

# *Young beginning learners' vocabulary learning via input and output tasks: The role of working memory*

Mark Feng Teng ✉

Macao Polytechnic University, Macau SAR, China

<https://orcid.org/0000-0002-5134-8504>

[markteng@mpu.edu.mo](mailto:markteng@mpu.edu.mo)

### Abstract

Working memory (WM) is essential to vocabulary learning. However, limited attention has been paid to young beginner learners' vocabulary development under various task conditions from the perspective of WM. This study investigates how two types of WM – complex WM and phonological short-term memory – may influence two instructional approaches (i.e., input and output tasks) on picking up new words. 93 young learners studying English as a foreign language (EFL) participated in input and output tasks and four vocabulary assessments. These assessments functioned as a pretest, immediate posttest, and delayed posttest. The participants also took two WM tests: an operation span test for complex WM and a word repetition test for phonological short-term memory. The results demonstrated that: (1) both input and output tasks significantly influenced the learning and retention of new words, (2) complex WM did not substantially predict task effects on the learning and retention of new words, and (3) phonological WM had a notable impact on learning and retaining new words under the input and output task conditions. The findings emphasize the role of WM for EFL young learners' vocabulary learning through tasks.

*Keywords:* input tasks; output tasks; vocabulary learning; vocabulary retention, working memory

## 1. Introduction

Instruction in English as a foreign language (EFL) emphasizes the importance of words and lexical expressions. Vocabulary development occurs gradually, encompassing an expansion in vocabulary size and a command of diverse aspects of lexical items (Schmitt, 2010). Acquiring vocabulary in a foreign language can be demanding, especially for young learners. A learner's vocabulary size is expected to grow with exposure to the target language (e.g., during instruction). Substantial exposure is thus required for notable progress (Bergström et al., 2022): Learners cannot obtain lexical items through one-time exposure, as the process of acquiring words and lexical expressions should be accumulative (Henriksen, 2008). Scholars have investigated accumulative vocabulary learning by examining relevant techniques and outcomes. Input and output tasks are two such methods (Duong et al., 2021). However, research comparing how these tasks influence young learners' vocabulary acquisition is limited.

Comparing input and output tasks can engender a rich understanding of the processes involved in vocabulary learning. By exploring two task types, researchers and educators can better determine how learners develop vocabulary knowledge. Measuring task effectiveness can also help educators and curriculum developers tailor their instructional approaches to young learners. Incorporating both input and output tasks into EFL instruction offers a richer, more engaging language learning experience for young learners.

Effective vocabulary learning calls for integrated cognitive and linguistic resources. Individual differences, for example, working memory (WM), are associated with behavior and outcomes in language learning (Li et al., 2022). WM is a cognitive function that enables people to access and process information for second language (L2) vocabulary development (Atkins & Baddeley, 1998). Acquiring vocabulary may exert cognitive pressure on EFL learners' WM resources. Therefore, learners with different WM capacities might display varying levels of effectiveness in organizing and executing vocabulary learning processes. This disparity can lead to diverse outcomes from input or output tasks. Even though this hypothesis is grounded in theory, it has not been extensively examined in EFL vocabulary development, especially among young beginner learners.

The present research identifies a significant gap in understanding how WM, encompassing phonological short-term memory and complex WM, relates to vocabulary acquisition and retention through input and output tasks. The current research defines vocabulary learning as the process of gaining and broadening knowledge of word forms/meanings, and how they are used within a language. In the present study, vocabulary learning captures aspects including vocabulary recognition and production. Retention refers to the ability to recall previously

learned vocabulary over time. In this research, it reflects the preservation and maintenance of new words in memory. I focus on whether learned words, as measured based on recognition and production, remain accessible in a delayed test. Theoretically, this study extends the literature on WM and vocabulary learning through input and/or output tasks. Pedagogically, results can guide the development of efficient EFL vocabulary teaching by outlining the benefits and constraints of task-based vocabulary acquisition. Educators and curriculum designers who consider WM can create more supportive strategies to optimize language learning for students with diverse cognitive abilities. Unraveling the relationship between WM and vocabulary learning enables teachers to identify areas where learners need further support.

