Inspecting Task-Induced Involvement from the Perspective of Sociocultural Theory

Shiva Kivanpanah
Associate Professor
University of Tehran
Shkaivan@ut.ac.ir

Mowla Miri *
PhD in TEFL
Allameh Tabataba'i University
molamiri84@gmail.com

Abstract

Grounded in sociocultural theory (SCT), this study explored whether the hypothesized difference in task-induced involvement could affect the actual realization of evaluation, one of the cognitive dimensions of the Involvement Load Hypothesis (ILH). A group of 24 Iranian EFL learners participated in the study. They were paired up to write a composition including ten unknown words in the first session and then completed a cloze task with another set of ten new words in the second one. Collaborative dialogues in both sessions were audio-recorded, transcribed verbatim and micro-genetically analyzed to trace how the value of hypothesized evaluation could affect the manifestation of evaluation during collaborative dialogues. In line with the tenets of ILH, the results of the micro-genetic analysis demonstrated that using target words in the composition task could induce a higher degree of evaluation than using them in the cloze task. In light of the findings, researchers are suggested to look at issues from different standpoints rather than restricting themselves to one single theoretical perspective.

Keywords: Task-induced involvement, evaluation, sociocultural theory (SCT), micro-genetic analysis

Despite the plethora of suggestions on ways to foster incidental vocabulary learning, there existed a lacuna regarding a viable method to measure the efficiency of word-focused tasks. To narrow down this gap, Laufer and Hulstijn (2001) ventured to propose the Involvement Load Hypothesis (ILH), which is premised on the tenet that tasks with varying

degrees of 'need,' 'search', and 'evaluation' bring about different gains in vocabulary learning since they induce different attentional and cognitive processing loads. To examine this hypothesis, Laufer and Hulstijn called for conducting empirical studies. However, their call has been responded by few studies to date (e.g., Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008, Yaqubi, Rayati, & Allemzade Gorji, 2010). Besides the scarcity in number, the studies have placed their prime focus on task products or test outcomes, without seeking to detect the underlying processes of word-focused tasks in action or tracking down how tasks with distinct involvement load prompt students to obtain different vocabulary gains.

It could be argued that this overemphasis on outcomes largely stems from the theoretical framework, the cognitive approach, within which these studies have been conducted. From this standpoint, learning is primarily an individualistic undertaking which occurs employing internal cognitive processes such as memory and attention (Ellis, 2008). Accordingly, participants in the aforementioned studies on ILH were requested to complete some tasks with different involvement loads (e.g., fill-in-the-blanks and composition writing); subsequently, the learners received a post-test on some target words used in the tasks, which was followed by explaining the differences in vocabulary gains in light of the hypothesized differences in the involvement loads. Hence, being restricted to the tenets of cognitive theory, tracking processes stimulated by the tasks received marginal attention, if not gone unnoticed altogether. To ameliorate this shortcoming, the present study drew upon theoretical tenets and methodology of Vygotskian socio-cultural theory (henceforth SCT) since this theory is concerned with "the processes of attention and memory and how these are revealed in learner engagement in L2 interaction" (Foster & Ohta, 2005, p. 403).

SCT holds that all types of learning and cognitive functions like selective attention and problem solving are initially co-constructed in a social milieu and they are then gradually internalized by the individuals and stored internally (Lantolf & Thorne, 2006; Vygotsky, 1978). From the SCT viewpoint, the processes giving rise to an outcome can be traced

through micro-genetic analysis, meaning the moment-to-moment study of learners' collaborative dialogues (Ellis & Barkhuizen, 2005; Lantolf, 2000; Swain, 2000, 2006). A unique segment of collaborative dialogues in which the formation of cognitive processes can be better viewed is known as Language Related Episodes (LRE). LREs are, in fact, parts of collaborative dialogues in which "the students talk about the language they are producing, questioning their language use, or correct themselves or others" (Swain & Lapkin, 1998, p. 326). Tocalli-Beller and Swain (2005) highlighted that the LREs in which co-participants weigh their proposed options against each other to reach an agreed response stimulate a larger amount of cognitive evaluation; they went on to label them as cognitive conflict episodes (CCEs). Thus, it can be argued that by micro-analysis of the CCEs in actuality, the process of evaluation, which is induced by using a word in a suitable context, can be observed and documented.

In brief, this study adopted a sociocultural perspective to document how word-focused tasks with different involvement loads affect the realization of evaluation, one of the cognitive dimensions of ILH. However, before proceeding to report on the study, presenting an account of ILH and pertinent tenets of SCT is in order.

Review of the Literature
Involvement Load Hypothesis

For years, researchers shared the opinion that vocabulary retention is primarily contingent on the quantity and quality of attention allocated to various aspects of vocabulary knowledge; that is, "the more attention is paid to the semantic and formal aspects of the words and the richer the associations are made with existing knowledge … the higher are the chances that the new information is retained" (Laufer & Hulstijn, 2001, p. 2). What is more, it was assumed that motivation could play a noticeable part in vocabulary learning. Nonetheless, research had failed to operationalize and examine the role of attention and motivation in vocabulary learning up until the time Laufer and Hulstijn (2001) proposed ILH.

ILH constitutes three components, namely 'need', 'search', and 'evaluation'. Need is the motivational dimension of ILH, which is closely associated with the 'need to achieve'; it is conceptualized as "a drive to comply with task requirements, whereby the task requirement can be either externally or internally imposed or self-imposed" (Laufer & Hulstijn, 2001, p. 14). Two degrees of prominence or strength are suggested for need: strong and moderate. When learners are internally motivated, they feel a robust intrinsic need to find out an unknown word, whereas the learners may not experience such a strong need when an external agent asks them (e.g., a teacher) to look up a word and retain it. Therefore, need is conceived as moderate for this externally demanded vocabulary search (Laufer & Hulstijn, 2001).

Search and evaluation, the cognitive components of ILH, are closely intertwined with noticing and thoughtful attention to form and meaning associations. Search refers to an attempt undertaken to find the meaning of an unknown L2 word or to find a suitable form to express an intended concept. It can be materialized by referring to a dictionary, a teacher, or a peer. Search is said to be either present or absent. It is present when learners seek to obtain knowledge about a word, but it is absent when required information about a target word is given to them, and they feel no need to make any quest. A good case in point for the latter is when the meaning of new words is glossed in the margin of passage.

