investigation since the human-related variables are best understood when they are analyzed through multiple lenses and methodological perspectives (Jang et al., 2014).

3.1. Participants

The cohorts included 34 male and female learners selected from a pool of 59 learners from Pardis English Institute, Mahmoudabad, Iran, in 2019. They were recruited according to the results of Quick Oxford Placement Test (QOPT) version 2. They were Iranian EFL learners ranging in age between 14 and 17. They had the same L1 background and studied English for about two years. The selected learners based on the results of QOPT were randomly included in an experimental (elaborate engagement) group (N= 16) and a comparison (limited engagement) group (N= 18).

To foster a higher degree of interaction and contribution among the learners, a small sample size was preferred, which helped the researchers monitor the learners and the intervention process more suitably. Given the small sample size, more students had the opportunity to run metatalk and respond to each other's contributions.

3.2. Materials and Instruments

3.2.1. Quick Oxford Placement Test (QOPT)

At the outset, the participants took the QOPT. The test includes 60 multiple choice and cloze test items measuring the learners' knowledge of grammar and vocabulary. The learners were provided with 30 minutes to answer the questions.

3.2.2. Test Tasks

A narrative task adapted from *Family and Friends 3* (Thompson, 2010) was selected as the pretest. The learners narrated the story of the pictures. The narrative task used as the immediate posttest was adapted from the same book. Another narrative task adapted from the same book was selected as the delayed posttest. The three tasks chosen for the pretest, the immediate posttest, and the delayed posttest were pilot-tested, and the difficulty index was obtained, which equaled .81 and denoted the desirable reliability. The learners' scores were collected and analyzed by SPSS software. To see whether the scoring was reliable, the data were given to a second rater who had an MA degree in TEFL and had almost ten years of experience in language teaching. Pearson Product Moment Correlation revealed significant relationships between the raters' accuracy scores on the pretest ($p = .000$, $R = .993$, $R^2 = .98$), immediate posttest ($p=.000$, $R= .999$, $R^2 = .99$), and the delayed posttest ($p = .000$, $R = .998$, $R^2 = .99$). The tests were adapted from the same book and were given to an expert to make sure they enjoyed desirable content validity.

3.2.3. Interventional Tasks

Ten narrative tasks adapted from *Family and Friends 3* (Thompson, 2010), *Family and Friends 4* (Simmons, 2010), and *Steps to Understanding* (Hill, 1980) were selected. Tavakoli and Skehan (2005) defined such tasks as "stories based on a sequenced set of picture prompts, which are given to participants in order to elicit language performance" (p. 248). Before administering the tests and to achieve reliability purposes, a pilot study was done with an existing sample who were similar in terms of such conditions as age, proficiency level, gender and educational background.

3.3. Procedure

At the start, the QOPT was administered to measure the participants' level of L2 proficiency and ensure their homogeneity. Selection of the participants was made according to the number of their correct answers and the learners with scores between +1 and –1 standard deviation (SD) were selected. Thirty-four learners were selected based on the results. They were

randomly assigned to two groups. The pretest included a narrative task. Ten narrative tasks were put to use during the ten instructional sessions. The participants were given the required instructions before starting the treatment sessions. Both groups were free to pose questions about the meaning of the unfamiliar vocabulary from the teacher or the other learners during the task performance.

The experimental group was asked to write the stories based on the pictures and they were provided with the reformulated and native-like model of their own drafts by the teacher the next session. However, no metalinguistic explanation for the applied changes was included in the sheets. The learners were then given 20 to 30 minutes to compare their own drafts with the reformulated versions, discuss the applied changes, give confirmation, explanations, and alternatives and verbalize reasons for accepting or rejecting the applied changes.

The comparison group wrote the stories of the tasks and were provided with the reformulated versions as well. They were given 20 to 30 minutes to compare their initial drafts with their reformulated versions. However, they only repeated or noted the differences and were not allowed to discuss the items or verbalize reasons. Both groups' metatalk was accompanied by teacher intervention, and they were allowed to run metatalk with or without metalinguistic terminology using either their L1 (Farsi) or L2 (English).

