THE ROLE OF VOCABULARY LEARNING CONDITIONS IN WORD LEARNING AND RETENTION

O PAPEL DAS CONDIÇÕES DE APRENDIZAGEM VOCABULÁRIA NA APRENDIZAGEM E RETENÇÃO DE PALAVRAS

EL PAPEL DE LAS CONDICIONES DE APRENDIZAJE VOCABULARIO EN EL APRENDIZAJE Y LA RETENCIÓN DE PALABRAS

Seyyed Nasser Mousavi
Islamic Azad University – Iran

Nasser Ghafoori
Islamic Azad University – Iran

Mahnaz Saeidi
Islamic Azad University – Iran

Abstract: The present study is an attempt to examine the learning processes of noticing, retrieval, and generating and their possible contribution to the process of vocabulary learning and retention among intermediate students. One hundred and twenty intermediate students were randomly assigned into four groups, namely Noticing through Input enhancement (n=30), Input Enhancement plus Input-Based Reviewing (n=30), Input Enhancement plus Output-Based Reviewing (n=30) and Input Enhancement plus Input-Based and Output-Based Reviewing. The Academic Words contextualized in Focus on Vocabulary 2: Mastering the Academic Word List (Schmitt, Schmitt, & Mann, 2011) were the target words of the study. A pretest composed of VLT items was administered to the participants. The first group encountered the target words that have been already highlighted to absorb their attention. Encountering the already highlighted words, the second group reviewed the words through researcher-made word cards. The third group, besides encountering the already highlighted words, reviewed the words through rewriting the sentences including the target words. The fourth group experienced noticing through input enhancement; retrieval through using researcher-made word cards; and generation through rewriting the sentences containing the unknown words. One week after the last treatment session, an immediate posttest, and after two weeks, a delayed posttest were administered. Based on the results of four one-way repeated measures ANOVAs and three one-way ANOVAs, it was revealed that all types of input-based, output-based and input+output-based reviewing have positive effect on vocabulary learning. However, their positive effect on vocabulary retention was fairly vague. Moreover, the group treated through input enhancement+input- and output-based reviewing outperformed the other groups.

Keywords: Generation; Noticing; Retrieval; Vocabulary Learning; Vocabulary Retention

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

404

Resumen: El presente estudio es un intento de examinar los procesos psicológicos de aviso, recuperación y generación y su posible contribución al proceso de aprendizaje y retención de vocabulario entre los estudiantes intermedios. El método de investigación fue experimental. Ciento veinte estudiantes intermedios fueron asignados aleatoriamente en cuatro grupos, a saber, mediante Mejora de entrada (n = 30), Mejora de entrada más Revisión basada en entrada (n = 30), Mejora de entrada más Revisión basada en salida (n = 30) y Mejora de entrada más revisión basada en entrada y en salida. Las palabras académicas contextualizadas en Focus on Vocabulary 2: Mastering the Academic Word List (Schmitt, Schmitt y Mann, 2011) fueron las palabras clave del estudio. Se administró a los participantes una prueba preliminar compuesta de ítems VLT. El primer grupo encontró las palabras objetivo que ya se han resaltado para atraer su atención. Al encontrar las palabras ya resaltadas, el segundo grupo revisó las palabras a través de tarjetas de palabras hechas por investigadores. El tercer grupo, además de encontrar las palabras ya resaltadas, revisó las palabras reescribiendo las oraciones, incluidas las palabras objetivo. El cuarto grupo experimentó darse cuenta a través de la mejora de entrada; recuperación mediante el uso de tarjetas de palabras hechas por investigadores; y generación reescribiendo las oraciones que contienen las palabras desconocidas. Una semana después de la última sesión de tratamiento, se realizó una prueba posterior inmediata, y después de dos semanas, se administró una prueba posterior retrasada. Sobre la base de los resultados de cuatro ANOVA de medidas repetidas unidireccionales y tres ANOVA unidireccionales, se reveló que todos los tipos de revisión basada en insumos, resultados e insumos tienen un efecto positivo en el aprendizaje de vocabulario. Sin embargo, su efecto positivo en la retención de vocabulario fue bastante vago. Por otra parte, el grupo tratado a través de la mejora de entrada basada en la revisión basada en entrada y salida superó a los otros grupos.

Palabras clave: Generación; Darse cuenta; Recuperación; Aprendizaje de vocabulario; Retención de vocabulario

Resumo: O presente estudo é uma tentativa de examinar os processos psicológicos de observação, recuperação e geração e sua possível contribuição para o processo de aprendizado e retenção de vocabulário entre estudantes intermediários. O método de pesquisa foi experimental. Cento e vinte alunos intermediários foram divididos aleatoriamente em quatro grupos: Nota através do aprimoramento de entrada (n = 30), aprimoramento de entrada mais revisão baseada em entrada (n = 30), aprimoramento de entrada mais revisão baseada em saída (n = 30) e Aprimoramento de entrada mais revisão baseada em entrada e baseada em saída. As Palavras Acadêmicas contextualizadas em Foco no Vocabulário 2: Dominando a Lista de Palavras Acadêmicas (Schmitt, Schmitt, & Mann, 2011) foram as palavras-alvo do estudo. Um pré-teste composto por itens do VLT foi administrado aos participantes. O primeiro grupo encontrou as palavras-alvo que já foram destacadas para absorver sua atenção. Encontrando as palavras já destacadas, o segundo grupo revisou as palavras por meio de cartões de palavras feitos pelo pesquisador. O terceiro grupo, além de encontrar as palavras já destacadas, revisou as palavras reescrevendo as frases, incluindo as palavras-alvo. O quarto grupo experimentou perceber através do aprimoramento das entradas; recuperação através do uso de cartões de texto feitos pelo pesquisador; e geração através da reescrita das frases que contêm as palavras desconhecidas. Uma semana após a última sessão de tratamento, um pós-teste imediato e após duas semanas, um pós-teste tardio foi administrado. Com base nos resultados de quatro ANOVAs de medidas repetidas unidirecionais e três ANOVAs unidirecionais, foi revelado que todos os tipos de revisões baseadas em entradas, baseadas em resultados e baseadas em resultados têm efeito positivo na aprendizagem de vocabulário. No entanto, seu efeito positivo na retenção de vocabulário foi bastante vago. Além disso, o grupo tratado por meio do aprimoramento das entradas e análises baseadas em entradas e saídas superou os outros grupos.

Palavras-chave: Geração; Notando; Recuperação; Aprendizagem de vocabulário; Retenção de Vocabulário

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

405

1. Introduction

The emphasis on the importance of input emerged in the 1960s as a reaction to traditional production-based teaching methodology (Shintani, 2012). The main claim was that comprehension practice should precede production practice. It was argued that "foreign-language learners-like first-language learners-need an extended period of receptive learning to comprehend the language they are learning before they begin producing in that language" (Gary & Gary, 1981, p. 332). Following Gary and Gary (1981), Krashen's Input Hypothesis (Krashen, 1982) was the first model which treated input as the main factor in L2 acquisition. According to the hypothesis, production serves only for generating comprehensible input and does not make a direct contribution to acquisition. Later on, in line with Krashen, Long (1983) proposed the Interaction Hypothesis based on which interactionally modified input as well as simplified input plays an important role in facilitating L2 acquisition. Nowadays, too, there is a broad consensus among L2 researchers that input is the key to developing L2 knowledge. For almost all L2 researchers (e.g., Gass, 1997; Mackey, 2012; Mitchell, Myles, & Marsden, 2013) the concept of input is perhaps the most important concept of second language acquisition and, as Gass and Mackey (2015) consider it as the "sine qua non of acquisition" (p.181), no second language learning is supposed to happen without input of some sort. As a matter of fact, no L2 acquisition theory has ever denied the importance of learners' access to input and has addressed the role of input in one way or another.

Following Krashen's (1985) strong claim that "if input is understood and there is enough of it, the necessary grammar is automatically provided" (p. 2), through the weaker version of the Noticing Hypothesis, Schmidt (2001) claimed that "people learn about the things they attend to and do not learn much about the things they do not attend to" (p. 30). Given the importance of noticing, the key question is: How can learners be helped to notice the target linguistic forms in the input? Sharwood-Smith (1993) wondered why L2 learners appear to ignore a vast mass of evidence and continue to operate with a system that has glaring inconsistencies with the target norms in the input. For him the answer was multi-faceted. First, L2 learners lack ample sensitivity to the linguistic features of the target structures in the input. So, even with a large amount of target linguistic forms, they might not grasp the forms in focus. Second, some linguistic features are inherently non-salient and for learners they might go unnoticed in spite of their abundant availability. Finally, learners' first language may act as an obstacle to notice certain linguistic features in the input. To remove these obstacles, Sharwood-Smith (1991, 1993) hypothesized that we need to stimulate input processing for language learning in general and for form as well as meaning in particular and this stimulation can be carried out through improving the quality of input. Therefore, through reanalyzing the notion of consciousness-raising in language learning, Sharwood-Smith (1991) introduced 'input enhancement' as an operation to augment the saliency of linguistic features, "the process by which language input becomes salient to learners" (Sharwood Smith, 1991, p. 118). Engineered by a third party, say, a researcher or a teacher, via typographic) for written input) or phonologic (for oral input) means, input enhancement seeks only to heighten the chances of detection or noticing (Robinson, 2013). Different external input manipulations have been developed to enhance the salience of input, including manipulation of frequency (input flood), visual salience (typographical or textual manipulation), and corrective feedback in discourse (e.g., repetition or recast). Textual enhancement, an implicit input manipulation technique employed in the current study, uses visual enhancement methods such as underlining, bold facing, color-coding, italicizing, CAPITALIZING or using different fonts.

Based on her experience with long-term French immersion students in Canada, Swain (1985) proposed Output Hypothesis through which she claimed that the provision of 'comprehensible input' (Krashen, 1985) alone is not enough for acquisition to take place. She

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

406

believed that, it is through the output production that learners notice the gap or the hole between their interlanguage and the target language. Put differently, "by attempting to produce output, learners are forced into noticing what they do not know, or what they know only partially" (Russell, 2014, p. 26).