Evaluation is interpreted as, "[…] a comparison of a given the word with other words, specific meaning of a word with its other meanings, or comparing words with other words in order to assess whether a word does or does not fit the context" (Hulstijn & Laufer, 2001, p. 544). Evaluation can carry two degrees of prominence, namely moderate and strong. It is moderate when learners should choose a word or one meaning of a polysemous word (i.e., a word with multiple related meanings) which best suits a given context (as in the case of a fill-in-the-blank task with some words given). By contrast, when the learners are required to decide how to combine a target word with other words to produce an original sentence, evaluation is robust since they have to weigh the target word against other

words in its context to see whether it is appropriate in terms of semantic, syntagmatic, and pragmatic features (Laufer & Hulstijn, 2001).

Laufer and Hulstijn (2001) argued that each of these three dimensions could be present or absent for a given task and they can also take different levels of prominence. The combination of these factors with their degrees of prominence is referred to as 'task-induced involvement' or 'involvement load'. Then to set the scene for objective measurement of a task effectiveness, a numerical involvement index was allocated to each task, "where absence of a factor is marked as 0, a moderate presence of a factor as 1, and a strong presence as 2" (Hulstijn & Laufer, 2001, p. 544). A task in which learners are expected to read a text with some words glossed in the margin at their teacher's request induces an involvement load of 1 since the need is moderate and search and evaluation are both absent $(1 + 0 + 0 = 1)$. Another task in which the learners are required by their teacher to write a composition including some words given along with their glossed meanings induces involvement index of 3 as the teacher externally imposes need, search is absent, and evaluation is reliable due to the fact that new words are weighed against other words in a self-generated context $(1 + 0 + 2 = 3)$.

ILH is based on the following assumptions. First, "retention of words when processed incidentally is conditional upon the following factors in a task: need, search, and evaluation" (Laufer & Hulstijn, 2001, p. 14). Second, "other factors being equal, words which are processed with higher involvement load will be retained better than words which are processed with lower involvement load" (Laufer & Hulstijn, 2001, p. 16). As a result, it can be argued that the composition task can be more conducive to vocabulary learning than fill-in-the-gap since it induces a higher involvement load or cognitive processing.

Empirical Studies on ILH

Some studies (e.g., Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Yaqubi et al., 2010) have directly tested the fundamental tenets of ILH. Initially, Hulstijn and Laufer (2001) conducted two parallel

experiments with young adult EFL learners in the Netherlands and Israel. The learners were assigned to six groups, three groups in each country to perform the ensuing tasks: reading comprehension, reading comprehension plus fill-in-the-blanks, and composition writing with some words given. Results of both studies corroborated the assumption that the composition task with the highest involvement load (involvement index = 3) would produce the highest gains in vocabulary retention, but only the results of the study in Israeli context supported the second prediction that task 2 (reading plus fill-in-the-blanks) would result in significantly higher vocabulary gains than task 1 (reading only). This study was, nevertheless, criticized on several grounds. For example, Folse (2006) claimed that the superiority of the composition task could have been associated with more time on task rather than task-induced involvement since it took different amounts of time for the three groups to complete their tasks: 40-45, 50-55, and 70-80 minutes for the reading, fill-in-the-blank, and composition tasks, respectively.

Therefore, Folse conducted a study by controlling time on task, showing that when the time was controlled, the task of completing three fill-in-the-blank sentences was more conducive than writing one original sentence with a word given. She associated the difference between her findings and those in Hulstijn and Laufer's study to the higher amount of practice in a longer time bracket than higher task-induced involvement load.

Keating (2008) also criticized Hulstijn and Laufer (2001) for restricting their study to advanced learners, testing only passive vocabulary retention, and not controlling time on task. In an attempt to remove these limitations, Keating conducted a study with 79 beginner learners of Spanish. The learners were asked to complete three tasks: a reading comprehension, fill-in-the-gaps, and a sentence writing task. Passive and active knowledge of the words was tested immediately and two weeks after the treatment. Results provided some evidence supporting the assumption that tasks with higher involvement load would foster more significant gains in vocabulary learning. The sentence writing task

(evaluation= 2) with the highest involvement load generated the best retention, the reading plus fill-in task (evaluation = 1) was the second best, and the lowest retention was achieved in the reading comprehension task (evaluation= 0). Keating (2008) stated that the result for passive knowledge was durable over time, but this was not the case for actual knowledge of the words. The results of the immediate post-test showed that task 2 was more effective than task 3 in fostering short-term vocabulary learning, but the delayed post-test did not show any significant difference between task 2 and task 3 for active vocabulary knowledge. He argued that this mismatch could be because the participants in his study were supposed to write a separate sentence using each word whilst the participants in Hulstijn and Laufer (2001) were required to write a composition with the words given. Thus, "it might be the case that producing connected discourse involves more elaborate processing of the target words than producing disconnected sentences" (Keating, 2008, p. 379).

Another study exploring the assumptions of ILH was Kim (2008), including two experiments. The first experiment was a partial replication of Hulstijn and Laufer (2001), while the second experiment examined their claim that writing new sentences and composing sentences with target words would induce the same involvement load. In the first experiment, 64 undergraduate students were assigned to three groups: reading (involvement index= 1), gap-fill (involvement index= 2), and composition (involvement index = 3). Results of this experiment were in line with those of Hulstijn and Laufer (2001) in that the composition task fostered the highest gains on both immediate and delayed posttest, vocabulary knowledge scale developed by Wesche and Paribakht (1996), but the difference between the reading and the fill-in-the-gap tasks as in Hulstijn and Laufer's (2001) Dutch-English experiment was not significant. It should be noted that time on task was equal for all groups, so the intervening effect of time on task was controlled. The second experiment examined the assumption that tasks with the same involvement loads lead to the same results in vocabulary learning; in fact, the assumption that

sentence writing and composition writing induce the same involvement loads was explored. The participants representing two proficiency groups performed either a composition writing or a sentence writing task. It was found that both tasks produced the same outcomes on immediate and delayed posttests across proficiency levels.

Moreover, Yaqubi et al. (2010) made an effort to examine the tenets of ILH in Iranian EFL context with three groups of EFL learners. The first group worked on an input-based task with the overall index of 3 whereas the second group carried out an output-based task with the same index. However, the other group performed the same input-based task but with the index of 2. Results of the study indicated that the output-based task was more conducive to developing gains in vocabulary learning and retention, which was not congruent with the predictions made by ILH. One plausible explanation for this discrepancy could be that the participants in the output-based group initially worked on a gap-fill task and then they were asked to use the ten target words to write an opinion essay on the top of the gap-fill text. That is, it could be argued that this is a "sequence of tasks" rather than one single task. Hence, the learners might have been urged to get more involved with the target words. This is acknowledged by the researchers where they maintained that "it seems to be a plausible explanation to suggest that in this study word retention was a product of repeated exposure" (Yaqubi et al., 2010, p. 161). Moreover, the time which each task took to be completed has not been specified; thus, it is not clear to claim if the edge of the output-based group was due to the task type or the amount of time spent on the task.