A different narrative task was selected as the immediate posttest to measure the effectiveness of the interventions. Another narrative task, as the delayed posttest, was given after two weeks to measure the effectiveness of the interventions in the long run. To qualitatively measure the effectiveness of the treatment, the learners' interactions were audio-recorded and transcribed by the researchers. The patterns existing in their LREs were coded considering the two indices, i.e., equality and mutuality, proposed by Damon and Phelps (1989) and the frequency of each pattern was calculated.

3.4. Data Analysis

Following Ellis and Yuan's (2004) claims, the learners' written grammatical accuracy was assessed by calculating the percentage of clauses with no error, including errors in grammar, morphology, and vocabulary. To see whether the scoring was reliable, the data were given to a second rater who had an MA degree in TEFL and had almost ten years of experience in language teaching. The correlation between the main rater and the second rater's scores was obtained for the accuracy on the pretest ($p = .000$, $R = .993$, $R^2 = .98$), immediate posttest ($p = .000$, $R = .999$, $R^2 = .99$), and the delayed posttest ($p = .000$, $R = .998$, $R^2 = .99$). Shapiro-Wilk test was applied

to ensure the normal distribution of the data. The mean and SD were obtained for the scores on the pretest, immediate and delayed posttest. To spot any difference between the two groups' performances, an independent-samples t-test was run. Separate one-way repeated-measures (RM) ANOVAs were run for the accuracy scores of each group to find out how the treatment improved the groups' performance. A 2(group)*3(time) repeated-measures ANOVA was then used to make a comparison between the groups' performance in terms of accuracy. To confirm the results, independent-samples t-test was also conducted to compare the two groups' immediate and delayed posttest accuracy scores. The produced LREs were audio-recorded and transcribed. The existing interactional patterns in the learners' LREs were coded and the frequencies, including the number and percentage of each pattern were calculated by entering the data into SPSS. To examine the reliability of coding the interactional patterns in the learners' pair talk, Spearman's rho was conducted to measure the correlation between the main coder- and the second coder's scores on all the episodes where a significant relationship was found revealing that there was a very large correlation between the two coders' coding (p =.000, R = .978, R^2 =.95).

4. Results and Discussion

4.1. Results

4.1.1. Quantitative Analyses

Shapiro-Wilk goodness-of-fit test was run to examine the assumption of normal distribution. The probability values were .862, .628, and .070 for the pretest, immediate posttest, and delayed posttest of the Elaborate Engagement group and .268, .872, and .138 for the pretest, immediate posttest, and delayed posttest of the Limited Engagement group. The probability values were non-significant (p >.05), which attested to the non-violation of normality assumption. Hence, the data could be analyzed by the parametric tests.

RQ1: Does elaborate engagement through metatalk lead to higher grammatical accuracy in writing among Iranian L2 learners than limited engagement?
The accuracy scores of the groups were first compared to find out whether the groups were similar at the beginning of the study.

Table 1

Descriptive Statistics for the Groups' Pretest Scores of Accuracy

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Pretest Accuracy	Elaborate Engagement	16	.2394	.12524	.03131
	Limited Engagement	8	.1839	.10483	.02471

The preliminary results show a higher mean score for the Elaborate Engagement group on the pretest ($M = .23$, $SD = .12$) than the Limited Engagement group ($M = .18$, $SD = .10$). It seems that the accuracy of the Elaborate Engagement group was higher than the limited one (see Table 1). To seek any significant difference, we used an independent-samples t-test.