In the realm of cognitive linguistics, language is conceived of as being constructed by the learner from individual instances of contextualized language input (e.g., Tomasello, 2003). As the learner meets more instances of contextualized language, he or she starts to construct new and creative structures generalizable to other contexts of use. In a similar vein, cognitive psychology deals with the mental processes and conditions involved in language learning, understanding, and production (Scovel, 1998).When it comes to vocabulary learning, too, learning depends on mental conditions. So, it is not what happens in the mouth or in the hands. What matters is what happens within the head. Having this cognitive orientation in mind, Nation (2001) has put forward three important general processes that may lead to a word being learned, namely noticing, retrieving, and generating. According to Nation (2001) the first proposition refers to the fact that learners need to notice the word, and be aware of it as a useful language item. Several factors were suggested by Nation to affect the quality of noticing: the salience of the word in the textual input or in the discussion of the text, previous contact that the learners have had with the word, and learners' realization that the word fills a gap in their knowledge of the language. Once a word is noticed, the memory of that word will be strengthened if the word is subsequently retrieved during a task, i.e., the second major process to learn vocabulary. The third major process that may lead to a word being remembered is generation or generative use of the words in focus. "Generative processing occurs when previously met words are subsequently met or used in ways that differ from the previous meeting with the word" (Nation, 2001, p. 105).

Almost no study, up until now, has been launched to compare the three processes of vocabulary learning with each other simultaneously. That is, studies in the domain of vocabulary learning and instruction have mostly focused on one of the mentioned conditions at a time. Some of them (Alanen, 1995; Lee & Benati, 2009; Leow, 2001; Leow, Nuevo, & Tsai, 2003), for example, have attempted to measure the effect of textual input enhancement on noticing in vocabulary learning. The noticing function of Output Hypothesis (Swain, 1993,1995, 1998) has also attracted little attention in SLA research and as Russell (2014) claimed "thus far, only three studies have been found) Izumi, 2002; Izumi & Bigelow, 2000; Izumi, Bigelow, Fujiwara, & Fearnow, 1999) that have attempted to test the noticing function of output" (p. 28). In almost all mentioned studies retention of vocabulary has not been taken into account as a variable. Having reviewed some empirically related studies, the present study is an attempt to deeply examine the vocabulary processes of noticing, retrieval, and generation and their possible contribution to the process of vocabulary learning and retention.
Bearing the variables of the study in mind, the researchers put forward the following research questions:

1. Does input enhancement have any effect on vocabulary learning and retention of Iranian intermediate EFL learners?
2. Does input-enhancement combined with input-based reviewing have any effect on vocabulary learning and retention of Iranian intermediate EFL learners?
3. Does input-enhancement combined with output-based reviewing have any impact on vocabulary learning and retention of Iranian intermediate EFL learners?
4. Does input-enhancement combined with input-based and output-based reviewing have any impact on vocabulary learning and retention of Iranian intermediate EFL learners?
5. Does input-enhancement, input-enhancement combined with input-based reviewing, input-enhancement combined with output-based reviewing, and input-enhancement

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

407

combined with input-based and output-based reviewing have differential effects on vocabulary learning and retention of Iranian intermediate EFL learner?

2. Literature Review

2.1. Information processing: From input to output

When it comes to studying vocabulary learning/acquisition, examining the role of linguistic input and output looms large. Input refers to pieces of language to which the learner is exposed. It is based on such linguistic evidence that the language learner forms linguistic hypotheses. From a comprehension point of view, for ease of access, input is simplified at the lexical, phonological, and syntactic levels in order to facilitate the processing of linguistic features (Hatch, 1983). There are strong arguments in the literature for the role of input in L2 acquisition in general and vocabulary learning/ acquisition in particular and most studies argue that L2 development depends on the quality of the input to which the learner is exposed (Ellis & Collins, 2009). Based on the findings of the related studies, the linguistic features in the input cannot sufficiently explain learning/acquisition by themselves, however (Swain, 1995). There is also another side to input which includes learner-based variables as noticing, processing, storing, and production that is output.

The other side of the coin is output or produced stretches of language by which the learners can move from lexical to syntactic processing. In addition, it provides the language learners with the opportunity to test their hypotheses with regard to how syntactic forms function in L2 (Swain, 1985). Also, output can help the learners notice the gap in their interlanguage consequently it facilitates learning and makes the learners sensitive to upcoming patterns of the target language (Swain, 1995). Furthermore, output has been demonstrated to increase frequency and accuracy of grammatical structures (Toth, 2006; Izumi & Bigelow, 2000).

2.2. Input, noticing, input-enhancement and vocabulary learning

In the domain of SLA, Lee and VanPatten (2003, p. 26) put forward a metaphor about language input: "input is to language acquisition what gas is to a car… an engine needs gas to run; without gas, the car would not move an inch… likewise, input in language learning is what gets the 'engine' of acquisition going…without it, acquisition simply doesn't happen."

Second Language Acquisition (SLA) has attracted a considerable interest believing that holding learners' attention to the formal features of second language input is beneficial and essential for second language development. Sharwood-Smith (1993) stated that input is the 'potentially processible language data which are made available, by chance or by design, to the language learner' (p.167). Furthermore, Gass and Mackey (2015) defined language input as the language data a learner is exposed to. In order that second language acquisition occurs, the learners should necessarily attend the L2 input. Since late 1970's, input has been under investigation in second language acquisition (SLA) research. It is believed that input is a necessary and sufficient condition, while output is a facilitative, but not a necessary condition for L2 acquisition. In 1991, Sharwood-Smith suggested another term, which is '*input enhancement*', to discuss and highlight the role of grammar in second language acquisition. Additionally, the term "input enhancement" was suggested to demonstrate a learning strategy through which second language learners are exposed to the input for its certain features so that they become perceptually salient and noticeable. Reinders and Ellis (2009) defined input enhancement as input process in which the target structure forms are highly frequent. Generally, input enhancement can be closely related to be one of the focus on form approaches in second language acquisition. At first, input enhancement was considered to be like consciousness-raising and a type of focus on form technique. Nassaji and Fotos (2011) believe that input enhancement is used not only to raise learners' awareness of target forms but also to improve

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

408

learners' acquisition. If the learners are exposed to enhanced input repeatedly, the target vocabulary items are most likely to be noticed. As literature indicates, one type of input enhancement which is common in studies is textual enhancement (TE). According to Wong (2003), there are two features of textual enhancement including: Firstly, second language learners read texts to achieve propositional content. Secondly, learners' attention can be drawn through such typographical cues as **bolding**, <u>underlying</u> and *italicizing*. Additionally, evidence shows that typographical cues can also be beneficial in learning and perception of new information and forms.

As far as noticing is concerned, Schmidt (2001) put forward that "people learn about the things that they attend to and do not learn much about the things they do not attend to" (p.30). Nassaji and Fotos (2011) stated that textual enhancement is believed to be an "implicit and unobtrusive way" and keep learners' attention in targeted forms. Learner's attention is directed to the meaning of the text, and incidentally on form-meaning relationships (Ellis, 2008). In a contrary manner, Lee (2007) believed that target forms which are textually enhanced may distract learners' attention from meaning, thereby having a harmful effect on comprehension. Godfroid, Boers, and Housen (2013) suggested that even though the perception of external stimuli is considered to be an essential part of L2 learning, the main functions of noticing, attention and awareness in second language learning are controversial and not that clear. In the same manner, Nassaji and Fotos (2011) proposed that it is possible to notice the target forms without taking the meaning of the form into account.

Noticing, which can be done through input enhancement, is described as "the process of the learner picking out specific features of the target language input which she or he hears or reads, and paying conscious attention to them so that they can be fed into the learning process" (Cullen, 2012, p. 260).

At first, Schmidt (2001) believed that conscious attention is an essential part of second language acquisition meaning that conscious attention is a necessary process through which second language acquisition occurs. However, he has changed the view in such a way that noticing and conscious awareness are separated. He proposed that noticing is limited to "awareness at a very low level of abstraction" (Schmidt, 2001, p. 5). Therefore, in order to put Schmidt's and Tomlin and Villa's (1994) views in the same line, Robinson (as cited in Song, 2007) defined noticing as "detection plus rehearsal in short-term memory, prior to encoding in long-term memory" (p. 296). Schmidt (1994) has also suggested that if a target feature in the input is noticed, it might become intake. In other words, the input becomes intake if learners notice and attend to input.

2.3. Input-based activities and vocabulary learning

Task-based language teaching (TBLT) has been known to include production-based tasks. Swan (2005) put TBLT under question due to this belief that it focuses on 'pushed output' and learners cannot grasp opportunities to acquire 'new linguistic material'. On the contrary, Ellis (2009) suggests that, tasks can also be 'input-based'. He believes that these kinds of tasks can be used to give the learners a chance of learning the 'new linguistic material' that Swan considers it to be necessary. Shintani (2012) argues that:

> An input-based task aims to promote interlanguage development by directing learners' attention to second language (L2) input through listening or reading without requiring them to produce the L2. However, L2 production is not prohibited in an input-based task; learners may elect to respond to the input they receive by engaging in language production (p. 254).

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

409

The main purpose of the task enriched with input is to provide the learners with samples of target forms and items but does not require the learners to indicate that they have successfully processed the input.

Input-Based Reviewing is a process through which the learners try to review and retrieve the target words via such instruments as word cards in which a target word is on one side and on the other side either its translation in learners' mother tongue, or its equivalent in second language or a picture is appeared. Takač (2008) states that "strategies of retrieval play a very important role in learning: every recall of a previously learnt word strengthens the link between knowledge and retrieval cue" (p. 75). As mentioned earlier, Nation (2001) argues that retrieval can take two different forms, namely receptive and productive. The former deals with the cue which is written form then meaning should be retrieved. The latter deals with the cue which is meaning then the word form should be retrieved. Teachers should make learners more aware of the importance of retrieving for their vocabulary learning and encourage them to integrate this repetition technique into their learning activities (Nation, 2001). Retrieval plays an important role in the strategy of using word cards for vocabulary learning and it makes the word cards more favorable for learners compared to other strategies such as notebooks or lists of vocabulary items (Schmitt & Schmitt, 1995; Waring, 2004). Because the target words and their meanings are put on different sides of word cards, retrievals with them will be easier for learners compared with word lists where L2 words and their meanings are presented at the same time. The strategy of using word cards for vocabulary learning will be examined in the following section.

The retrieval process is an indispensible part of flashcard learning. So, learners using flashcards should be encouraged to retrieve the meaning of the target word from memory, which leads to a more permanent learning (Barcroft, 2007; McNamara & Healy, 1995; Nation, 2001). In addition to the retrieval process, the order of the flashcards is another factor which affects the learning process. According to Baddeley's (1990) primacy and recency effects, the items at the beginning and the end of a list are memorized better than the words in the middle. Taking into account this finding, and also the fact that learners have the freedom to change the order of words if they study with flashcards, learners should put difficult words near the beginning, so these words can get more attention.