As the previous studies show, in exploring the underlying assumption of ILH from a cognitive perspective, the question of how the difference in task involvement loads can affect the actual activities occurred during task performance has been left out.

Task-Induced Involvement from the Sociocultural Theory Perspective
Sociocultural theory (SCT) of mind is fundamentally based on Vygotsky's (1978) view that human beings' relationships with internal and

external worlds are not direct, but it is mediated via physical (e.g., a shovel or a dictionary) or symbolic tools (e.g., language and arithmetic rules). Working through a concrete example about human beings' mediated relationship with the physical world can help us better illustrate human-mediated psychological relationship via symbolic tools. Digging a hole to plant a bush requires the use of bare hands or a shovel to enhance the efficiency and precision of work and save time and energy. To further increase the ease and efficiency, a mechanical backhoe can be utilized. Humans can utilize a varied range of mediatory tools to attain their goal. However, they are not entirely free to use the selected tools in any way they desire; in fact, these socio-culturally constructed tools impose some rules and regulations on the way they are to be utilized (Lantolf & Thorne, 2007). Put it another way, using a shovel requires different patterns of use in comparison to implementing a backhoe. In the words of Lantolf and Thorne, "physical tools allow us to change the world in ways that simple use of our bodies does not. Moreover, by transforming our social and material environment, we also change ourselves and the way we live in the world" (2007, p. 199).

From an SLA perspective to language development, Lantolf (2000) maintained that mediation falls into three categories: first, mediation through social interaction with other human beings; second, mediation by self via private speech (i.e., an audible speech addressed to self rather than others) and third, mediation through artifacts such as tasks. Lantolf added that mediation could be accomplished either externally or internally. External mediation involves providing help to a novice by an expert or an artifact such as dictionary whereas internal mediation entails drawing upon one's resources to gain further mastery or to control over a function.

Returning to Lantolf and Thorne's (2007) line of argument, it can be suggested that though humans use word-focused tasks to modify and reorganize the process of vocabulary learning, the tasks can, in turn, affect our higher-order cognitive processes such as selective attention and evaluation. There is a reciprocal relationship between mediatory tools and higher-order cognitive abilities. In other words, tasks with different

involvement loads impose different methods of use and cognitive processing on our internal world as "a shovel requires one type of motion and a backhoe another" (Lantolf & Thorne, 2007, p. 199). In brief, it can be assumed that the word-focused tasks work as artifacts mediating between our internal cognitive processes such as attention and the external domain of vocabulary knowledge.

A Critical Look at ILH

Although ILH is conceived as an effective means of evaluating the effectiveness of word-focused tasks, some criticisms can be leveled against it. One of the significant concerns with ILH lies in the fact that it was initially developed from some speculative cognitive assumptions. Moreover, it has been only supported by differences in task outcomes. In fact, as Al-Hadlaq (2003) noted, developers and proponents of ILH have failed to provide evidence for the actual occurrence of supposedly undertaken processes which underlie the real task outcomes or pertinent test scores. As a result, one can argue that the outcomes or products of each word-focused task can be due to factors other than the three dimensions of ILH, i.e., need, search and evaluation since

> [the] same products can be as a result of a variety of different processes. In the second language acquisition context, to say that a learner has full control of a particular structure or idiom tells us little about acquisition processes. Processes are socio-cognitive events that occur along the road from novice to expert. (Ohta, 2001, p. 3)

As noted above, word-focused tasks can mediate to generate various types of cognitive processes such as evaluation, which in turn can be realized at different levels. It can be noticed that this drawback mainly lies in the theoretical standpoint informing the studies on ILH, for instance, cognitive accounts of language learning with its primary focus on cognitive processes occurring inside one's head. Thus far, the difference between moderate and robust evaluation and how they differ in varied

tasks have not received due attention; therefore, the principles of SCT were employed to address this lacuna.

The Present Study

SCT can unfold the various processes resulting in different indices assigned to evaluation. The micro-analysis of learners' collaborative dialogues from the SCT perspective can open up opportunities for observing how learning is shaped and internalized at inter-psychological level (Lantolf, 2000; Swain, 2000; Swain & Lapkin, 1998). A fragment of collaborative dialogues, recognized as LRE, can afford optimal opportunities during which co-construction of knowledge and cognitive processes can be traced and viewed (e.g., Swain & Lapkin, 1998). For this reason, the LREs over the target words and how they are resolved are considered in the current study to depict how task type can influence the process of evaluation.

Depending on the ways learners resolve them, LREs are classified as Cognitive Conflict Episodes (CCEs) and non-CCEs (Tocalli-Beller & Swain, 2005). The former induces initial disagreement over their resolution and require learners to discuss the points of conflict by drawing on available resources (e.g., glossary) and strategies (e.g., request for clarification and pauses); as a result, the parties involved in a CCE can develop an agreement and resolve the matter of conflict. The latter requires learners to share the same opinion in the target forms and acknowledge each other's contributions rather than discussing and debating over them. Learners confirm or repeat their partners' comments and suggestions. Thus, it can be argued that engagement in CCEs is more conducive to increasing evaluation degree since learners have an opportunity to weigh their choices against their partners' options. They have to weigh proposed options against each other, offer arguments and counter-arguments, and deliberate upon the choices at hand in order to find a solution to the conflicts they are facing, so cognitive conflicts can be indicative of a higher degree of evaluation. The larger the number of CCEs is, the higher the degree of evaluation. On the other hand, it stands to reason to

hypothesize that agreement and consensus are less likely to motivate learners to be engaged in serious cognitive challenges in order to evaluate possible options and alternatives against each other; as a consequence, the level of induced evaluation in the case of non-CCEs tends to be lower than CCEs.

In light of the preceding discussion, it was hypothesized that the effect of task type on evaluation could be viewed in the LREs produced in collaborative dialogues over word-focused tasks; in fact, the evaluation could be documented and operationalized by both qualitative micro-genetic analysis of pertinent LREs and counting the frequency of CCEs. To accomplish the objectives of the study, the following question is addressed:

How can task type affect the realization of evaluation in word-focused tasks?