## 2. Literature review

### 2.1. Theories on input- versus output-based instruction

Input tasks facilitate one's sense of word form and meaning connections. Output tasks promote active language production, repetition, and refinement of vocabulary knowledge. Ellis (2012) stated that input-based instruction represents a type of guidance in which input changes; these tasks underscore the need to process meaning. The concept is rooted in Krashen's input hypothesis (1982, 1985, 1998): L2 acquisition occurs when learners grasp the meaning of input. By contrast, production pertains to comprehensible input from conversation partners. Output might not be essential for most language learners. Krashen (1998) proposed three principles in this regard: (1) output is limited; (2) output may not be necessary to achieve high levels of linguistic proficiency; and (3) output does not guarantee language acquisition. VanPatten's (2004) input processing theory also stresses learners' understanding of input and highlights the importance of input-based instruction.

Apart from input-focused instruction, some scholars stress output in L2 acquisition. Swain's (1985) comprehensible output hypothesis exemplifies this perspective. She asserted that "pushed output," which comes into play when people generate spoken language in unfamiliar situations, forces learners to notice language attributes and adjust their output. This revised output enhances language acquisition. The theory acknowledges the importance of input but contends that input alone is insufficient for developing linguistic systems. In accordance with the noticing (Schmidt, 1990) and interaction (Johnson & Johnson, 1990) hypotheses, students must consciously recognize forms and negotiate meanings linked to those forms to transform input into effective acquisition during

interaction. Pushed output is hence vital for drawing learners' attention to form. Put simply, students must be prompted to observe target language structures and modify output. This process relates to DeKeyser's (2007) skill acquisition theory (SAT), according to which L2 acquisition takes place through explicit processes. Subsequent practice converts explicit processes into implicit ones. Based on this theory, language acquisition entails using declarative knowledge (i.e., conscious understanding of facts, concepts, or ideas) and subsequently automating procedural knowledge, that is, the unconscious knowledge of executing a task. This transformation is suggested to be accomplished via the present-practice-production (PPP) approach (Ur, 1996).

## 2.2. Vocabulary learning via input- and output-based instruction

Task-based instruction facilitates vocabulary learning. Ellis et al. (1994) presented the outcomes of a pair of studies involving Japanese high school students who were learning English. Both investigations featured the same multifactorial design. Tests covered listening comprehension and vocabulary learning. The authors found that modified interaction positively influenced participants' comprehension and vocabulary development. Moreover, students who managed to negotiate input meaning achieved higher scores than students who encountered unmodified or simplified input in relation to comprehension and vocabulary development. Students who were given the chance to negotiate meaning performed better when acquiring new items. The process of negotiating meaning might have stimulated the cognitive functions of the participants, leading them to recognize the differences between the input they observed and their personal output. For instance, input-oriented tasks expanded learners' comprehension of specific linguistic forms, resulting in superior learning outcomes.

Multiple empirical studies have revealed that learners typically acquire limited vocabulary through repeated exposure to text (Pellicer-Sánchez, 2017; Teng, 2020; Webb & Chang, 2015). Laufer (2006) evaluated the effectiveness of word-focused exercises for acquiring new L2 words among 158 Israeli high schoolers learning English. During input instruction, students engaged in a reading task and answered comprehension questions. As for output instruction, learners completed word-focused exercises. Participants were expected to provide target words' meanings in the vocabulary test. Learners who completed word-focused exercises performed significantly better than learners in the input-oriented group. Laufer (2006) thus deemed output-based instruction crucial for helping people build vocabulary knowledge. Later studies came to similar conclusions (e.g., Laufer & Rozovski-Roitblat, 2015): Learners who engaged in

word-focused exercises after reading generally demonstrated stronger word knowledge than learners who only participated in reading comprehension. One reason is that, as argued by Teng (2022), a higher level of involvement in word-focused exercises can lead to greater gains in vocabulary learning.