2.4. Output-based activities and vocabulary learning

There is now a growing consensus on the role that output plays in the language acquisition (Russell, 2014). Simply stated, output is the language a learner produces. Output occurs when a learner discusses and writes within a group of learners who then give immediate feedback for the purpose of solving a problem (Swain, 2000). The findings of Swain's (1985) seminal paper about output countered Krashen's input hypothesis (1980, 1985). The major tenet of her hypothesis (Output Hypothesis) was that the provision of comprehensible input alone is not enough for acquisition to take place and that the production of output forces learners to process language more deeply, attending to both meaning and linguistic form simultaneously (Swain, 1985, 1993, 1995, 1998). As Swain (1985) argues, producing the target language may serve as "the trigger that forces the learner to pay attention to the means of expression needed in order to successfully convey his or her own intended meaning" (p. 249).

The Output Hypothesis can be discussed on the ground of some psycholinguistic and cognitive models of processing as well. According to Anderson's (1983) ACT model, for example, practicing to produce the language is likely to free up the cognitive resources needed for attending to other aspects of language via the process of automatization. Two important concepts of McLaughlin's (1987) Information Processing Model are 1) controlled / automatic processing, and 2) restructuring. Briefly stated, the most important variable facilitating these two mental processes is production practice, which provides the impetus for the shift from

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

410

controlled to automatic processing and the occurrence of restructuring. Later on, McLaughlin and Heredia explained the functions of practice as follows:

> (Production) Practice can have two very different effects. It can lead to improvement in performance as sub-skills become automated, but it is also possible for increased practice to lead to restructuring and attendant decrements in performance as learners reorganize their internal representational framework. It seems that the effect of practice does not accrue directly or automatically to a skilled action, but rather cumulates as learners develop more efficient procedures (McLaughlin & Heredia, 1996, p, 218).

According to Swain (1995, 2005), one important function of output, among others, is helping learners notice the gap between their linguistic resources and the target language system. To help students reach this level of noticing, several out-based activities can be employed. Output-Based Reviewing is one of them in which learners are asked to generate the word forms. Following productive skills, the concept of the word is changed when they make their own expressions, and grasp other features of the word. Furthermore, learners are required to use those words in such activities as paraphrasing the sentences containing the unknown words. Nation (2001) suggests that in generative phase, getting the learners to reconstruct what was in the text rather than repeat it, learners can retell of the written input in different ways. Another exemplar of output-based reviewing is that learners review the target forms and are required to tell stories with pictures. Nation (2001) introduced a generation activity, paraphrasing, which can be used in new contexts across the four skills. According to Swain (2005) depth of processing refers to the degree of analysis and elaboration carried out on input through activities like paraphrasing rather than mere repetition, "with greater depth being associated with longer term and stronger memory traces" (Swain, 2005, p. 475).

3. Method
3.1. Participants
One hundred and fifty intermediate English learners (males and females) at Rezvan and Khan-e-Zaban Institutes in Ardabil were the final participants of the current study. They were selected based on a simple random sampling. The participants were randomly assigned into five groups, namely Noticing through Input enhancement (n=30), Input Enhancement plus Input-Based Retrieving (n=30), Input Enhancement plus Output-Based Reviewing (n=30), Input Enhancement Coupled with Input-Based and Output-Based Reviewing (n=30), and a pilot group (n=30). The participants' age ranged from 16 to 28.

3.2. Materials and Instruments
3.2.1. Preliminary English Test (PET)
In order to select a homogenous set of participants a general proficiency test, Preliminary English Test (PET), was used. It included five sections, namely reading (including 3 parts and 30 items), writing (including 2 parts and 5 items), use of English (including 4 parts and 42 items), listening (including 4 parts and 30 items), and speaking (including 4 parts). Due to time constraints and subjectivity in scoring, the speaking and writing sections were excluded.

3.2.2. Academic Word List
Another instrument was the Academic Word List (AWL) (Coxhead, 2000). The AWL is based on frequency and range data from a corpus of 3.5 million words from academic texts. It includes 570 word families- defined by Coxhead (2000) as "a stem plus all closely related affixed forms"

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

411

(p. 218). It was employed both in the treatment sessions as the target words and assessment sessions as the pre- and posttests items.

3.2.3. Focus on Vocabulary 2 Mastering the Academic Word List

It is a research-based vocabulary textbook developed by Schmitt, Schmitt, and Mann (2011) that gives high intermediate to advanced students the advantage they need to succeed in academic environments and master academic words. It has featured 504 out of 570 word families Coxhead (2000) had identified as frequently used words in academic texts across a wide range of topics. Eight units of this book have been chosen through which 56 academic words have been contextualized, 7 words in each unit. These passages were used as an instrument in which the target words were encountered.

3.2.4. Vocabulary Levels Test (VLT)

It was originally designed by Nation (1990) and has been wildly used in a number of vocabulary research studies (e.g., Cobb, 1997; Laufer & Nation, 1999; Laufer & Paribakht, 1998; Schmitt & Meara, 1997). Later it was modified by Schmitt, Schmitt, and Clapham (2001) and was developed in two versions, where the validation evidence is also presented. Each section of the revised VLT consists of 30 items in a multiple matching format. Items are clustered together in 10 groups for this, so that learners are presented in each cluster with six words in a column on the left and the corresponding meaning senses of three of these in another column on the right. Learners are asked to match each meaning sense in the right-hand column which one single word from the left-hand column. Within each level, there is a fixed ratio of word classes to represent the distribution of English word classes. This ratio is 3 (noun) : 2 (verb) : 1 (adjective) in the latest revised versions (Schmitt et al., 2001). Word classes are not mixed within any one cluster. For the 2001 version of the VLT, two parallel test versions are available that have been established to be relatively equivalent. Both versions will be used both as an assessment instrument and in the validation phase as an established test. The VLT includes four frequency-based levels (the 2,000 word level, 3,000 word level, 5,000 word level, 10,000 word level). It also includes the AWL.

As Brown (1990) argued reliability strategies for norm-referenced tests, e.g., test-retest and equivalent-from strategies, all depend, in one way or another, on the magnitude of the standard deviation. On the other hand, a relatively high standard deviation is one result of developing a test that spreads students out along a continuum of abilities which is a feature of NRT tests. However, the same strategies may be inappropriate for CRTs because they are not developed for the purpose of producing variance among scores.

Of the three categories applied specifically to the estimation of CRT consistency, threshold loss agreement, squared-error loss agreement or domain score dependability, the first one has been used to find the reliability or dependability of the VLT. To do so, two common threshold loss agreement statistics, i.e., agreement coefficient (ρ_o) and kappa coefficient (k) have been employed. These two coefficients measure the consistency of mastery-non-mastery classifications across two test administrations. The agreement coefficient is simply the proportion of examinees consistently classified (as masters or as non-masters) on both administrations and kappa reflects the proportion of consistent classifications observed beyond that expected by chance. However, as Brown (2005) argues, administering a test twice is cumbersome and hard on everyone involved. So several approaches (e.g., Huynh, 1976; Subkoviak, 1976; Peng & Subkoviak, 1980) have been worked out to estimate threshold agreement from one administration. Since these methods required access to computer facilities and appropriate software and assume a somewhat advanced background in test theory, they were also difficult for practitioners to implement. More recently, Subkoviak (1988) presented practical approaches for approximating both the agreement and kappa coefficients. To reach

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

412

either of these coefficients from a single test administration, two values are required. The first is the cut-point score converted to a standard score. The second value is one of the NRT internal-consistency reliability estimates. Once the standardized cut-point score and an internal-consistency reliability are in hand, it is just a matter of checking the appropriate tables developed by Subkoviak (1988, pp. 49-50).

Following "the well-established contrasting group method" proposed by Brown (2005, pp. 236-7) to reach the cut-points, converting these points to standardized cut-point scores according to the formula developed by Brown (2005, p. 203), estimating the internal-consistency reliability for each test through calculating Cronbach alpha coefficient (a), and checking the appropriate tables mentioned above, the agreement coefficient (ρ_o) and kappa coefficient (k) for the reliability of each test have been approximated as follows:

Table 1

Agreement coefficient and kappa coefficient for the pre- and posttests

group	tests	ρ_o	k
G1	Pretest	0.90	0.69
	Immediate posttest	0.95	0.63
	Delayed posttest	0.96	0.62
G2	Pretest	0.92	0.67
	Immediate posttest	0.95	0.63
	Delayed posttest	0.93	0.66
G3	Pretest	0.89	0.70
	Immediate posttest	0.88	0.70
	Delayed posttest	0.91	0.68
G4	Pretest	0.92	0.67
	Immediate posttest	0.87	0.70
	Delayed posttest	0.93	0.66

3.3. Procedure

The initial participants of the present study were 212 male and female students from intermediate classes at the above-mentioned institutes in Ardabil. They sat for the proficiency test (PET) based on which 171 learners were identified as intermediate. Finally, 150 learners whose scores fell within one standard deviation above and below the sample mean were randomly selected as the final participants of the study. Thirty out of the 150 learners were randomly selected and assigned into a pilot group who sat for the VKS consisting of 150 vocabulary items (extracted from the AWL, 60 words in Sublist 8, 60 words in Sublist 9, and 30 words in Sublist 10, respectively) in order to discover the unknown words. Given the results of the VKS, the words the participants recognized through the first and the second categories came to be 71 words out of which 56 words were randomly selected as the target words. Regarding the sample size of lexical items, it is self-evident that the more samples you obtain from your participants, the more valid and reliable your results should be. However, this measurement ideal is always in contrast with the practical reality that the amount of time to administer a study is always limited (Schmitt, 2010) and in practice, a researcher is usually able to measure only a small proportion of lexical items in focus. Furthermore, there is no set answer to the tricky question of how large (or how small) a target sample is acceptable. However, bearing in mind Schmitt's (2010) guidance that the researcher needs to be able to argue that the sample which is measured gives reliable and meaningful information about the vocabulary being discussed and consulting Nation (2017, personal communication), the researchers have come up with the final decision in this regard.

The remaining 120 participants were randomly assigned into four groups each including thirty learners who had already been proved to be homogeneous by meeting the criteria. The

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

413

four treatment conditions included noticing through input enhancement, input enhancement plus input-based reviewing, input enhancement plus output-based reviewing, and input enhancement plus input-based and output-based reviewing. Each group sat for the pretest the items of which were adopted from the VLT developed by the researchers. The items for the four groups were composed of the words coming from Sublists 8, 9, and 10.

A point must be clarified here that by noticing through input enhancement the first condition is meant i.e., noticing, by input enhancement plus input-based reviewing the second one is meant i.e., noticing combined with retrieval, by input enhancement plus output-based reviewing the third one is meant i.e., noticing plus generation, and in the last group input enhancement is coupled with both input-based and output-based reviewing. What follows is a brief account of what was performed in each group:

*Group one***: Noticing through Input enhancement**: Each session the learners in this group were given a reading passage including at least seven out of 56 target words. Seven of those words were highlighted to enhance input and call learners' attention. So participants of this group took part in eight treatment sessions.