Method

Setting and Participants

Participants were an entire class (n = 24) of Iranian EFL learners in a private language institute in Iran. Their age ranged from 16 to 18 years, and they were at the intermediate level of language proficiency based on the institute placement test results. According to registration documents, it was evidenced that all the participants had been studying in the Institute for about 18 months and had read the same textbooks. The class was regularly held twice a week, 105 minutes per session. Most of the instructional time was devoted to developing listening and speaking skills since the majority of the learners intended to learn English for oral communication purposes; nonetheless, reading and writing activities were covered as part of their institution curriculum as well. The teacher checked out their reading exercises and reviewed them in class. Additionally, he took their writing assignments home and provided them with written corrective feedback; then, the learners were required to revise their initial writings based on the feedback and resubmit them to the teacher.

The data was collected by the second researcher (henceforth T-R) who was the regular teacher of the class. He had five years of experience in teaching English as a Foreign Language (EFL). The participants are referred to as S1, S2, S3 ..., to maintain their anonymity.

Target Words

The following steps were undertaken to select 20 words used in the current study. Initially, the T-R chose 50 words which were likely to be unknown to the participants; half of the words were semantically related to the task topics whereas the other half were not semantically related to the tasks (general words). The list of words was given to a number of students (n= 16) homogeneous with participants in the study; they were asked to tick the words which they already knew. It was assumed that they might overestimate their vocabulary knowledge, so they were obliged to write the Persian translation or synonym for the words they checked as known. Then the words ticked as known by more than two of the students were deleted from the list. Afterward, from the remaining words, five words semantically related to crime (the topic of the cloze) and five general words were utilized to develop the cloze task.

Furthermore, another set of words, five semantically related to diseases (the topic of the composition task) along with five unrelated words were allocated to the composition task. It is worth noting that general words were supplied to provide the composition group with an opportunity to develop their ideas more easily (Al-Hadlaq, 2003) and also to moderate the difficulty which might be created by the semantic similarity of new words. Nation (2001) maintained that learning a large set of words which bear close semantic similarities can be puzzling for language learners.

Tasks

A cloze task entailing ten gaps which needed to be filled by ten of the target words was utilized. This task was followed by five comprehension questions to extend the time on task to be similar to that of the composition

task. After its initial drafting by the researchers, the cloze task was given to three EFL teachers to evaluate the appropriacy of the cloze concerning vocabulary, structure, content, and difficulty level for the participants. Based on their suggestions and comments, the task was modified and later piloted with a similar group of students (n = 16) to the participants in the main study regarding language proficiency, age, and gender. Then in light of their performance, some modifications were made to the language of the cloze.

Additionally, a glossary was developed for the target words which included L1 and L2 translations of the target words, supporting examples, and one synonym for each of them. The involvement load for the cloze is three since learners were required to look up the target words and use them in a given context (i.e., moderate need= 1, search= 1, and moderate evaluation= 1).

For the second task, the participants were asked to write a composition on the familiar topic of *the effects of a dangerous disease on society* (see Appendix A). They were required to use all ten target words within 30 minutes. The teacher's hands-on experience and the results obtained from the performance of the pilot group indicated that the task was at the level of the participants' language proficiency.

The same procedures were followed to prepare a glossary of the target words for the composition (see Appendix B). Given the fact that writing a composition involves learners in generating new texts, its involvement index is 4 (i.e., moderate need= 1, search= 1, strong evaluation= 2).

Procedures

It took three sessions to carry out the current study. Before the instructional sessions, the learners were briefed on how to collaborate with their partners by adopting strategies such as questioning, pausing, commenting, repeating, confirming, and making use of the glossary. Then, the learners were asked to pair up with their self-selected partners and write a joint composition including the words given. It should be noted that each pair was given only one glossary. Afterward, in keeping with Swain and

Lapkin (2007), the T-R read the target words aloud to assist the participants to obtain a partial familiarity with the words and hence lessen the task difficulty. Next, each started writing a joint composition. In order to track the manifestation of evaluation processes, all the collaborative dialogues by each pair were recorded by a separate voice recorder. At the end of the session, the students' recorded files were collected and stored on the T-R's personal computer for later transcription and analysis.

In the third session, the participants were paired up with their partners once more and did the cloze task. To be more specific, initially, the T-R distributed the cloze and glossary papers and then explained the task instruction in Persian, mother tongue of the participants, to make sure that they had understood task objectives. Each pair worked collaboratively to complete one cloze task in 30 minutes; all the collaborative dialogues over completing the cloze were audio-recorded again and then collected and stored by the T-R.

Identification and Classification of LREs

Students' collaborative dialogues over both tasks were transcribed verbatim (see Appendix C for transcription symbols), reviewed, and analyzed to identify LREs and Cognitive Conflict Episodes (CCEs). It should be mentioned that only the LREs over the target words were examined in this study and categorized as either CCEs or non-Cognitive Conflict Episodes (henceforth non-CCEs) as proposed by Tocalli-Beller and Swain (2005).

To code the resolution of LREs and tally their frequency, the following measures were adopted. Initially, all the collaborative dialogues were transcribed verbatim by the T-R and reviewed several times. Then the first researcher and T-R categorized the LREs based on their resolution (i.e., CCEs and non-CCEs) individually. The inter-coder reliability for this codification calculated by Cronbach's Alpha was high ($\alpha= .95$). The coders then discussed the points of difference until they reached consensus.

Micro-genetic Analysis

Vygotsky (1978) developed a new research methodology to account for how cognition is socio-culturally mediated and co/constructed; his method attaches particular importance to the processes through which a new function comes into existence (Ellis & Barkhuizen, 2005; Lantolf & Thorne, 2006). Vygotsky's methodology constitutes four genetic domains of development, the smallest of which is conceived as micro-genesis during which "how development takes place throughout a particular interaction in a specific sociocultural setting" is traced and examined (Ellis, 2008, p. 521). To directly borrow from Vygotsky (1978),

> Any psychological process, whether the development of thought or voluntary behavior, is a process changing right before one's eyes. The development in question can be limited to a few seconds or even fractions of seconds …Under certain conditions, it becomes possible to trace this development. (p. 61)

Similarly, Gutiérrez (2008) held that the focus of micro-genetic analysis was on how the overt examples of learning occur, and this could be traced in learners' collaborative dialogues "over a relatively short span of time (for example…learning a word, sound, or grammatical features of a language)" (Lantolf, 2000, p. 3). To Robbins (2001), "the microgenetic domain focuses on the overt, in flight, an instance of learning as it happens during interpsychological activity" (as cited in Gutiérrez, 2008, p. 2).

By moment-to-moment analysis of learners' collaborative dialogues from Vygotskian SCT perspective, this study intended to track how task type could affect the underlying processes and activities which generate varying degrees of evaluation, and consequently task-induced involvement. In a word, this study intended to document and "grasp the process" of evaluation "in flight" (Vygotsky, 1978, p. 68). The transcriptions which were coded as CCEs and non-CCEs for both tasks were scrutinized to spot the manifestation of evaluation. Some of the LREs are micro-genetically analyzed here to present in flight manifestations of evaluation.