De la Fuente (2006) conducted a study to investigate how three different vocabulary teaching methods, one traditional and two task-based, affect the learning of word forms, meanings, and morphological features. 30 university students with basic Spanish knowledge were divided into three groups: PPP, task-based without explicit form-focused instruction (TB-NEF), and task-based with explicit form-focused instruction (TB-EF). The PPP method encompassed introducing new words (presentation), engaging in both spoken and written exercises (practice), and participating in paired role-playing activities (production). The TB-NEF group was exposed to pretask, task cycle, and task repetition stages. The TB-EF group received the same pretask and task cycle as the TB-NEF lesson but with an explicit stage replacing task repetition. Vocabulary learning was assessed through an oral production test. There were no notable differences in immediate word retrieval tests among the three groups. The TB-NEF and TB-EF groups significantly outperformed the PPP group on delayed tests, but there were no significant differences between the TB-NEF and TB-EF groups. These findings underline the importance of output in vocabulary acquisition. De la Fuente (2002) also investigated three task conditions on L2 learners' vocabulary knowledge. The three task conditions were non-negotiated premodified input, negotiation without "pushed output," and negotiation plus pushed output. Although negotiated interaction positively influenced receptive vocabulary knowledge, only negotiated interaction with pushed output facilitated the learning of receptive and productive vocabulary knowledge.

Shintani (2011) compared input- and production-based instruction among 36 Japanese children learning a set of English concrete nouns. Four vocabulary tests, including a multiple-choice listening test, a category task test, a discrete item production test, and a "same or different" task test, were used to measure word learning. The input and output groups displayed similar rates of vocabulary knowledge development. However, input-based instruction was more pronounced on the category task test than output instruction. Teng (2018) also employed two instructional groups, output-based instruction and input-based instruction, to evaluate students' learning of phrasal verbs. The participants were 90 Chinese tertiary-level students. Output-based instruction that centered on PPP phases led to better learning gains than the input group. Duong et al. (2021) also assessed the effects of input and output task instruction on vocabulary learning. The dependent variables were vocabulary tests that assessed target word use, form recall, meaning recall, and meaning recognition. The tasks focused

on input and output influenced vocabulary acquisition in distinct ways. For example, tasks centered on input improved scores in recalling meanings, while those emphasizing output enhanced the ability to recall word forms.

The aforementioned studies produced varied results regarding input and output tasks, prompting researchers to delve into the need for deeper information processing in a task. Others have reported that input and output tasks do not produce significant variation in vocabulary learning. Additionally, most studies in this vein have revolved around adolescent learners with a certain level of English proficiency; only one study appears to have involved young beginner EFL students (Shintani, 2011). Clark (1993) accentuated the differences between young and adolescent learners, suggesting that young learners might benefit more from input tasks than output tasks. However, adults exhibit greater WM capacity than young learners (Vogan et al., 2016), which raises questions about the advantages of input- and output-based instruction in L2 vocabulary acquisition. Thus, the present study investigated how input and output tasks influence young beginner EFL students' vocabulary learning.

### 2.3. Working memory (WM) and vocabulary learning

The psychological processes underlying L2 vocabulary acquisition have garnered academic attention. WM refers to the limited cognitive resources utilized for temporarily holding and processing information during the execution of additional mental activities (Baddeley, 2003). WM is multifaceted, including a phonological loop for short-term maintenance of auditory or speech-based information and a visuospatial sketchpad for managing visual and spatial information (Baddeley & Hitch, 1974). These two subsystems are managed by the central executive, which oversees the entire process. The intricacies of WM complicate its measurement (Miyake & Friedman, 1998). Assessments typically focus on phonological WM (PWM) and executive WM, also called complex WM (Gathercole & Baddeley, 1993). PWM tasks include nonword repetition (Gathercole, 2006; Gathercole et al., 2001) and digit span (Gary & Macken, 2015). Executive WM tasks consist of reading span (Daneman & Carpenter, 1980), operation span (Conway et al., 2005), and *n*-back (Owen et al., 2005).