*Group two***: Input enhancement plus input-based reviewing**: Learners in this group followed the same procedure as the first group did, that is noticing the target words. What is more in this group is that every session the learners reviewed and retrieved seven words introduced in the previous session by using researcher-made word cards out of those words. The idea of using word cards originates from the fact that retrieval or reviewing "… can involve an informal schedule for returning to previously studied items on word cards and the recycling of old material" (Nation, 2013, p. 302). According to Nation (2013), in order to provide students with a chance of reviewing learnt vocabulary, word cards can be applied in different forms including a foreign language word on one side and either its first language translation, or its second language definition or a picture on the other side of the card. From the three procedures to use flash cards the first one, a foreign language word on one side its first language translation on the other side, was used in this study. It is imperative to mention that the input-based reviewing of seven words presented in the first session was done in the second session and the treatment of the second seven words in the third session and so on. Consequently, the participants of this group took part in nine treatment sessions i.e., the treatment of the last seven target words was done in the ninth session.

*Group three***: Input enhancement plus Output-based reviewing**: Learners in this group followed the same procedures as the first group did; that is, noticing the target words through input enhancement. Furthermore, they were required to use those words in such a generative activity as writing new sentences containing the unknown words. As mentioned above, this approach originates from Nation (2013) who introduced paraphrasing or using words in new contexts across the four skills as a kind of generation activity. It goes without saying that, like the second group, the output-based reviewing of seven words presented in the first session was done in the second session and the treatment of the second seven words was done in the third session and so on. So the participants of this group, as referred to before, took part in nine treatment sessions i.e., the treatment of the last seven target words was done in the ninth session.

Group four: **Input enhancement coupled with input-based and output-based reviewing**: This condition was the sum of the three conditions provided for the first three groups, that is, they had noticing through input enhancement; retrieval through using researcher-made word cards; and generation through rewriting the sentences containing the unknown words. Needless to say, this group, too, took part in nine treatment sessions like the second and third groups.

The immediate posttest including the target words presented in the class was administered one week after the last treatment session. Vocabulary learning is incremental in nature, and only research designs considering this element can truly describe it (Schmitt, 2010). One way to meet this requirement is simply adding one or more delayed posttests. So, the

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

414

participants sat for a delayed posttest two weeks after the immediate one. To exclude any interim immediate posttest exposures and their subsequent practice effect on the delayed posttest, the practical question of the length of the delay arises. There is no clear-cut answer to this question in the literature and as Schmitt (2010) argued "there is no standard period of delay, and that any delay beyond the immediate posttest is better than nothing" (p. 156). Following Schmitt (2010), the delay between the posttests in this study was two weeks.

3.4. Data Analysis

Since the study enjoys a delayed posttest and two dependent variables, to measure within-group differences across different measurement times (from pretest to the immediate posttest and from the immediate posttest to the delayed posttest) four one-way repeated measures ANOVAs were run to analyze the data obtained from the VLT. Moreover to measure the between-group differences, three one-way ANOVAs followed by subsequent post-hoc comparisons were run for the three time intervals among the four groups for the data obtained from the VLT.

4. Results and Discussion

4.1. Results

Within-group analyses

First of all, the normality of the distribution of scores was assessed to decide on using either parametric statistics or their non-parametric equivalents. So, the data obtained was assessed through a series of Kolmogorov–Smirnov tests confirming that the scores obtained are normally distributed. Since each subject was measured on the continuous scale on three occasions, one-way repeated measures ANOVA was conducted to detect the possible significant differences in the participants' performance through the three time intervals.

Table 2
Descriptive Statistics obtained from VLT for the first group

	Mean	Std. Deviation	N
Pretest	8.00	2.913	30
Posttest	12.63	2.965	30
Delayed	12.53	2.874	30

Table 2 shows the descriptive statistics for the scores obtained from VLT for the first group over the pretest, the immediate posttest, and the delayed posttest. It is revealed that the first group's scores increased from the pretest to the immediate and delayed posttests but there was a slight decline from the immediate posttest to the delayed one.

Table 3
One-way repeated measures ANOVA for the VLT obtained from the first group over three time intervals[b]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.990	1.361E3[a]	2.000	28.000	.000	.990
	Wilks' Lambda	.010	1.361E3[a]	2.000	28.000	.000	.990
	Hotelling's Trace	97.182	1.361E3[a]	2.000	28.000	.000	.990
	Roy's Largest Root	97.182	1.361E3[a]	2.000	28.000	.000	.990

a. Exact statistic

b. Design: Intercept
Within Subjects Design: Time

Table 3 shows the results of the one-way repeated measures ANOVA for the VLT obtained from the first group over the pretest, the immediate posttest, and the delayed posttest.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

415

The results revealed that there is a significant difference in the performance of the participants in the first group over the three time intervals (Wilks' Lambda = .10 with a probability value of .000, F (2, 28) = 1.361, $p < .0005$, multivariate partial eta squared (η^2) = .990). To determine the exact location of the difference, the pairwise comparisons are also required (Table 4).

Table 4
Pairwise Comparisons for the first group across the three time intervals

Measure:MEASURE_1

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-4.633[*]	.089	.000	-4.861	-4.406
	3	-4.533[*]	.093	.000	-4.769	-4.298
2	1	4.633[*]	.089	.000	4.406	4.861
	3	.100	.056	.249	-.042	.242
3	1	4.533[*]	.093	.000	4.298	4.769
	2	-.100	.056	.249	-.242	.042

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Based on the results of descriptive statistics (Table 2) coupled with the post-hoc Pairwise Comparisons for the first group (Table 4), it was revealed that the group made gains significantly from the pretest ($M = 8$, $SD = 2.913$) to the immediate posttest ($M = 12.63$, $SD = 2.965$) ($p = .000$) and the delayed posttest ($M = 12.53$, $SD = 2.874$) ($p = .000$) but not from the immediate posttest to the delayed posttest ($p = .249$).

To analyze the data to answer the second research question, that is, the possible effect of the second condition, i.e., input-enhancement combined with input-based reviewing, the same procedures followed for the first research question have been followed the findings of which are presented in the tables 5 through 7.

Table 5
Descriptive Statistics obtained from VLT for the second group

	Mean	Std. Deviation	N
Pretest	7.63	3.068	30
Posttest	21.83	3.291	30
Delayed	19.50	2.418	30

Based on table 5, it is revealed that the second group's scores increased from the pretest to the immediate and delayed posttests but there was a drop from the immediate posttest to the delayed one.

Table 6
One-way repeated measures ANOVA for the VLT obtained from the second group over three time intervals[b]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.996	3.436E3[a]	2.000	28.000	.000	.996
	Wilks' Lambda	.004	3.436E3[a]	2.000	28.000	.000	.996
	Hotelling's Trace	245.404	3.436E3[a]	2.000	28.000	.000	.996
	Roy's Largest Root	245.404	3.436E3[a]	2.000	28.000	.000	.996

a. Exact statistic

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

416

Table 6
One-way repeated measures ANOVA for the VLT obtained from the second group over three time intervals[b]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.996	3.436E3[a]	2.000	28.000	.000	.996
	Wilks' Lambda	.004	3.436E3[a]	2.000	28.000	.000	.996
	Hotelling's Trace	245.404	3.436E3[a]	2.000	28.000	.000	.996
	Roy's Largest Root	245.404	3.436E3[a]	2.000	28.000	.000	.996

b. Design: Intercept
Within Subjects Design: Time

Based on the results of the one-way repeated measures ANOVA for the VLT obtained from the second group over the pretest, the immediate posttest, and the delayed posttest (Table 6), it was revealed that the difference in the performance of the participants in the second group over the three time intervals was significant (Wilks' Lambda = .004 with a probability value of .000, F (2, 28) = 3.436, $p < .0005$, multivariate partial eta squared (η^2) = .996).

Table 7
Pairwise Comparisons for the second group across the three time intervals

Measure:MEASURE_1

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-14.200[*]	.169	.000	-14.629	-13.771
	3	-11.867[*]	.313	.000	-12.663	-11.070
2	1	14.200[*]	.169	.000	13.771	14.629
	3	2.333[*]	.268	.000	1.651	3.015
3	1	11.867[*]	.313	.000	11.070	12.663
	2	-2.333[*]	.268	.000	-3.015	-1.651

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Merging the results of the descriptive statistics (Table 5) with the post-hoc Pairwise Comparisons for the second group (Table 7), it was revealed that the group made gains significantly from the pretest ($M = 7.63$, $SD = 3.068$) to the immediate posttest ($M = 21.83$, $SD = 3.291$) ($p = .000$) and the delayed posttest ($M = 19.50$, $SD = 2.418$) ($p = .000$). It was also revealed that, unlike the first group, the second group's decline from the immediate posttest to the delayed one is significant ($p = .000$).

The results obtained from the third group across three time intervals replicate those for the second group- that is, the third group's scores soared from the pretest to the immediate and delayed posttests but there was a decrease from the immediate posttest to the delayed one (Table 8).

Table 8
Descriptive Statistics obtained from VLT for the third group

	Mean	Std. Deviation	N
Pretest	7.40	2.283	30
Posttest	29.93	4.218	30
Delayed	27.80	3.718	30

To ensure that these changes are statistically significant, the results of the multivariate tests are displayed in Table 9:

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

417

Table 9
One-way repeated measures ANOVA for the VLT obtained from the third group over three time intervals[b]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.995	2.715E3[a]	2.000	28.000	.000	.995
	Wilks' Lambda	.005	2.715E3[a]	2.000	28.000	.000	.995
	Hotelling's Trace	193.961	2.715E3[a]	2.000	28.000	.000	.995
	Roy's Largest Root	193.961	2.715E3[a]	2.000	28.000	.000	.995

a. Exact statistic

b. Design: Intercept
 Within Subjects Design: Time

The within-subjects results (Table 9) revealed that the difference in the performance of the participants in the third group over the three time intervals was significant (Wilks' Lambda = .005 with a probability value of .000, F (2, 28) = 2.715, $p < .0005$, multivariate partial eta squared (η^2) = .995). The results yielded from the pairwise comparisons (Table 10) coupled with the descriptive statistics mentioned above (Table 8) replicated those of the second group, that is, that the third group made significant improvement from the pretest ($M = 7.40$, $SD = 2.283$) to the immediate posttest ($M = 29.93$, $SD = 4.218$) ($p = .000$) and the delayed posttest ($M = 27.80$, $SD = 3.718$) ($p = .000$). It was also observed that the third group's fall from the immediate posttest to the delayed one is significant ($p = .000$).