Results

CCEs in the Composition Task

To portray a picture of how evaluation materializes in two different word-focused tasks, micro-genetic analysis of 6 LREs is presented below.

The first episode is taken from a pair's collaborative dialogue during joint composition writing. The students are attempting to extend their writing by using one of the given words, *'curfew*.'

Episode 1: *A CCE provoked in the composition task*

1 S1: …now we should say this illness is their curfew.

2 S2: ↑Mozhgan! This illness IS their curfew::, is it ↑ right?

3 S1: I'm not sure about this.

4 S2: Do you think it is grammatical?

5 S1: *Man'e raft o amadeshoon (.) in bimari hastesh ke baes man'e raft o amadeshoon shode.*

     Prohibition of their coming indoors and going outdoors (.) it is the illness which has prohibited them from coming indoors and going outdoors.

6 S2: Let's check the glossary.

7 S1: [reading out from the glossary] A law that forces people to stay indoors after a particular point at night, or the time people must stay indoors, time limit, *man'e raft o amad* [the Persian translation of curfew]. Look at the example (3 seconds)

8 S2 Caused!

9 S1: Caused them a curfew, right?↑

10 S2: eh:: um:: (8 seconds) OK! caused them a curfew

11 S1: OK!

12 S2: OK! What is the next word?

The opening line of this CCE indicates that S1 aims to proceed with the writing by uttering a sentence containing '*curfew*', but S2 does not seem satisfied with the use of *curfew* in this context. Thus, she tries to attract her partner's attention by calling her first name in a rising intonation; afterward, she draws S2's attention to the problematic part by stressing 'IS' and pronouncing it in a louder voice than natural speech. She thinks that it

is not grammatically appropriate to use 'is' with 'curfew'. When S2 expresses some reservations about the accuracy of her contribution, as noticed in line (4), S1 voices that she has some doubts about grammatical soundness of the sentence. Seemingly, this conflict of ideas adds to task difficulty, so S1 resorts to their L1 to elucidate the point of difference. Afterward, to resolve the existent cognitive conflict, both learners agree to look up the target word in the glossary. Line (7) indicates that they check all information in the entry for *curfew*. After consulting the glossary, S2 proposes collocating 'curfew' with 'cause', line (8), but she still has some hesitations if 'cause' fits the context. Having recognized S1's doubts, S2 strives to assist her by complementing her previous turn and seeking her agreement by saying '…right?↑' in a rising intonation. However, uttering 'eh:: um::' (line 10) illustrates that S2 is yet to be fully convinced; hence, she takes more time to reflect on the sentence. After an eight-second reflection, she verifies the accuracy of the sentence by uttering an acknowledgment token ('OK!') and repeating the segment they were dubious about.

This extract from the students' collaborative dialogues highlights that the demand for using the word given in a self-generated context is strong enough to urge the students to weigh the target word against other words in its surrounding context and to attend to collocational and structural behaviors of the word in conjunction with its semantic properties. The demand creates some cognitive conflicts which, the students are stimulated to regard the definition, synonym, and examples for the target word included in the glossary to get them resolved.

The following extract taken from the collaborative dialogues of another pair during the composition task constitutes a vivid example of evaluation in making use of one of the target words, *'exasperating'*.

Episode 2: *A CCE provoked in the composition task*

13 S3: *That illness was very (.) very umm (.) (5 seconds) khob aziyat
           konandeh ro chetoor estefade konim?*
           That illness was very (.) very umm (5 seconds) so how can we
              use exasperating?

14 S4: *Aziyat koonandeh. Bezar checkesh konim.*
           Exasperating. Let me check it.

15 S3: Here here!

16 S4: Let me see (.) uha!

17 S3: Extremely annoying, frustrating [reading out from the
           glossary]

18 S4: *Aziyat konandeh adjective hast, chetor estefade mishe hala?*
           Exasperating is adjective, so how is it used?

19 S3: Yes it's adjective (.) umm (.) You have this exasperating habit
           of never looking at me! [Reading out the given an example in
           the glossary]

20 S4: Exasperating habit (.) exasperating habit, exasperating habit
           (.) yes adjective.

21 S3: It very (.) umm.

22 S4: *Are haminjori dorosteh.*
           Yes, this way is correct.

23 S3: *Chi shodesh?*
           What was it?

24 S4: /Ig/ (.) /Ig/ (.) /Ig zæs / (.) eh / ɪɡˈzæspəreɪtɪŋ/ OK? [She tries
           to pronounce the phonetic symbols from the glossary].

25 S3: Yeh / ɪɡˈzæspəreɪtɪŋ/.

26 S4: Exasperating people so?

27 S3: It exasperating people.

28 S4: It EXASPERATING ↑people?

29 S3: It very exasperating for people.

30 S4: *Ama exasperating SEFAT hastash*
           However, exasperating is ADJECTIVE

31   S3:   *Khob, midonam.*
          OK! I know.

32   S4:   *Ama sefat ba'd az fe'l e rabti ya ghabl az esm miad.*
          However, adjectives come after linking words or before nouns.

33   S3:   Exasperating (.) aha was.

34   S4:   It was very exasperating for people, so again doctors started (.) *ba semajat chi mishe*? (how do you say *'with tenacity' in English?*)

The opening of the CCE shows that S3 hesitates in the use of '*exasperating'*. Also, since the pair does not know the word completely, they draw on the glossary as an external mediatory tool to jointly complete the unfinished sentence. Lines (16) to (23) demonstrate that they attend to different aspects of vocabulary knowledge; they check the meaning, pronunciation, and grammatical function of the word in order to use it in an original context; further, they study the examples given in the glossary to learn how the word is used in context. Line (28) portrays S4's attempts to raise and draw S3's attention to the written segment of the text by posing a question, but S3 fails to detect the problem. Therefore, she adds two other words to the sentence to complete it. When, as seen in line (30), S4 realizes that S3 has not grasped the point, she resorts to explicit metalinguistic explanation in their L1.

Nonetheless, S3 is still unable to understand what her partner intends to communicate, so S4 continues to explain the function of adjectives explicitly. Expressing 'aha' (Line 33) shows that S3 finally identifies the source of the problem and then removes it with the help of her partner. This episode offers further evidence corroborating that using one of the words given in an appropriate self-generated context encourages the students to weigh the word given against other words in its surroundings to see whether it fits the context. The demand for using the word in a self-generated context prompts them to pay attention to various aspects of vocabulary knowledge. Furthermore, since the learners are stimulated to

weigh their ideas against each other in order to resolve the cognitive conflict, it can be argued that evaluation value is enhanced.