Table 2

Independent-Samples t-test Statistics for the Groups' Pretest Scores of Accuracy

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Eq. var. assumed	.438	.513	1.406	32	.169	.05549	.03946	-.02490	.13587
Eq. var. not assumed			1.391	29.429	.175	.05549	.03989	-.02604	.13701

The results of the Levene's test showing non-significant results revealed that the variances can be assumed to be equal ($p= .513$). The results of the t-test table, hence, show no significant difference between the groups in terms of accuracy on the pretest ($p = .169$, $t = 1.40$, $df = 32$). The mean difference equaled .05 and the 95% CI ranged from -.02 to .13 (see Table 2). Therefore, the groups were similar in accuracy at the inception. One-way repeated-measures ANOVA was run to examine the development in accuracy in the Elaborate Engagement group.

Table 3

Descriptive Statistics for the Accuracy Scores of the Elaborate Engagement Group

	Mean	Std. Deviation	N
Pretest	.2394	.12524	16
Immediate Posttest	.5163	.26262	16
Delayed Posttest	.4656	.15144	16

The results of descriptive statistics revealed that elaborate engagement enhanced the learner's accuracy scores on the immediate posttest (*M*= .51, *SD*= .26) compared with the pretest (*M*= .23, *SD*= .12). However, a decrease was observed in the mean score of the delayed posttest (*M*= .46, *SD*= .15), although being still higher than that of the pretest (see Table 3).

Table 4

One-Way Repeated-Measures ANOVA Statistics for the Accuracy Scores of the Elaborate Engagement Group

Source		Type II Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Greenhouse-Geisser	.696	1.507	.461	12.470	.001	.454
Error (Time)	Greenhouse-Geisser	.837	22.607	.037			

One-way repeated-measures ANOVA was utilized to find the significance of the difference in the accuracy of the Elaborate Engagement group across time. The results showed statistically significant difference across time ($F1.50, 22.60 = 12.47, p = .001$) which was very large (partial Eta squared= .45) (see Table 4). In accordance with Larsen Hall's (2010) recommendation on using a correction to degrees of freedom, Greenhouse-Geisser was used to report the data.

Table 5

Pair-Wise Comparisons for the Accuracy Scores of the Elaborate Engagement Group

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	-.277*	.068	.003	-.460	-.094
	3	-.226*	.039	.000	-.331	-.122
2	1	.277*	.068	.003	.094	.460
	3	.051	.066	1.000	-.127	.228
3	1	.226*	.039	.000	.122	.331
	2	-.051	.066	1.000	-.228	.127

The Bonferroni's method showed that the difference between the pretest and the immediate posttest was statistically significant (p= .003, 95% CI = -.46 to -.09). Similar results were found for the difference between the pretest and the delayed posttest (p = .000, 95% CI = -.33 to -.12). However, a non-significant difference was observed between the mean scores of the immediate and the delayed posttest (p = 1.000, 95% CI = -.12 to .22) (see Table 5). Therefore, elaborate engagement enhanced the learners' accuracy, and the effects were stable after two weeks, although a decline was seen.

One-way repeated-measures ANOVA was also run to examine the development in accuracy in the Limited Engagement group.

Table 6

Descriptive Statistics for the Accuracy Scores of the Limited Engagement Group

	Mean	Std. Deviation	N
Pretest	.1839	.10483	18
Immediate Posttest	.3333	.19906	18
Delayed Posttest	.3989	.24961	18

The results of descriptive statistics revealed that limited engagement enhanced the learner's accuracy as well. The increase is clear in the scores of the immediate posttest (M = .33, SD = .19) and the delayed posttest (M= .39, SD= .24) compared with the pretest (M = .18, SD = .10) (see Table 6).

Table 7

One-Way Repeated-Measures ANOVA Statistics for the Accuracy Scores of the Limited Engagement Group

Source		Type II Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Greenhouse-Geisser	.437	1.823	.240	9.793	.001	.365
Error (Time)	Greenhouse-Geisser	.759	30.987	.024			

The results of one-way repeated-measures ANOVA showed statistically significant results regarding differences across time (F1.82, 30.98 = 9.79, p = .001). The value of partial Eta squared also showed that the difference was very large (partial Eta squared= .36) (see Table 7).