Table 10
Pairwise Comparisons for the third group across the three time intervals

Measure:MEASURE_1

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-22.533*	.367	.000	-23.467	-21.600
	3	-20.400*	.278	.000	-21.106	-19.694
2	1	22.533*	.367	.000	21.600	23.467
	3	2.133*	.157	.000	1.734	2.532
3	1	20.400*	.278	.000	19.694	21.106
	2	-2.133*	.157	.000	-2.532	-1.734

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

The fourth research question aimed to find the possible effect of the last condition introduced for the first time in this work, that is, the impact of input-enhancement combined with input-based and output-based reviewing on vocabulary learning and retention. To do so, the same procedures conducted for the first three groups were followed for the last group.

Table 11
Descriptive Statistics obtained from VLT for the fourth group

	Mean	Std. Deviation	N
Pretest	7.67	2.657	30
Posttest	41.23	5.425	30
Delayed	36.50	4.841	30

Like the second and third groups, the fourth group, too, gained improvement from the pretest to the immediate and the delayed posttests (Table 11). Based on the results of the

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

418

multivariate tests (Table 12), the difference between the three time intervals is significant (Wilks' Lambda = .005 with a probability value of .000, F (2, 28) = 2.590, $p < .0005$, multivariate partial eta squared (η^2) = .995).

Table 12
One-way repeated measures ANOVA for the VLT obtained from the forth group over three time intervals[b]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.995	2.590E3[a]	2.000	28.000	.000	.995
	Wilks' Lambda	.005	2.590E3[a]	2.000	28.000	.000	.995
	Hotelling's Trace	184.994	2.590E3[a]	2.000	28.000	.000	.995
	Roy's Largest Root	184.994	2.590E3[a]	2.000	28.000	.000	.995

a. Exact statistic

b. Design: Intercept
Within Subjects Design: Time

To find the exact location of the differences between the pretest and the immediate and delayed posttests, the results of the pairwise comparisons are displayed in Table 13 according to which it can be asserted that all three time intervals are significantly different from one another. Furthermore, taking the results of the previous findings into account (Tables 11 and 12), it can be concluded that the scores soared significantly from the pretest ($M = 7.67$, $SD = 2.657$) to the immediate posttest (M 41.23, $SD = 5.425$) ($p = .000$) and the delayed posttest ($M = 36.50$, $SD = 4.841$) ($p = .000$). The difference between the immediate posttest and the delayed one is also significant ($p = .000$). The forth research question is also partially rejected, that is, the impact of input-enhancement combined with input-based and output-based reviewing on vocabulary learning is statistically significant but its effect on vocabulary retention is under question.

Table 13
Pairwise Comparisons for the fourth group across the three time intervals

Measure:MEASURE_1

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	-33.567[*]	.513	.000	-34.871	-32.262
	3	-28.833[*]	.407	.000	-29.868	-27.799
2	1	33.567[*]	.513	.000	32.262	34.871
	3	4.733[*]	.143	.000	4.369	5.097
3	1	28.833[*]	.407	.000	27.799	29.868
	2	-4.733[*]	.143	.000	-5.097	-4.369

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Between-group analyses

To see whether the four conditions have differential effects on vocabulary learning and retention or not, three one-way ANOVAs were run for the three time intervals among the four groups. The first one was conducted for the data obtained from the pretests of the four groups:

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

419

Table 14
Descriptive Statistics for the pretests of the four groups
Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
G1	30	8.00	2.913	.532	6.91	9.09	3	13
G2	30	7.63	3.068	.560	6.49	8.78	3	13
G3	30	7.40	2.283	.417	6.55	8.25	4	11
G4	30	7.67	2.657	.485	6.67	8.66	4	12
Total	120	7.68	2.720	.248	7.18	8.17	3	13

Table 15
One-way ANOVA for the pretests of the four groups
Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5.492	3	1.831	.243	.866
Within Groups	874.833	116	7.542		
Total	880.325	119			

Based on the one-way ANOVA run (Table 15), the four groups were taken to be equal in their initial proficiency in the target words, *F* = .24, *p* = .86.

To discover the potential immediate outperformance of the groups over one another, the data of the immediate posttest was also analyzed through a one-way ANOVA accompanied by post-hoc comparisons the results of which are presented in tables 16-18:

Table 16
Descriptive Statistics for the immediate posttest of the four groups
Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
G1	30	12.63	2.965	.541	11.53	13.74	8	17
G2	30	21.83	3.291	.601	20.60	23.06	17	27
G3	30	29.93	4.218	.770	28.36	31.51	23	36
G4	30	41.23	5.425	.990	39.21	43.26	33	51
Total	120	26.41	11.311	1.033	24.36	28.45	8	51

Table 17
One-way ANOVA for the immediate posttest of the four groups
Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13286.625	3	4428.875	265.042	.000
Within Groups	1938.367	116	16.710		
Total	15224.992	119			

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

420

Table 18
Post hoc Multiple Comparisons for the immediate posttest of the four groups

Dependent Variable: Score

	(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	G1	G2	-9.200*	1.055	.000	-12.19	-6.21
		G3	-17.300*	1.055	.000	-20.29	-14.31
		G4	-28.600*	1.055	.000	-31.59	-25.61
	G2	G1	9.200*	1.055	.000	6.21	12.19
		G3	-8.100*	1.055	.000	-11.09	-5.11
		G4	-19.400*	1.055	.000	-22.39	-16.41
	G3	G1	17.300*	1.055	.000	14.31	20.29
		G2	8.100*	1.055	.000	5.11	11.09
		G4	-11.300*	1.055	.000	-14.29	-8.31
	G4	G1	28.600*	1.055	.000	25.61	31.59
		G2	19.400*	1.055	.000	16.41	22.39
		G3	11.300*	1.055	.000	8.31	14.29

*. The mean difference is significant at the 0.05 level.

Based on the results of the one-way ANOVA run (Table 17), the between-group differences in the immediate posttest were significant ($F = 265.042$, $p = .000$). The post-hoc test (Table 18) run showed that the differences between the groups were significant. More specifically, taking into account the descriptive statistics (Table 16), it was concluded that the second group outperformed the first group, the third group outscored the first and the second ones and the fourth group outperformed the three previous groups.

The same procedures were followed to find the potential long-term differences in the performance of the four groups.

Table 19
Descriptive Statistics for the delayed posttest of the four groups

Score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
G1	30	12.53	2.874	.525	11.46	13.61	8	17
G2	30	19.50	2.418	.441	18.60	20.40	16	24
G3	30	27.80	3.718	.679	26.41	29.19	22	34
G4	30	36.50	4.841	.884	34.69	38.31	29	45
Total	120	24.08	9.685	.884	22.33	25.83	8	45

Table 20
One-way ANOVA for the delayed posttest of the four groups

Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9671.900	3	3223.967	251.117	.000
Within Groups	1489.267	116	12.839		
Total	11161.167	119			

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

421

Table 20
One-way ANOVA for the delayed posttest of the four groups

Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9671.900	3	3223.967	251.117	.000
Within Groups	1489.267	116	12.839		

Table 21
Post hoc Multiple Comparisons for the delayed posttest of the four groups

Dependent Variable:Score

	(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	G1	G2	-6.967*	.925	.000	-9.59	-4.34
		G3	-15.267*	.925	.000	-17.89	-12.64
		G4	-23.967*	.925	.000	-26.59	-21.34
	G2	G1	6.967*	.925	.000	4.34	9.59
		G3	-8.300*	.925	.000	-10.92	-5.68
		G4	-17.000*	.925	.000	-19.62	-14.38
	G3	G1	15.267*	.925	.000	12.64	17.89
		G2	8.300*	.925	.000	5.68	10.92
		G4	-8.700*	.925	.000	-11.32	-6.08
	G4	G1	23.967*	.925	.000	21.34	26.59
		G2	17.000*	.925	.000	14.38	19.62
		G3	8.700*	.925	.000	6.08	11.32

*. The mean difference is significant at the 0.05 level.

Based on the results of the one-way ANOVA run (Table 20), the between-group differences in the delayed posttest were also significant ($F= 251.11$, $p = .000$). These results coupled with the descriptive statistics (Table 19) and the results obtained from the post-hoc test (Table 21) showed that in the delayed posttest, like the immediate posttest, the second group outperformed the first group, the third group outscored the first and the second ones and the fourth group outperformed the three previous groups.

4.2. Discussion

Regarding the first condition, it was observed that the exposure to input which is textually enhanced through boldfacing does facilitate the learning of vocabulary and has immediate effect on it. In the long term, however, the results are fairly vague. Although the group gained significant improvement from the pretest to the delayed posttest, there was a slight decline in the performance of the group from the immediate posttest to the delayed one. Putting all mentioned findings into account, it can be argued that the first condition, i.e. noticing through input-enhancement, has significant positive effects on vocabulary learning but its effect on vocabulary retention should be dealt with greater care and conservatism. Theoretically speaking, the first part of this finding i.e., positive effects of noticing through input-enhancement on vocabulary learning, confirms the original psycholinguistic underpinnings of attention and/or awareness in SLA in general and second language vocabulary learning in particular. As Leow (2015) argues one irrefutable key process in all of the psychology-based theoretical underpinnings is the role attention plays in the L2 learning process. Particularly

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

422

stated, the finding is in line with Schmidt's (1990, 1994, 2001) Noticing Hypothesis through which he argues that only L2 data that are noticed can be converted into intake and noticing is "the necessary and sufficient condition for the conversion of input to intake" (Schmidt, 1994, p. 209). The finding is also in line with Sharwood-Smith's (1991, 1993) "input enhancement" by which he meant the guidance teachers provide for promoting second language learners' self-discovery or conscious awareness of the formal features of the L2. Regarding the empirical side of the coin, this finding is in line with Baleghizadeh, Yazdanjoo, Fallahpour (2018), Izumi (2002), Jourdenais et al. (1995), Leow (2001), Lee and Lee (2012), and Rassaei (2014). Regarding the second finding, i.e. the lack of positive effects of noticing through input-enhancement on vocabulary retention, several facts should be referred to. First of all, there are few empirical studies that have used delayed posttests to see if the positive effects of input enhancement actually persist in the longer term (Sharwood-Smith & Truscott, 2014). However, theoretically speaking, the finding can be justified from some perspectives. An item of information can be consolidated and transferred from working memory to long-term memory by rehearsal. The more often we retrieve a particular item of information from long-term memory, the easier it becomes to access it and the less likely it is to be lost. Information that is rarely retrieved may decay and language attrition is an example of such decay (Field, 2004). On the other hand, vocabulary knowledge is more likely to experience such attrition compared to other linguistic aspects (Schmitt, 2010). So, we can argue that the decline from the immediate posttest to the delayed one can be the result of language attrition because between the two posttests there was no language exposure.