To summarize, it should be argued that using words in a new context creates a suitable opportunity for attending to a higher number of associations between form and meaning. That is, apart from semantic features, the students were prompted to pay attention to other features of the target words such as collocational behavior, syntactic qualities, and functional properties. The episodes illustrate that deciding how to combine new words with other words to generate a well-formed and appropriate context results in more conflicts, and hence more evaluations. These episodes provide evidence for Laufer and Hulstijn's (2001) argument that using words in an original context induces robust evaluation.

CCEs in the Cloze Task

To compare the effect of task type on evaluation, two CCEs produced during the cloze task are micro genetically analyzed. Episode 3 is taken from S1 and S2's collaborative dialogues during the cloze task to fill in the gaps in its first line.

Episode 3: *A CCE provoked in the cloze task*

1 S1: Nowadays crime is an eh eh (.) (3 seconds) problem uh in many countries.

2 S2: A (2 seconds) problem.

3 S1: Frigid problem, OK::?

4 S2: *Fekr konam moshkel e SARD doros nis.*
        I think FRIGID problem is not correct.

5 S1: *Pervasive problem, ↑OK?*

6 S2: *Moshkel e faragir... OK!*
        Pervasive problem...OK!

As illustrated in the opening line, S1 is reading out a part of the cloze. When reaching the blank in the text, she hesitates and pauses for about three seconds to find the appropriate word. Meanwhile, S2 utters the words around the gap (Line 2), which can be indicative of the fact that she is mainly considering the next words around the gap to fill it. Afterward, S1

proposes '*frigid*' to fill the gap, but S2 translates this word into their L1 and states that the combination of 'frigid' and the next word, i.e., 'problem' does not make sense, so she utters '*frigid'* with a higher tone of voice in order to underscore the source of problem. This expression of dissent, in fact, creates a cognitive conflict. The conflict encourages S1 to reconsider her original proposal and suggest a new word with a rising intonation (Line 5). By translating S1's contribution into L1, S2 examines its contextual appropriacy in this other-generated context. Line (5) substantiates that they are mainly evaluating the intended word concerning semantic features. The provoked cognitive conflict is resolved by considering the meaning of the words given and the students are not stimulated to pay attention to other dimensions of vocabulary knowledge such as syntactic and collocational behaviors.

Moreover, this episode depicts since the students have to fill in the blank in an already available context, they fail to make a connection between the target word and its broader surrounding context; therefore, in keeping with Laufer and Hulstijn (2001), it can be argued that the amount of generated evaluation is moderate. More supporting evidence for this stance is given in the following episode, where two of the students are trying to fill in another blank in the cloze. This episode also indicates that the cloze task principally stimulates learners to translate the next words surrounding the blank to decide on the correct choice.

Episode 4: *A CCE provoked in the cloze task*

7 S3: He then went to the drawer where the money was kept and (2 seconds) uh [the blank in the cloze]

8 S4: And um (.) perpetrate.

9 S3: *Um (.) no (.) um in ja serghat kardan* doroste.

10 S4: *Are serghat kard pool ha ro* [translating the rest of the sentence after the blank]

    Yes, purloined all the money

11 S3: Here is purloin.

When S3 reaches the blank in the text, she takes two seconds to reflect upon it; meanwhile, S4 proposes one of the words given, '*perpetrate*', but S3 disagrees with S4's suggestion. Then, to express the rationale behind her dissent, she shifts to their L1 and finally suggests an appropriate word for the blank. As a result, S3 adopts the same strategy and translates the next words after the blank to make sure that the selected word fits the context. As observed, the learners mainly weigh two of the possible words (i.e., 'purloin' and 'perpetrate') against each other to see whether it fits the gap regarding meaning. This is because the rest of the sentence is given to the learners, so they are not driven to examine whether the chosen word matches the context regarding other dimensions of vocabulary knowledge like collocational patterns.

To reiterate, it was found that the micro-analysis of in-flight evaluation processes in the cloze task lends support to Laufer and Hulstijn's (2001) assumption that evaluation is moderate in cloze task as students are required to weigh the words given against each other to select the correct word. It is illustrated that the students are not pushed to evaluate intended words against other words in their surroundings nor are they required to examine their suitability in terms grammatical function or collocational behavior.

Non-CCEs in Both Tasks

The analysis of non-CCEs can provide a clear picture of the underlying processes of evaluation in both tasks. This episode is taken from the students' joint work on the composition task where they intend to use the target words *'tenacity'* and *'gawky'*.

Episode 5: *A non-CCE provoked in the composition task*

1   S5:   *man ba semajat be madaram goftam ke man dast o pa cholofti nistam.*

I said to my mother with the tenacity that I'm not gawky.

2   S6:   I said tenacity I'm not gawky.

3   S5:   OK.

Initially, S5 resorts to her L1 to determine the content of what they are writing about; afterward, S6 translates the intended sentence into English while S5 only complies with it. S6 confirms her partners' contribution instantly without reflecting enough on its appropriacy. That is, they do not discuss different aspects of word knowledge to make associations between form and meaning nor do they attempt to link the word to its surrounding words; as a consequence, it can be argued that the amount of induced evaluation becomes lower than that generated in the case of CCEs.

In the following episode, two students are attempting to fill in the gap in the first line of the cloze task.

Episode 6: *A non-CCE provoked in the cloze task*

4   S3:   On a freezing night a um (.) [blank in the task]

5   S4:   A (.) man

6   S3:   *Ghavi heikal, ghavi heikal* [Persian equivalent for 'hefty,' one of the given words]

7   S4:   Yes, here [pointing to the glossary] hefty.

8   S3:   OK! Hefty man entered the jewelry store (.)

As observed, when S3 gets to the blank, she takes some time to think. Line (5) indicates that S4 is considering the blank as well. Line (6) portrays that while S3 resorts to their L1 to make sense of the context and determine the meaning of the sentence, S4 offers the English equivalent for the Persian word proposed by S3. Since they share the same opinion about the correctness of this word, they do not pay attention to other aspects of the word apart from its meaning. Hence, the amount of induced evaluation declines.

The total number of the LREs and the way they were resolved for each task type is presented in Table 1.

Table 1.

Distribution of the lexical LREs across cloze and composition tasks

	CCEs	Non-CCEs	Total
Cloze	36 (30%)	84 (70%)	120
Composition	72 (60%)	48 (40%)	120

It was found that the composition task was twice as useful as the cloze task in encouraging the students to turn the LREs over the words given into CCEs. More specifically, 60% of the lexical LREs over the target words for the composition were turned into CCEs, while only 30% of them for the cloze task was changed into CCEs. This was in correspondence with evaluation degrees allocated to these two tasks by Laufer and Hulstijn (2001).