The Bonferroni's method showed that the difference between the pretest and the immediate posttest was statistically significant (p = .011, 95% CI = -.26 to -.03). Similar results were found for the difference between the pretest and the delayed posttest (p = .001, 95% CI = -.34 to -.08). However, a non-significant difference was observed between the mean scores of the immediate and the delayed posttest (p = .794, 95% CI = -.21 to .08) (see

Table 8). Therefore, limited engagement enhanced the learners' accuracy and the effects were stable but non-significant after two weeks although an increase was seen.

Table 8

Pair-Wise Comparisons for the Accuracy Scores of the Limited Engagement Group

		Mean			95% Confidence Interval for Difference	
(I) Time	(J) Time	Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
1	2	-.149*	.044	.011	-.267	-.032
	3	-.215*	.048	.001	-.341	-.089
2	1	.149*	.044	.011	.032	.267
3	3	-.066	.057	.794	-.216	.085
	1	.215*	.048	.001	.089	.341
	2	.066	.057	.794	-.085	.216

A 2(groups)*3(time) repeated-measures ANOVA was used to check the differential performances of the two groups.

Table 9

Mixed Repeated-Measures ANOVA Statistics for the Accuracy Scores of the Groups

Source		Type II Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Greenhouse-Geisser	1.048	1.770	.592	21.026	.000	.397
Time * Groups	Greenhouse-Geisser	.084	1.770	.048	1.691	.196	.050
Error (Time)	Greenhouse-Geisser	1.596	56.636	.028			

The results indicated a significant effect for time (F1.77, 56.63 = 21.02, p = .000). Partial Eta squared also showed that the effect was very large (partial Eta squared = .39). Therefore, the two groups were largely affected by the instructions across time. The interaction effect was, however, non-significant (F1.77, 56.63 = 1.69, p = .196) and the value of partial Eta squared showed a negligible effect size (partial Eta squared = .05) (see Table 9). Therefore, the groups did not equally benefit during the treatment sessions.

The tests of between-subject effects, on the other hand, showed significant results (F1, 32 = 4.34, p = .045) and a medium difference (partial Eta squared = .119) (see Table 10). Hence, the groups were different with respect to accuracy. Independent-samples t-test was separately conducted on the scores of the immediate and the delayed posttest to spot the source of difference.

Table 10

Tests of Between-Subject Effects Statistics for the Accuracy Scores of the Groups

Source	Type II Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	12.727	1	12.727	210.155	.000	.868
Groups	.263	1	.263	4.341	.045	.119
Error	1.938	32	.061			

Table 11

Independent-Samples t-test Statistics for the Groups' Immediate Posttest Scores of Accuracy

	Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Eq. var. assumed	2.742	.107	2.304	32	.028	.18292	.07938	.02122	.34462
Eq. var. not assumed			2.267	27.828	.031	.18292	.08070	.01757	.34826

The results of the Levene's test showing non-significant results revealed that the variances can be assumed to be equal (p= .107). The results of the first line of the t-test table, hence, show that the groups did not show difference with respect to accuracy on the immediate posttest (p= .028, t= 2.30, df = 32). The mean difference equaling .18 showed a large difference (95% CI = .02 to .34, Eta squared = .14) (see Table 11).

The results of the Levene's test showing non-significant results revealed that the variances can be assumed to be equal (p = .061). The results of the first line of the *t*-test table, hence, show that the groups were not different in terms of accuracy on the delayed posttest (p = .361, t = .92, df = 32). The mean difference equaling .06 showed a small difference (95% CI = -.07 to .21, Eta squared = .02) (see Table 12).