Regarding the second condition, first, it is asserted that input-enhancement combined with input-based reviewing has immediate significant positive effect on the participants' vocabulary knowledge (here called vocabulary learning). Ellis (2003) explained that two types of input-based approaches, structured input and consciousness-raising, can be best used in teaching grammar. The first approach, structured input, has been adapted to the teaching of L2 vocabulary through making use of flash cards. Such structured input task occupies an important part in processing instruction (Takimoto, 2007). VanPatten (1996) argues that processing instruction entails an explanation of the relationship between a given form and the meaning it can convey and "structured input activities in which learners are given the opportunity to process form in the input in a controlled situation so that better form-meaning connections might happen compared with what might happen in less controlled situations" (p. 60). So, the substantial gains for the participants after reviewing the words through form-meaning connections and the controlled activity can be attributed to what Ellis (1997) termed "structured input". Furthermore, this finding lends support to previous studies (De la Fuente, 2002; De Jong, 2005; Ellis & He, 1999; Shintani, 2012), which have shown that input-based instruction benefits vocabulary knowledge. Second, comparing the results of the pretest with the delayed posttest, we observed that in the long-term, too, the participants maintained the positive effect of the treatment. However, they experienced a significant decline from the immediate posttest to the delayed one leading us to cast doubt on the long-term significant positive effect of the second treatment on vocabulary knowledge as well (here called retention). Like the first case, this decline can be justified by the lack of exposure between the immediate and the delayed posttests. Furthermore, we can argue that not only was input-based reviewing unable in compensating this lack of exposure but also the situation got worse, that is, the decline from the immediate posttest to the delayed one was significant. This conclusion is in line with van Zeeland and Schmitt (2013). They compared the effects of four levels of input frequency in one

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

423

learning session. Although they found small positive effects of repeated input on the immediate posttest, no significant difference was found on the posttest conducted 2 weeks after learning. Based on their finding, Nakata (2016) argued that "although increasing frequency in one learning session enhances learning in the short term, its advantage diminishes over time" (p. 4). This finding can also be analyzed through the distinction drawn between working memory and long-term memory (Baddeley, 1999, 2007; Baddeley & Hitch, 1974; Gathercole & Alloway, 2008) that will be discussed at end of this section. According to Nation (2013), flash cards can be employed as a tool to review the input already provided. However, in order for such a strategy to succeed students should enjoy some skills to benefit its advantages. As Nakata (2008) claimed about his non-native Japanese participants, it is likely that the subjects of the current study, too, could not take full advantage of benefits offered by cards due to lack of experience.

The findings pertinent to the third condition, i.e. input-enhancement combined with output-based reviewing, replicated those of the second condition discussed above. Specifically speaking, it can be concluded that the input-enhancement combined with output-based reviewing positively affects the participants' vocabulary knowledge (here called vocabulary learning). Furthermore, based on the significant difference between the pretest and the delayed posttest, it can be argued that the treatment has long-term significant positive effect on vocabulary knowledge as well (here called retention). However, there was a significant decline from the immediate posttest to the delayed one. The positive effects of input-enhancement combined with output-based reviewing on vocabulary learning can be discussed with regards to Swain's (1985) Output Hypothesis through which she argued that learners must be pushed to produce output in the second language in order to develop accuracy. This finding is also consistent with some works studying and confirming the noticing function of output (Adams, 2003; Kwon, 2007; Horibe, 2003; Izumi & Bigelow, 2000). The possible effect of this treatment on vocabulary retention, to the best of our knowledge, has not been the subject of empirical research. The finding reveals that although the treatment has worked regarding the participants' improvement from the pretest to the immediate posttest, its effectiveness is under question when the delayed posttest is concerned. To put it differently, the positive effect of this treatment like the previous one on the long-term retention of vocabulary is rejected and we can claim that output-based reviewing, too, couldn't save the words from being faded away.

Putting the findings related to the second and third conditions into account, the positive effect of the fourth condition, input-enhancement combined with input-based and output-based reviewing, on vocabulary learning is predictable. The condition had positive effect on vocabulary learning, that is, the group's improvement from the pretest to the immediate and delayed posttests was significant. However, the participants experienced a significant decline from the immediate posttest to the delayed one. Reviewing the theoretical perspectives discussed above, we can claim that the first finding is consistent with Noticing, Input and Output hypotheses. However, the second finding, the decline from the immediate posttest to the delayed one, is intriguing. Since the amount of exposure in this group was more than the previous three groups, it was predicted that the groups' performance would not decline from the immediate posttest to the delayed one. However, the prediction was not confirmed and retention has not been achieved through this condition. In contrast to the comparative studies, the potential efficacy of combining input-based and output-based practices has not received much research attention. Although some studies were conducted to find the effect of such synergy on

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

424

grammar learning (Gass & Torres, 2005; Kirk, 2013; Jabbarpoor & Tajeddin, 2013; Tanaka, 1999, 2001), no such study was conducted for vocabulary learning.

What discussed above was concerned with the within-group differences over three testing periods for each condition. To discuss the differential effects of the vocabulary teaching procedures in focus, the between-group differences should be taken into account too.

Based on the date gained in the immediate posttest, the results revealed that 1) the students treated through input enhancement+input-based reviewing outperformed those treated through input enhancement alone, 2) the students treated through the input enhancement+ouput-based reviewing performed better than those treated through the input enhancement alone and those treated through input enhancement+input-based reviewing and finally, 3) the students treated through input enhancement+input- and output-based reviewing outperformed those treated through input enhancement alone, those treated through input enhancement+input-based reviewing and those treated through input enhancement+out-based reviewing.

With respect to the first finding, as Sharwood-Smith (1993) himself cautioned the enhanced input might or might not be further processed into the language system, that is, L2 knowledge. This caution can be interpreted that while "enhanced input is noticed and taken in by the learner, such linguistic information may not be further processed due to the kind of evidence to which learners are exposed" (Leow, 2015, p. 166). Moreover, Leow's (2015) assertion about the grammatical benefits of input enhancement can be employed to justify the superiorly of input enhancement coupled with input-based reviewing over input enhancement without such reviewing in the realm of vocabulary learning. After having a critical review of the literature on grammatical benefits of input enhancement, he argued that a conflation of input enhancement with one or more variables (e.g., instruction, feedback, metalinguistic information, oral discussion, etc.) appears to contribute to better grammatical development. In terms of vocabulary instruction, one of those variables is input-based reviewing operationalized, in this study, through using flash cards. The finding is in line with Kornell (2009) and Komachali and Khodareza (2012) which found the positive effect of using flashcard-based instruction on vocabulary learning. The finding also confirmed two recent studies investigating the efficacy of flashcard-based instruction, Lin-Fang (2013) and Sinaei and Asadi (2014).

Regarding the second finding, most of the studies investigating the efficacy of output-based activities have mainly focused on grammar and compared its effect with some other grammar instructions the results of which are far from conclusive and mixed enough to urge researchers to do more works to clarify the point. The story is blurrier in terms of vocabulary instruction and the studies on input- and output-based instruction of vocabulary have revealed mixed results. While many studies confirm the positive role of output production in the development of learners' vocabulary (DeKeyser, 1997; Hashemi & Kassaian, 2011; Izumi & Bigelow, 2000; Jalilifar & Amin, 2008; Kwon, 2007; Soleimani, Ketabi, & Talebinejad, 2008;), in some other studies, input-based instruction appears to be more effective in L2 vocabulary development (Long, Inagaki, & Ortega, 1998; Shintani, 2011). Some others, like Rassaei (2012), concluded that L2 knowledge might develop through both input-based and output-based instruction. This finding of the current study confirms the noticing function of output in vocabulary learning. As Swain (2005) claims while attempting to produce the target language, learners may notice that they do not know how to say precisely the meaning they wish to convey. Consequently, they might pay more attention to something they need to discover about their L2. Output not only makes learners notice their interlanguage defects, but also activates

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

425

the internal cognitive process of second language acquisition, so as to promote language acquisition. It also provided the participants with the chance of hypothesis testing and the consolidation of interlanguage knowledge.

With regard to the third finding, we can argue that continuing to investigate the two types of practice in order to determine which should predominate the other may not be the best avenue to pursue, especially since both input and output are widely used in language classrooms. Differently stated, uniquely focusing on one approach or the other may be unfruitful; rather, we should look after the best ways to merge these two approaches in order to exploit the unique benefits of each simultaneously. Research combining input- and output-based instructional approaches has been scarce. There are few studies (e.g., Gass & Torres, 2005; Hanaoka, 2007; Izumi, 2002; Kirk, 2013; Leeser, 2008; Tanaka, 1999, 2001) investigating such synergy based on which we can claim that the combination seems to be favorable on the whole whenever grammar instruction is concerned. However, to the best of our knowledge, no attempt has been made to study the effect of the integration of input- and output-based practices on vocabulary learning. Integrating input- and output-based practices, the participants enjoyed the retrieval advantage of the first one and the noticing function of the latter.

Regarding the between-group differences in the long term, i.e. the differences reported in the delayed posttest, the results replicated those for the immediate posttest. Regarding the superiority of input enhancement combined with input-based reviewing through flash cards over input enhancement without such intervention, it is worth mentioning that vocabulary learning through flash cards can be categorized as a kind of rehearsal. Rehearsal is defined as an activity through which learners encode new information into their long-term memory through overt or silent articulation (Tanaka, 2008). Based on cognitive psychology literature, Tanaka (2008) concludes that "a successful recall from memory yields superior retention to mere presentation of the target item because the very act of retrieving information from memory strengthens retrieval routes to memory" (p. 5). According to the phenomenon known as the "retrieval practice effect" (Baddeley ,1997; Ellis, 1995; Nation, 2013), testing one's memory to recall the L2 word form or its meaning is beneficial to long-term retention.

The second argument drawn from the between-group differences is the superiority of the input enhancement combined with output-based reviewing over the input enhancement combined with input-based reviewing. Although the argument that vocabulary used in production tasks was recalled better than words practiced in non-production tasks has been questioned by some studies (e.g. Dekeyser & Sokalski, 1996; Horibe, 2003; Sakai, 2004), it is supported by a number of empirical studies (Ellis & He, 1999; Hulstijn & Laufer, 2001; Hulstijn & Trompetter, 1998; Joe, 1998; Nobuyoshi & Ellis, 1993). One rationale underlying the long-term effectiveness of output is that production induces high cognitive involvement, and is more efficient in leaving durable imprints in memory than simple exposure to target words in the input (Kwon, 2007). Furthermore, production provides L2 learners' with an opportunity to promote noticing on their lexical knowledge, either on their deficiency of vocabulary or on the gap between their receptive and productive vocabulary knowledge.