Discussion

This study intended to examine how task type could affect the actual manifestation of evaluation induced by two word-focused tasks, which were different concerning assumed evaluation indices. Micro-genetic analysis of the LREs indicated that using the target words in a new context could more strongly urge the students to evaluate the words given against each other and against preceding and following words in their contexts. This supported Laufer and Hulstijn's (2001) claim that writing words in a new context not only entails making decisions about the words given but also requires "additional syntagmatic decisions about the precise collocations of the word which the learner is trying to use" (Laufer & Hulstijn, 2001, p. 15).

What is more, micro-genetic analysis of the LREs which occurred in the composition task also indicated that deciding on collocating the words given with other words (i.e., their syntagmatic relations with other words) provoked a large number of the cognitive conflicts, which can, in turn, stimulate the students to notice a higher number of form-meaning associations, and consequently a higher level of evaluation could be generated. On the contrary, micro-genetic analysis of the LREs stimulated

by the cloze task illustrated that the students were not urged strongly enough to evaluate a word given with its preceding and following words regarding syntagmatic and collocational properties. That is, to decide on an appropriate word to fill a gap, the students primarily took semantic properties of the words given into account. For this reason, it stands to reason to assume that the level of evaluation for the cloze task decreased. In a similar vein, Laufer and Hulstijn (2001) maintained that the richness of associations between form and different aspects of word knowledge could lead to stronger involvement load. This has also been supported by Maftoon and Sharifi (2012), where they indicated that output-oriented tasks could push learners to notice a wider web of form-meaning associations because the learners have to encode their concepts.

Additionally, comparing the CCEs and the non-CCEs suggested that the former could bring about a higher degree of evaluation since resolving CCEs can prompt the students to weigh the words given against other words in their context whereas the students mostly acknowledged and confirmed their partners' proposals in the case of non-CCEs. Therefore, the researchers tend to regard the higher frequency of CCEs as a reliable indicator of stronger evaluation. In line with Tocalli-Beller and Swain (2005), it can be argued that engagement in CCEs prompted the students to reflect upon the points of conflict, and as a result, this deeper reflection could contribute to deeper understanding. In other words, conflict-solving draws and maintains students' attention and motivation to "re-examine and clarify their thoughts" (Tocalli-Beller & Swain, 2005, p. 22) by adopting strategies like offering suggestions and counter-suggestions, posing questions and offering solutions, and seeking for clarification. The students used their language as a mediatory tool to challenge and complement their partners' views and understanding. Moreover, they expressed and exposed their thoughts to critical evaluation in light of common cognitive and linguistic resources. Thus, it can be argued that participation in more different tasks entails the process of self- and other-evaluation which in turn contributes to an increase in the depth of evaluation.

The CCEs in the composition task indicated that the dyads were stimulated to consider a target word from various perspectives. Different dimensions of vocabulary knowledge were assessed in a joint undertaking to use the word in an appropriate context which was in tune with Tocalli-Beller's (2003) argument that conflict allows learners to look at issues from different angles and evaluate them with more precision. Likewise, Pathinathan (2012) noted that cognitive or 'substantive' conflicts during collaborative writing could encourage group members "to consider alternative word choices and supporting details" while thinking and debating more extensively (p. 16). As viewed in the analyzed CCEs, the composition encouraged the students to pay more attention to various relationships between form and meaning, which is a critical component of deeper cognitive processing and vocabulary learning (Laufer & Hulstijn, 2001). Hence, both the quantity and quality of evaluation are enhanced by involving students in tasks requiring them to use the words given in a self-generated context.

It should be noted that engaging in CCEs may also generate, strengthen and maintain students' motivation to evaluate differing opinions and alternatives more effectively (Tocalli-Beller, 2003); to borrow from Yong (2010), "Conflict also plays a crucial role in problem-solving processes. It provides a broader understanding of the problem or issue. Students generate alternative ideas, which in turn, maintain their interest and participation" (p. 23). Based on this line of argument, it can be claimed that engaging in CCEs could have stimulated the students to turn the obvious need for searching various aspects of the words given into a strong internal need. In other words, although the students in this study were initially asked by their teacher to look up the words in the glossary, the CCEs could have deepened and converted this externally-driven need into an internal need for searching and using the words. As a result, the levels of evaluation (i.e., by considering further possible alternatives) and need (i.e., by converting externally to internally driven need) were raised. However, the active status of need in word-focused tasks and vocabulary learning warrants further investigation.

Another quality of CCEs which could affect task-induced involvement is explainable by referring to various levels of noticing and attention. Qi and Lapkin (2001) identified two levels of noticing: 'perfunctory' and 'substantive.' Perfunctory noticing is defined as where the students merely notice a discrepancy between their original writing and a reformulated version by a native speaker without managing to express the reasons underlying the difference. On the contrary, the substantive noticing is taken as the situations where the students not only notice a difference in the writings but also articulate the reasons behind it. Similarly, studying the quality of LREs in collaborative writing task, Storch (2008) distinguished two levels of attention: 'elaborate' and 'simple'. The former is realized where students reflect upon possible alternatives, critique their partners' suggestions, and offer counter-suggestions, whereas the latter refers to where one student makes a proposal and the other repeats, acknowledges, or do not respond to the proposal (Storch, 2011). Storch (2008) maintained that simple attention or engagement "leads to shallow processing since it "may be quite mechanical with little attention invested in the act" (p. 110). This position is in keeping with Hulstijn and Laufer's claim that "Rich (qualitative) and numerous (quantitative) associations with existing knowledge ,(e.g., in the form of establishing similarities and contrasts between old and new information)" enhances task-induced involvement (2001, p. 541). Likewise, some recent studies have accentuated that tasks which require productive generations can help better prompt learners to establish a larger number of strong links between form and meaning of a word (Hu & Nassaji, 2016). As demonstrated, the CCEs, especially those in the composition task, entailed the students' attempts to evaluate suggested words against the other words in the context and examine their suitability in terms of meaning, syntax, and function; also, the CCEs prompted the students to articulate the reasons underlying the appropriacy of their choices and offer argument and counter-argument. As a result, the CCEs increased the level of cognitive processing and consequently task-induced involvement.

Conclusion

This study examined the degree of evaluation in two different task types by adopting the tenets and methodology of Vygotskian SCT. Micro-genetic analysis of the students' collaborative dialogues indicated that task type could affect the way evaluation unfolds. It was found that using the target words in a 'self-generated context' (i.e., the composition) could encourage the students to weigh the words given against each other and other words in their surrounding context to produce an appropriate piece of discourse. Moreover, the students were encouraged to pay attention to a larger number of form-meaning associations.