Table 12

Independent-Samples t-test Statistics for the Groups' Delayed Posttest Scores of Accuracy

| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
| | | | | | | | | 95% Confidence Interval of the Difference | |
	F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Eq. var. assumed	3.786	.061	.928	32	.361	.06674	.07195	-.07982	.21329
Eq. var. not assumed			.954	28.463	.348	.06674	.06996	-.07647	.20994

4.1.2. The Qualitative Analysis of the Learners' LREs

4.1.2.1. Frequency of Patterns of Interaction in the Experimental Group

The findings of the qualitative analysis answered the second research question, which was as follows:

RQ$_2$: What types of interactional patterns could be found in Iranian L2 learners' language-related episodes with elaborate and limited engagement in the course of metatalk?

All interactional patterns in both groups' LREs were closely examined and coded, considering the amount of feedback provided by the learners, the number of phatic utterances used by the participants, and other indices reported in the concerned literature (Damon & Phelps, 1989). The most frequent pattern was considered as the dominant pattern in each group. Generally, 98 LREs were found in the experimental group, including 38 collaborative, five dominant-dominant, 53 expert-novice and two dominant-passive patterns of interaction. The comparison group produced 124 LREs containing eight collaborative, five dominant-dominant, 45 expert-novice, and 66 dominant-passive pattern of interaction among which dominant-passive was the most frequent and the dominant pattern in the comparison group.

Table 13

The Frequency of Patterns of Interaction in the Elaborate Engagement Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	collaborative	38	38.8	38.8	38.8
	dominant-dominant	5	5.1	5.1	43.9
	expert-novice	53	54.1	54.1	98.0
	dominant-passive	2	2.0	2.0	100.0
	Total	98	100.0	100.0	

As Table 13 shows, expert-novice was the dominant pattern in the experimental group where the learners had more opportunities to help each other and the expert, sometimes the teacher, tried to encourage the novice's contribution through providing explanations, suggestions and alternatives. For example, when the novice faced a problem during the task performance, it was followed by a follow-up question. The following excerpt depicts the expert-novice pattern:

Excerpt 1: expert-novice pattern
1. S_1: I wrote: 'suddenly, Amy saw a mouse and she was shocked'.
2. T: is it correct?
3. S_s: No! No!, shocked.
4. T: shocked or was shocked?
5. $S_{4, 5, 6}$: was shocked.
6. T: Amy was happy. Amy was sad. Amy was … .so, Amy was ..?
7. S_s: shocked. '*Was*' is needed.
8. T: yes. We need a verb. So, she was shocked. Right?
9. S_s: Yes.

The excerpt above shows an expert-novice pattern of interaction where the teacher, as an expert, tries to invite the learners' contribution (lines 2, 4, 6, and 8) and persuades them by providing some more examples (line 6). The novice learners are also willing to attend the discussion (lines, 3, 5, and 7).

The collaborative pattern was the second frequent pattern in the treatment group in which more equality and mutuality patterns were observed and the learners were provided with a great opportunity for collaboration, group work and provision of feedback to their interlocutors. The learners worked on the task collaboratively and provided explanations, suggestions and alternatives to each other and confirmed each other's utterances. However, they not only did not try to impose their opinions but also intended to persuade each other through the provision of logical explanations.

Dominant-dominant was the third frequent pattern in the experimental group. All the learners made attempts to attend the discussion. Various suggestions and alternatives were provided by them and they tried to share their opinions. However, they did not agree with each other and were not willing to reach consensus. Actually, they did not participate in the contribution of others and the learners' suggestions were not accepted by their interlocutors. The word 'no' was repeated for many times and the learners' feedback was rejected by the other learners.

Dominant-passive was the least frequent pattern in the experimental group's LREs where the learners were not willing to have contribution and did not participate in the contribution of others. One of the students, as the dominant, engaged more in the dialogue, and there was a long monologue uttered by the dominant partner, while the other participants had more passive roles and only repeated and confirmed the explanation and the suggestions provided by the dominant partner. However, they were not willing to provide suggestions or engage with the dominant learner's contribution. The dominant, on the other hand, did not try to invite the other learners' participation and had the most control. Therefore, the results showed that, expert-novice was the most frequent pattern and the learners were more willing to take the roles of expert and novice in their metatalk.