Vocabulary retention gained through input enhancement coupled with input- and output-based reviewing has been reported to overtake both input enhancement without any of them and with one of them. In this condition, the input condition provides learners with a good chance to understand the meanings of the target words, and such comprehension is transferred to the production of words when they have extra opportunity to produce those words in the

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

426

output phase. According to Lee and Benati (2009), general consensus exists that the effects of instruction will fade, wear off, or diminish over time, but that the results of effective instruction should not disappear completely. They employed the term 'durative effects' to refer to posttesting that occurs beyond the time frame of examining the immediate effects of instruction. The last treatment employed in the current study is one of those effective instructions that had durative effects, preventing vocabulary diminution and resulting in durable learning.

5. Conclusion and Implications

Research on input- and output-based practice has aimed to determine which is superior and perhaps more important for the development of vocabulary. The present study was an attempt to move past the old fashioned dichotomy of input vs. output practice and introduce a new way to look at vocabulary instruction. By combining comprehension and production practice, we took a step away from the classical input vs. output debate to gain a more detailed perspective on the factors affecting successful vocabulary instruction. The results of this study showed that combining input and output practice was effective for vocabulary learning. The study showed that output-based practice in conjunction with input-based practice was effective in developing learners' ability to perceive the target vocabulary as well as produce it in a controlled fashion.

At the theoretical level, the outcomes from this study have reaffirmed the positive effects of input and output in SLA in general and in vocabulary learning in particular. Specifically speaking, the role and the importance of language input in enhancing vocabulary knowledge have been reemphasized through this study. In respect of this, the results of the present study provide further support for Input Hypothesis. More specifically, the results of the present study add further value to the effectiveness of word card using through which input has been reviewed. It can be also suggested that although it is often assumed that most vocabulary is learned from natural context (Krashen, 1989), vocabulary learning from context alone is not sufficient, particularly in an EFL context where learners have little exposure to the target language outside the classroom, and should be complemented with word-focused activities such word card, which is more certain and efficient (Hulstijn, 2001; Laufer, 2003; Nation, 2013; Waring, 2004). Regarding the other side of the coin, output, this study like some other studies (e.g., Ellis & He, 1999; Hashemi & Kassaian, 2011; Jalilifar & Amin, 2008; Kwon, 2007; Sarani, Mousapour, & Ghaviniat, 2013) lends support to Swain's (1995) Output Hypothesis and suggests that production facilitates vocabulary learning. It mainly emphasizes the importance of output in developing L2 vocabulary knowledge and considering output as essential for the acquisition of productive vocabulary. The main finding of this study is that combining comprehension and production practice can be a suitable alternative to dichotomous input- and output-based instruction. This finding not only paves the ground for building up further hypotheses regarding vocabulary instruction but it has also got some pedagogical implications.

As far as the pedagogical level is concerned, it was observed that when learners are pushed to paraphrase the sentences including the target words they mainly noticed lexical problems and partly noticed grammatical gaps. This finding is in line with some other studies (e.g., Alsulami, 2016; Mackey, Gass, & McDonough, 2000; Swain & Lapkin, 1995). So, output production should be taken into consideration as an essential condition for L2 vocabulary learning and output practice should necessarily be included in L2 vocabulary instruction in order for full development of learners' vocabulary knowledge. Accordingly, we can recommend teachers to employ output-based activities to draw learners' attention to their lexical gap between their interlanguage and the target language. Teacher's feedbacks regarding learners' output can be provided to ensure that learners avoid misunderstanding of and misuse of target vocabulary during paraphrasing. These feedbacks can be provided through some follow-up mini-sessions where the outcomes are generated by learners.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

427

Broadly speaking, deliberate decontextualized word learning from word cards that was once under criticism in the 1980s with the advent of communicative learning methodologies has been proved to be not only an efficient and convenient but also a very effective method of L2 vocabulary acquisition. Word cards using is, indeed, one of the uncommunicative, input-based and deconceptualized approaches recommended to be employed by teachers whenever vocabulary instruction is concerned through which they can short-cut the long learning process. Moreover, in order to have a more comprehensive teaching of vocabulary, teachers are suggested to present vocabularies in flashcard-based instruction which can cause a better learning of target words. The value of word-focused activities is emphasized from another perspective as well. As Maftoon and Haratmeh (2013) asserted vocabulary is considered as one of the important components of reading comprehension ability of learners. To take this importance into account, teachers can include some word-focused tasks in General English courses where the primary emphasis is placed on the knowledge of the reading comprehension.

The results of the present study also support the use of output practice as well as input-based practice in the L2 classroom environment as a means for building suitable vocabulary knowledge, not only in later stages of instruction but also during early stages of instruction of new words. Combining practice types may promote better learning than their separate use. The results of this study support the claim that combining comprehension and production practice can increase not only immediate comprehension and production abilities, but also may promote durability. Therefore it is suggested that design and organization of practice activities should incorporate both types of practice. Combining practice can pave the ground to success in second language vocabulary learning. Based on principled analysis of the lexical loading of teaching material (Milton & Alexiou, 2012) curriculum designers and/or materials writers can use the findings of the current study to reanalyze the commonly used materials and develop, adopt or adapt vocabulary exercises in which comprehension and production are incorporated to pave the ground for the teachers to help their students reach the possible highest point regarding their vocabulary knowledge.

In the current study, the input-based activity was implemented through word card using and output-based activity through paraphrasing. Future studies can use other approaches to operationalize these two practice types. Not as a cliché, but the individual differences of the participants (e.g., their age, gender, proficiency level, motivation, and self-confidence) can be considered as other variables to study. Other insightful studies might be done in the form of mixed-method research investigating the way(s) the participants would employ the approaches in focus and their attitudes towards them. The order in which the input-based and output-based activities were presented in the last condition was not considered as a variable in this study. Future researchers can study whether and to what extent different activity sequences have an effect on increasing learners' attention which finally would lead to different degrees of vocabulary learning and retention.

REFERENCES

Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research, 7*, 347–376.

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), Attention and awareness in foreign language learning (pp. 259–302). Honolulu: University of Hawai'i.

Anderson, R. W. (1983). Transfer to somewhere. In S.M. Gass & L. Selinker (Eds.), *Language learning* (pp. 177–201). Rowley, MA: Newbury House.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

428

Alsulami, S. Q. (2016). Testing the Noticing Function of the Output Hypothesis. *English Language Teaching, 9(2),* 136.

Baddeley, A. D. (1990). *Human memory: Theory and practice.* London: Lawrence Erlbaum Associates.

Baddeley, A. D. (1997). *Human memory: theory and practice*. East Sussex: Psychology Press.

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*: *Volume 8* (pp. 47–89). New York: Academic Press.

Baleghizadeh, S., Yazdanjoo, S., & Fallahpour, H. (2018), The Effect of Input Enhancement on Academic Vocabulary Learning among Intermediate EFL Learners in Iran. *TESL Reporter 50*(2), 31–48.

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*, 35–56.

Brown, J. D. (1990). Short-cut estimators of criterion-referenced test consistency. *Language Testing, 7*(1), 77–97.

Brown, J. D. (2005). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25,* 301–15.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34,* 213–238.

Cullen, R. (2012). Grammar instruction. In A. Burns & J. Richards (Eds.), *The Cambridge guide to pedagogy and practice in second language teaching* (pp. 258–266). New York, NY: Cambridge University Press.

De Jong, N. (2005). Can second language grammar be learned through listening?: An experimental study. *Studies in Second Language Acquisition, 27,* 205–234.

De la Fuente, M.J. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition, 24,* 81–112.

DeKeyser, R., & Sokalski, K. (1996). The differential role of comprehension and production practice. *Language Learning, 46,* 613–642.

DeKeyser, R., (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19(2),* 195–221.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

Ellis, R. & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21,* 285–301.

Ellis, N. C. (2008). Usage-based and form-focused language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition*, (pp. 372–405). New York: Routledge.

Ellis, N., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, norm, and function. Introduction to the special issue. *Modern Language Journal*, *93*, 329–335.

Field, J. (2004). *Psycholinguistics*. New York: Routledge.

Gary, J.O. & Gary, N. (1981). Comprehension-based language instruction: Theory. *Annals of the New York Academy of Sciences, 379,* 332–42.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

429

Gass, S, M. & Mackey, A. (2015). Input, Interaction and Output in Second Language Acquisition. In B. vanPatten & J. Williams (Eds.). *Theories in Second Language Acquisition – An Introduction.* (pp.180–206). New York: Routledge.

Gass, S. M. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.

Gass, S., & Torres, M. J. A. (2005). Attention when: An investigation of the ordering effect of input and interaction. *Studies in Second Language Acquisition, 27*, 1–31.

Gathercole, S.E., & Alloway, T.P. (2008). *Working memory and learning: A practical guide for teachers.* London: Sage Publications.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483–517.

Hanaoka, O. (2007). Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research, 11*(4), 459–479.

Hashemi, S., & Kassaian, Z. (2011). Effects of learner interaction, receptive and productive learning tasks on vocabulary acquisition: An Iranian case. *Procedia Social and Behavioral Sciences, 15,* 2165–2171.

Hatch, E. (1983). *Psycholinguistics: A second language perspective*. Rowley, MA: Newbury House.

Horibe, S. (2003). The output hypothesis and cognitive processes: An examination via acquisition of Japanese temporal subordinate conjunctions (Unpublished doctoral dissertation). Purdue University, Lafayette.

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–86). Cambridge: Cambridge University Press.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning, 51,* 539–558.

Hulstijn, J., & Trompetter, P. (1998). Incidental learning and second language vocabulary in computer-assisted reading and writing tasks. In D. Albrechtsen, B. Henriksen, I. Mees, & E. Poulsen (Eds.), *Perspectives on Foreign and Second Language Pedagogy* (pp. 191–202). Denmark: Odense University Press.

Huynh, H. (1976). On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement, 13,* 253–264.

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study of ESL relativization. *Studies in Second Language Acquisition*, *24*, 541–577.

Izumi, S. & Bigelow, M. (2000). Does output promote noticing and second language acquisition? *TESOL Quarterly, 34(2),* 239–278.

Izumi, S., Bigelow, M., Fujiwara, M., & Fearnow, S. (1999). Testing the Output Hypothesis. Studies in Second Language Acquisition, *21*(3), 421–452.