In contrast, filling the gaps in an 'other-generated context' (i.e., the cloze) prompted the students to mostly take the semantic properties of the words given into account and overlook their syntagmatic relations with other words surrounding them. It was also indicated that using the target words in an original context like the composition task led to producing a more significant number of CCEs than the cloze. Hence, given the fact that the level of evaluation in CCEs is higher than non-CCEs, the composition task could induce a higher degree of evaluation than the cloze. The materialized in-flight instances of evaluation in both tasks which were captured by the micro-genetic analysis supported the evaluation values proposed by Laufer and Hulstijn (2001).

In light of the findings, it is suggested that although the basic principles of cognitive and sociocultural accounts of second language learning seem incommensurable (Foster & Ohta, 2005), they can afford us to present a better picture of task-induced involvement. The richness of the results hints that second language researchers should benefit from all theories at their disposal rather than restrict themselves to one single theory or methodology of research. This resonates with Lantolf's (1996) position that 'letting all flowers bloom' is of critical importance to developing theories in second language acquisition. Teachers are also recommended to benefit from tasks which induce a larger number of cognitive conflicts because they can prompt students to pay attention to more dimensions of vocabulary knowledge.

Finally, future studies are suggested to employ other data collection methods such as think-aloud protocols to explore how evaluation for different word-focused tasks is materialized when tasks are performed individually. Investigating whether the quantity and quality of CCEs can affect vocabulary learning also warrants further attention.

References

Al-Hadlaq, M. S. (2003). *Retention of words learned incidentally by Saudi EFL learners through working on vocabulary learning tasks constructed to activate varying depths of processing.* Unpublished PhD dissertation. Indiana: Ball State University.

Ellis, R. (2008). *The study of second language acquisition* (2nd Ed.). Oxford: Oxford University Press.

Ellis, R., & Barkhuizen, G. P. (2005). *Analyzing learner language.* Oxford: Oxford University Press.

Folse, K. S. (2006). The effect of type of written exercise on l2 vocabulary retention. *TESOL Quarterly, 40(*2), 273-293.

Foster, P., & Ohta, A. S. (2005). Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics*, *26*(3), 402 -430.

Gutiérrez, A. G. (2008). Micro-genesis, method, and object: A study of collaborative activity in Spanish as a foreign language classroom. *Applied Linguistics, 29(1), 120-148.*

Hu, H. & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System, 56*, 28-39.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning*, *51*(3), 539–558.

Keating, G. (2008). Task effectiveness and word learning in a second language: the involvement load hypothesis on trial. *Language Teaching Research, 12* (3), 365-386.

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 58*, 285–325.

Lantolf, J. P. (1996). SLA theory building: "Letting all the flowers bloom!" *Language Learning, 46*, 713–749

Lantolf, J. P. (2000). *Sociocultural theory and second language learning* (1st Ed.). Oxford University Press, USA.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

Lantolf, J. P., & Thorne. S. L. (2007). Sociocultural theory and second language acquisition. In B. van Patten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 201-224). Mahwah, NJ: Erlbaum.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26.

Maftoon, P. & M. Sharifi Haratmeh. (2012). The relative effectiveness of input and output-oriented tasks with different involvement loads on the receptive and productive vocabulary knowledge of Iranian EFL learners. *The Journal of Teaching Language Skills (JTLS)*, 4 (2), 27-52.

Nation, I. S. P. (2001). *Learning Vocabulary in another Language (1 st Ed.)*. Cambridge University Press.

Ohta, A. S. (2001). *Second language acquisition processes in the classroom learning Japanese.* Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Pathinathan, S. (2012). Intragroup Conflicts during Collaborative Writing in an ESL/EFL Preparatory Programme. Inter*national Journal of Applied Linguistics & English Literature, 1*(7), 8-18.

Qi, D., & Lapkin, S. (2001) Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10 (4)*, 277–303.

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

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics, 31*, 275–288.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Eds.) *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford 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). London: Continuum.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, *82*(3), 320-337.

Swain, M., & Lapkin, S. (2007). The distributed nature of second language learning: Neil's perspective. In S. Fotos and H. Nassaji (Eds.) *Form-focused instruction and teacher education: Studies in honor of Rod Ellis (*pp. 73-86). Oxford: Oxford University Press.

Tocalli-Beller, A. (2003). Cognitive conflict, disagreement, and repetition in collaborative groups: Affective and social dimensions from an insider's perspective. *The Canadian Modern Language Review, 60*(2), 143–71.

Tocalli-Beller, A., & Swain, M. (2005). Reformulation: The cognitive conflict and L2 learning it generates — *International Journal of Applied Linguistics*, *15*(1), 5–28.

Vygotsky, L. S. (1978). *Mind in Society: The development of higher psychological processes* (14th Ed.). Harvard University Press.

Wesche, M., & Paribakht, T. S. (1996). Assessing second Language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*(1), 13–40.

Yaqubi, B., Rayati, R. A., & Allemzade Gorgi, N. (2012). The involvement load hypothesis and vocabulary learning: The effect of task types and involvement index on L2 vocabulary acquisition. *The Journal of Teaching Language Skills, 2*(1), 145–163.

Yong M. F. (2010). Collaborative writing features. *RELC Journal, 41*(1), 18-30.

Appendix A. Instructions of the composition task

Write a composition on the effects of a disease on people of society; for example, you can write about causes that have spread it and what measures governments and people can take to control and cure it. (***You must use all ten words in the table below).

*** Notice that you can refer to the attached glossary whenever needed.

> *Pernicious- appease- palpitate- exasperating- purvey- gawky-*
> *perpetual- herbal-tenacity - curfew*

Appendix B. A sample of the glossaries used in the cloze and composition tasks

exasperating / ɪgˈzɑːspəreɪtɪŋ $ -ˈzæs- / adjective
extremely annoying, **frustrating** (اذیت کننده)
EX. You have this exasperating habit of never looking at me!

Appendix C. Transcriptions conventions used in the current study are as follows: (*Adapted from Ellis and Barkhuizen, 2005*)

(2 seconds)	Numbers enclosed in parentheses accompanied with seconds represent a pause and its duration concerning seconds.
[ ]	Brackets indicate additional information or explanations given by the researchers.
↑	Up arrow is used to indicate that there is sharply rising intonation
CAPS	Capital letters indicate that the speaker pronounces the capitalized section at a higher volume than the speaker's average volume.
(.)	Pauses of less than one second