4.1.2.2. Frequency of Patterns of Interaction in the Comparison Group

The frequency of each pattern was also calculated in the comparison group and the explanations in each section were followed by an excerpt. As Table 14 shows, expert-novice was the second frequent pattern where the expert reads the two versions loudly, asks some questions to encourage the other participants' contribution and make sure that all the learners in the class were paying plenty of attention to what she was saying. The other students also listened carefully and were willing to run metatalk. They were able to find the differences between the two versions and repeat them. The learners' noticing was followed by the teacher and the expert's positive reinforcement afterward.

Although the students within the comparison group were not permitted to discuss the language items, the collaborative pattern of interaction was observed in some of the episodes as well in which the learners listened to their friends carefully and compared the two versions collaboratively. Their metatalk was not limited only to listening and repeating the difference. In essence, some alternatives were also provided by two of the learners, and there was a more elaborate form of engagement.

Dominant-dominant was the least frequent pattern in the comparison group's LREs. The learners had almost equal participation; however, the

learners' contribution and suggestions were ignored by the other participants and none of them were willing to encourage the other learners' contribution. Therefore, the upshots revealed that, dominant-passive was the most frequent pattern in the learners' LREs in the comparison group where the learners were not much willing to participate in the contribution of others.

Table 14

The Frequency of Patterns of Interaction in Limited Engagement Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	collaborative	8	6.5	6.5	6.5
	dominant-dominant	5	4.0	4.0	10.5
	expert-novice	45	36.3	36.3	46.8
	dominant-passive	66	53.2	53.2	100.0
	Total	124	100.0	100.0	

4.2. Discussion

The current study examined the role of whole class metatalk, at two levels of engagement, in improving L2 learners' grammatical accuracy. The first research question was posed to examine the effects of elaborate and limited engagement on Iranian EFL learners' grammatical accuracy.

The results revealed that both types of intervention significantly contributed to the learners' grammatical accuracy in writing with large effect sizes. However, the experimental group outperformed the comparison one in the posttest but the difference between their delayed posttest scores was not significant, showing a negligible effect size. The findings of the study revealed that through metatalk, at both levels of engagement, the learners had the opportunity to revise their output and discern the gaps in their interlanguage or their linguistic problems and the difference between the L2 input and their output, which confirmed the findings of several studies (Swain & Lapkin, 1995; Yilmaz, 2016). The learners' metatalk, reflecting Ishikawa (2013) and Swain's (1993) claims, functioned as an effective mediating tool for the learners to facilitate language learning. It could enhance the learners' conscious knowledge of rules, which helped them to exhibit a more accurate use of the L2 on their immediate and delayed posttest. These findings are also in line with Suzuki (2009), who emphasized the great effect of languaging on L2 learners' writing accuracy and as stated by Suzuki (2012), the learners' accurate use of language was the result of externalizing the knowledge through languaging. Commensurate with Arabgary and Izadpanah (2016), the learners' classroom interaction could enable the learners to be more communicatively competent and have better use of the language in real-life setting. Echoing the results of Abtahi et al.

(2020), the applied collaborative dialogue and peer feedback could function as a mediational tool for the learners.

The quantitative results were in favor of metatalk at both levels of engagement. However, the upshots revealed that the experimental group with elaborate engagement playing the role of high languagers (Swain et al., 2006) outperformed the comparison one, and their posttest scores were statistically different in terms of accuracy. In essence, the learners with more contributions were able to have a more accurate use of language. Thus, the quality of the learners' engagement or their level of attention played a great role in their performance which lends credence to the claims by Kuiken and Vedder (2002), Leow (1997), Storch (2008), and Swain et al. (2009) stating that metatalk at a more elaborate level provided more opportunities for L2 learning. The upshots indicated that the learners in the experimental group had more chances to discuss the linguistic items, note the differences between their drafts and the reformulated versions, verbalize reasons for accepting or rejecting the changes, and even provide their own suggestions and alternatives. The learners' performance on immediate and delayed posttest also indicated that even the learners with limited engagement benefited from their metatalk.