Jabbarpoor, S., & Tajeddin, Z. (2013). The effect of input enhancement, individual output, and collaborative output on foreign language learning: the case of English inversion structures. *RESLA 26*, 267–288.

Jalilifar, A.R., & Amin, F. (2008). Classroom vocabulary learning: Investigating the role of task types on Iranian high school learners of English as a foreign language. *Teaching English Language and Literature Society of Iran, 2(8),* 111–141.

Joe, A. (1998). What effect do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19,* 357–377.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

430

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. (1995). Does textual enhancement promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183–216). Honolulu: University of Hawaiʻi.

Kirk, R. (2013). The Effects of Processing Instruction with and without Output: Acquisition of the Spanish Subjunctive in Three Conjunctional Phrases. *Hispania, 96*(1), 153–169.

Komachali, M., & Khodareza, M. (2012). The Effect of Using Vocabulary Flash Card on Iranian Pre-University Students' Vocabulary Knowledge. *International Education Studies, 5(3),* 134–147.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23(9),* 1297–1317.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.

Krashen, S. D. (1985). *The Input Hypothesis: Issues and Implications*, New York: Longman.

Kwon, S. H. (2007). The Roles of Output on L2 Vocabulary Acquisition: Noticing, Retrieval and Retention. *English Education, 62*(4), 279–310.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal, 73*, 440–463.

Laufer, B. (2003). Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *The Canadian Modern Language Review, 59,* 567–587.

Laufer, B. & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*, 33–51.

Laufer, B. & Paribakht, T.,S. (1998). The relationship between passive and active vocabularies: effects of language learning context. *Language Learning, 48*, 365–91.

Lee, J. F., & Benati, A. G. (2009). *Research and perspectives on processing instruction.* Berlin: Mouton de Gruyter.

Lee, S. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning, 57*(1), 87–118.

Lee, Y.H. & Lee, H.K. (2012). Effects of input enhancement techniques on word learning of Korean elementary learners. *Primary English Education, 18*(3), 381–398.

Lee, J. F., & VanPatten, B. (2003). *Making communicative language teaching happen, 2nd edition*. New York, NY: McGraw-Hill.

Leeser, M. (2008). Pushed output, noticing, and development of past tense morphology in content-based instruction. *The Canadian Modern Language Review, 65,* 195–220.

Leow, R. P. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania, 84*, 496–509.

Leow, R. P. (2015). *Explicit learning in the L2 classroom: a student-centered approach*. New York: Routledge.

Leow, R., Egi, T., Nuevo, A., & Tsai, Y. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning, 13*, 1–16.

Lin-Fang, W. (2013). A study of factors affecting college students' use of ESL vocabulary learning strategies. *International Journal of Humanities and Social Sciences, 3*(19), 202–208.

Long, M.H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126–41.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

431

Long, M.H., Inagaki S,. & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal, 82(3),* 357–371.

Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning*. Oxford: Oxford University Press.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*(4), 471–497.

Maftoon, P., & Haratmeh, M. S. (2013). Effects of input and output-oriented tasks with different involvement loads on the receptive vocabulary knowledge of Iranian EFL learners. *Iranian Journal of Research in English Language Teaching, 1*(1), 24–38.

McLaughlin, B. (1987). *Theories of Second-Language Learning*. London, Baltimore, Melbourne, Auckland: Edward Arnold.

McLaughlin, B., & Heredia, R. (1996). Information processing approached to research on second language acquisition and use. In W. Ritchie & T. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 213–225). San Diego, CA: Academic Press.

McNamara, D. S., & Healy, A. F. (1995). A generation advantage for multiplication skill training and nonword vocabulary acquisition. In A. F. Healy & L. E. Bourne Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 132–169). Thousand Oaks, CA: Sage.

Milton, J., & Alexiou, T. (2012). Vocabulary input, vocabulary uptake and approaches to language teaching. *The Language Learning Journal, 40(1),* 1–5.

Mitchell, R., Myles, F. & Marsden, E. (2013). *Second Language Learning Theories.* Abingdon: Routledge.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: implications from cognitive psychology research for optimal spaced learning. *ReCALL, 20*(1), 3–20.

Nakata, T. (2016). Does Repeated Practice Make Perfect? The Effects of Within-Session Repeated Retrieval on Second Language Vocabulary Learning. *Studies in Second Language Acquisition, 39*(4)*,* 653–679.

Nassaji, H., & Fotos, S. S. (2011). *Teaching Grammar in Second Language Classrooms: Integrating Form-Focused Instruction in Communicative Context*. Florence: Taylor and Francis.

Nation, I. S. P. (1982). Beginning to learn foreign vocabulary: a review of the research, *RELC*, *13*, 14–36.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Massachusetts: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.)*.* Cambridge: Cambridge University Press.

Nobuyoshi, J., and Ellis, R. (1993). Focused communication tasks and second language acquisition. *English Language Teaching, 47*, 203–210.

Peng, C.-Y. J., & Subkoviak, M. J. (1980). A note on Huynh's normal approximation procedure for estimating criterion-referenced reliability. *Journal of Educational Measurement, 17,* 359–368.

Rassaei, E. (2014). Effects of Textual Enhancement and Input Enrichment on L2 Development. *TESOL Journal,6*(2), 281–301.

Rassaei, E., (2012). The effects of input-based and output-based instruction on L2 development. *The Electronic Journal for English as a Second Language, 16(3),* 1–25.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

432

Reinders, H., & Ellis, R. (2009). The effects of two types of input on intake and the acquisition of implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning and teaching* (pp. 281–302). Bristol, England: Multilingual Matters.

Robinson, P. (2013). *The Routledge Encyclopedia of Second Language Acquisition*. Abingdon: Routledge.

Russell, V. (2014). A Closer Look at the Output Hypothesis: The Effect of Pushed Output on Noticing and Inductive Learning of the Spanish Future Tense. *Foreign Language Annals, 47*(1), 25–47.

Sakai, H. (2004). Roles of output and feedback for L2 learners' noticing. *JALT journal, 26,* 25–54.

Sarani, A., Mousapour, G., & Ghaviniat, M. (2013). The role of task type in L2 vocabulary acquisition: a case of Involvement Load Hypothesis. *Language and Culture, 35*(4), 377–386.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). London: Academic Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Schmitt, N. & Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition, 19*, 17–36.

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: theoretical underpinnings and practical suggestions. *ELT Journal*, *49*, 133–143.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing 18*(1), 55–88.

Schmitt, D., Schmitt, N., & Mann, D. (2011). *Focus on vocabulary 2: Mastering the Academic Word List*. White Plains, NY: Pearson Education.

Scovel, T. (1998). *Psycholinguistics*. Oxford: Oxford University Press.

Sharwood-Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research, 7*, 118–132.

Sharwood-Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*(2), 165–179.

Sharwood-Smith, M., & Truscott, J. (2014). Explaining input enhancement: a MOGUL perspective. *International Review of Applied Linguistics in Language Teaching, 52*(3). 253– 281.

Shintani, N. (2012). Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research, 16*(2), 253–279.

Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research, 15(2),* 137–158.

Sinaei, M., & Asadi, J. (2014). The impact of two instructional techniques on EFL university learners' academic vocabulary Knowledge: Flash card-based instruction versus wordlists. *International Journal of Language Learning and Applied Linguistics World, 6*(4), 156–167.

Soleimani, H., Ketabi, S., & Talebinejad, M.R. (2008). The effect of output fronted activities to enhance noticing and acquisition of rhetorical structure of contrast paragraphs. *Iranian Journal of Applied Linguistics, 10(2),* 141–166.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

433

Song, S. (2007). Beginning ESL learners' noticing of morphological and syntactic changes in recasts. *TESOL & Applied Linguistics*, *7*(1), 1–25.

Subkoviak, M. J. (1976). Estimating reliability from a single administration of a mastery test. *Journal of Educational Measurement, 13*, 265–276.

Subkoviak, M. J. (1988). A Practitioners Guide to Computation and Interpretation of Reliability Indices for Mastery Tests. *Journal of Educational Measurement, 25*(1), 47–55.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input and second language acquisition* (235–253). Rowley, MA: Newbury House.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. Canadian *Modern Language Review, 50*, 158–164.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H.G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85–113). Cambridge, UK: Cambridge University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.

Swain, M. (2005). The output hypothesis; Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Swan, M. (2005). Legislation by hypothesis: The case of task-based instruction. *Applied Linguistics*, *26*(4), 376–401.

Takač, V. P. (2008). *Vocabulary learning strategies and foreign language acquisition*. Clevedon: Multilingual Matters.

Takimoto, M. (2007). The Effects of Input-Based Tasks on the Development of Learners Pragmatic Proficiency. *Applied Linguistics, 30*(1), 1–25.

Tanaka, T. (1999). The effect of combination of comprehension and production practice in grammar instruction. *JACET Bulletin, 30*, 119–133.

Tanaka, T. (2001). Comprehension and production practice in grammar instruction: Does their combined use facilitate second language acquisition*? JALT Journal, 23*, 6–30.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomlin, R., & Villa, V. (1994). Attention in Cognitive Science and Second Language Acquisition. *Studies in Second Language Acquisition, 16*(2), 183–203.

Toth, P. D. (2006). Processing Instruction and a Role for Output in Second Language Acquisition. *Language Learning, 56*(2), 319–385.

VanPatten, B. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition*. Norwood, NJ: Ablex.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41,* 609–624.

Waring, R. (2004). A study of receptive and productive learning from word cards. *Studies in Foreign Languages and Literature*, *21*, 94–114.

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

434

Wong, W. (2003). Textual enhancement and simplified input: Effects on L2 comprehension and acquisition of non-meaningful grammatical form. *Applied Language Learning*, *13*, 17–45.

SOBRE OS AUTORES:

Seyyed Nasser Mousavi
Department of English Language, Tabriz Branch, Islamic Azad University, Tabriz, Iran E-mail: mousavi.nasser1981@gmail.com
https://orcid.org/0000-0003-4445-6651

Corresponding Author: **Nasser Ghafoori**
Department of English Language, Tabriz Branch, Islamic Azad University, Tabriz, Iran.
E-mail: ghafoori@iaut.ac.ir
ORCID: http://orcid.org/0000-0002-8476-6534

Mahnaz Saeidi
Department of English Language, Tabriz Branch, Islamic Azad University, Tabriz, Iran
E-mail: m_saeidi@iaut.ac.ir
http://orcid.org/0000-0003-2366-8364

Revista Práxis Educacional, Vitória da Conquista – Bahia – Brasil, v. 16, n. 40, p. 404-435, jul./set. 2020.

435