The second research question aimed at detecting the interactional patterns that accrued from the learners' LREs throughout the instructional sessions. The analysis evidenced that the most frequent pattern was the dominant pattern in each group. Generally, 98 LREs were found in the experimental group among which expert-novice, with the most frequency (n = 53), was the overall pattern. The comparison group produced 124 LREs among which dominant-passive, as the most frequent pattern (n = 66), was the overall pattern. As observed in the learners' LREs, members of the experimental group showed more willingness to establish an expert-novice interactional pattern in which the experts tried to encourage the novice learners' contribution and the novice ones received feedback and assistance from the more expert ones. Moreover, the more expert learners tried to persuade the novice ones through providing suggestions and explanations, and they had more opportunities to run metatalk at a more elaborate level of engagement which led to more noticing and, confirming Storch (2002), more learning outcomes were observed.

In the comparison group, however, most of the LREs showed dominant-passive pattern in which the dominant learners preferred to produce long monologues and were not willing to encourage their interlocutors' contribution. As a result, fewer opportunities were provided to run metatalk and, compared with the experimental group, less noticing occurred. Thus, the qualitative results, combined with the quantitative analysis of the learners' scores revealed that expert-novice pattern led to

more successful language learning than the dominant-passive one, which lends credence to the findings of the study conducted by Storch (2002).

5. Conclusion and Implications

The present study had as its aim to assess the effects of metatalk at the levels of elaborate and limited engagement on Iranian L2 learners' grammatical accuracy. The upshots revealed that both types of interventions enhanced the learners' grammatical accuracy. The analysis of the LREs indicated that expert-novice was the dominant pattern in the experimental group, while the majority of the episodes in the comparison group showed the dominant-passive pattern. Theoretically, the study stressed the role of noticing (Schmidt, 1990) as the by-product of metatalk. The findings revealed that the provision of opportunities for learners to run metatalk and reflect on their language can help them to test their hypotheses and discern the incongruities between their produced sentences and L2 target structures, which can enhance the learners' control over the task. Thus, the learners exhibited more accuracy in their use of language.

Pedagogically, as evidenced by this study and following Myhill and Newman (2016), metatalk could be used by the teacher as an instructional technique to verbalize metalinguistic knowledge and discover the learning processes. Therefore, it could function as a tool to encourage students to notice the intended forms and expedite their grammatical accuracy development. The observed findings in this research have suggestions for language teachers and syllabus designers. The teachers can explicitly instruct the concept of metatalk and exemplify its functionality in the classroom. They can allocate a certain time to metatalk activity and encourage the learners to engage in elaborate metatalk with each other by giving feedback on each other's drafts and, in this way, mediate their thoughts and facilitate their writing. It is also suggested that instructors align themselves with the expert-novice interactional patterns and engage the students in high languaging.

Despite its important findings, the present study faced some limitations as well. First, the participants were the learners in an EFL context. Therefore, the findings reported here should be cautiously generalized to the ESL context. The present study only examined the role of metatalk, that is, languaging in an oral modality, and more studies need to be carried out to compare the role of metatalk with metanote in a written modality. The audio data used in this study imposed some constraints in interpreting the recordings and left some aspects of the authentic talk exchanges unknown. The contribution of paralinguistic features like facial expressions, body gestures, laughs, chuckles, nods, gazes and the like as instances of tacit

communication in collaborative dialogues remained elusive. A video recording of the exchanges is suggested to fully capture the multidimensional aspects of natural communication during task performance.

Acknowledgments

References

Abtahi, M., Abadikhah, S., & Dehqan, M. (2020). The influence of computer-based and collaborative handwritten peer feedback on the writing performance of EFL learners. *Journal of Modern Research in English Language Studies, 3*(7), 95-113.

Arabgary, H., & Izadpanah, S. (2016). The study of turn-taking processes through focus on form and focus on forms instructions: Incidental grammar acquisition. *English Language Teaching, 3*(3), 105-128.

Damon, W., & Phelps, E. (1989). Critical distinctions among three approaches to peer education. *International Journal of Educational Research, 58*, 9-19.

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Ellis, R. (1995). Interpretation tasks for grammar teaching. *TESOL Quarterly*, *29*(1), 87–105.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*, 59- 84.

Faerch, C. (1985). Metatalk in FL classroom discourse. *Studies in Second Language Acquisition, 7*, 184–99.

Hill, L. (1980). *Steps to understanding*. Oxford University Press.

Ishikawa, M. (2013). Examining the effect of written languaging: The role of metanotes as a mediator of second language learning. *Language Awareness, 22*(3), 220-233.

Jang, E. E., Wagner, M., & Park, G. (2014). Mixed methods research in language testing and assessment. *Annual Review of Applied Linguistics, 34,* 123–153.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

Kuiken, F., & Vedder, I. (2002). The effect of interaction in acquiring the grammar of a second language. *International Journal of Educational Research, 37*, 343–358.

Lado, R. (1979). Thinking and "languaging": A psycholinguistic model of performance and learning. *Sophia Linguistics, 12*, 3-24.

Leow, R. P. (1997). Attention, awareness, and foreign language behaviour. *Language Learning, 47*(3), 467–505.

Long, M. H. (1981). Input, interaction and second language acquisition. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 259-278). Annals of the New York Academy of Sciences.

Mackey, A., & Gass, S. (2005). *Second language research.* Lawrence Erlbaum Associates.

Mitchell, R., & Myles, F. (2004). *Second language learning theories*. Hodder Arnold.

Moradian, M. R., Miri, M., & Nasab, M. H. (2017). Contribution of written languaging to enhancing the efficiency of written corrective feedback. *International Journal of Applied Linguistics, 27*(2), 406-421.

Myhill, D., & Newman, R. (2016). Metatalk: Enabling metalinguistic discussion about writing. *International Journal of Educational Research*, *80*, 177–187.

Nation, P. (2002). Best practice in vocabulary teaching and learning. In J. C. Richards, & W. A. Renandya (Eds.), *Methodology in language teaching* (pp. 267-272). Cambridge University Press.

Nobuyoshi, J., & Ellis, R. (1993). Focused communication tasks and second language acquisition. *ELT Journal*, *47*(3), 203-210.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*, 165-179.

Simmons, N. (2010). *Family and friends 4*. Oxford University Press.

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning, 52*(1), 119–158.

Storch, N. (2008). Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness*, *17*(2), 95–114.

Suzuki, W. (2009). Improving Japanese university students' second language writing accuracy: Effects of languaging. *Annual Review of English Language Education in Japan, 20*, 81-90.

Suzuki, W. (2012). Written languaging, direct correction, and second language writing revision. *Language Learning, 62*(4), 1110-1113.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modem Language Review, 50,* 158-64.

Swain, M. (1998). Focus on form through conscious reflection. In C. Dougthy, and J. Williams (Eds.), *Focus on Form in classroom second language acquisition* (pp. 64–82). Cambridge University Press.

Swain, M. (2006). Languaging, agency and collaboration in advanced language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-108). Continuum.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*, 371–391.

Swain, M., Lapkin, S., Knouzi, I., Suzuki, W., & Brooks, L. (2009). Languaging: University students learn the grammatical concept of voice in French. *The Modern Language Journal, 93*(1), 5-29.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-273). John Benjamins.

Thompson, T. (2010). *Family and Friends 3*. Oxford University Press.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Yilmaz, M. (2016). Improving Turkish EFL learners' writing accuracy: Effects of written languaging and languaging types. *Procedia - Social and Behavioral Sciences, 232*, 413-420.