The WM system has been identified as crucial for understanding vocabulary knowledge acquisition (Teng, 2023; Teng & Zhang, 2023). Research on young learners' vocabulary development has reinforced this view, given that vocabulary knowledge can be linked to broader cognitive factors of PWM (Gathercole & Baddeley, 1989). Scholars have also connected WM with vocabulary acquisition. For instance, Martin and Ellis (2012) assessed phonological STM in vocabulary learning among 50 native English speakers at a U.S. university. The authors discovered that

phonological STM predicted participants' performance when learning singular vocabulary forms. In another study on young learners (Engel & Gathercole, 2012), WM tests were used to evaluate short-term storage, as well as complex WM capacity. Learners participated in tests in assessing vocabulary, grammar, and literacy. The findings highlighted the importance of phonological STM for first language and L2 vocabulary. Meanwhile, executive WM demonstrated weak or inconsistent predictive value for L2 vocabulary learning. Révész (2012) divided 90 beginner-level secondary school students into recast and nonrecast groups. Participants completed three tasks – a grammaticality judgment task, a written picture description task, and an oral description task – alongside phonological and complex WM tasks. No significant relationship was observed between WM and task performance for the non-recast group. For the recast group, higher reading spans correlated with improved performance on written tests; higher digit and nonword spans were tied to better progress on oral tests. Yang et al. (2017) studied the relationship between WM and vocabulary learning under multiple task conditions. Their sample included 85 university EFL students in China, and WM was evaluated through a reading span task. Results demonstrated the significant effects of WM on vocabulary learning in the comprehension-only and gap-fill groups. However, such effects were not observed in the sentence-writing group. WM did not predict the vocabulary performance at the delayed test for any of the three groups. Liu et al. (2021) investigated how WM influences vocabulary learning from input enhancement. Input enhancement was essential for vocabulary learning. However, WM modulated such effects. Teng (2022) assessed WM in vocabulary learning in reading comprehension, reading and gap-fill, and reading and sentence writing. The outcomes supported the pronounced effect of sentence writing on vocabulary learning; WM, as measured via an *n*-back test, significantly predicted this group of learners' vocabulary gains.

In summary, the preceding research highlights the relationship between WM and vocabulary acquisition. WM has varied impacts on vocabulary development – especially with respect to task-based learning. Tasks' characteristics can greatly affect EFL students' WM resources. The influence of WM on vocabulary learning is contingent on task requirements. Their diverse nature may place distinct processing demands on young EFL learners' WM capacities. This implies that more research is needed on the association between WM and vocabulary learning based on task conditions for young learners.

#### 2.4. Young learners

Effective and age-appropriate instructional strategies must be adopted to promote young learners' vocabulary acquisition. In the present study, "young learners" refer

to children in the early stages of their formal education, usually ranging from pre-school age to the early years of primary or elementary school. Young learners are characterized by their developing cognitive, social, and emotional abilities, and they are in the process of acquiring foundational skills in areas such as reading, writing, and arithmetic, as well as beginning to explore broader subject areas. Vocabulary growth is a subject worthy of investigation for young learners (Yeung et al., 2019). This early stage is pivotal for establishing the essential skills and knowledge that form the foundation of their English language education. The aim is to concentrate on discerning the distinctive learning requirements and potential of these young learners in order to deliver English language instruction that is both efficacious and captivating. However, focusing on form during tasks can be challenging for these students. Learners at the beginner level might find it challenging to divert their focus from understanding the meaning to grasping the form (Erlam & Ellis, 2018). Additionally, learners tend to concentrate on aspects of language that are more easily processed. In one study of young learners, Ma and Sin (2015) considered conditions involving either receptive or both receptive and productive exercises. Combining the two types of exercises led to more vocabulary retention than relying on receptive exercises alone. Ma and Sin (2015) also contended that young learners, due to their cognitive immaturity, require explicit instructional direction to effectively acquire vocabulary from exercises. This argument was also noted in McDonald and Reynolds (2023), who argued that learning vocabulary is challenging for very young learners and embedding either semantic or thematic clusters in storybook contexts can be helpful for those learners.

Research on young beginner EFL learners' task-based vocabulary acquisition is scarce. Although primary school EFL textbooks often include output tasks, in which students actively produce language by speaking or writing, few input tasks aim to help learners understand language input. Input tasks are critical for young learners to make sense of the language they encounter so that they may accomplish vocabulary learning outcomes (Shintani, 2011). These tasks offer specific advantages for young learners, particularly in terms of being able to rapidly connect word form and meaning – a core aspect of L2 vocabulary development. Young learners who engage in input tasks can discover ways to process new vocabulary more efficiently. This skill contributes to a solid foundation for later language acquisition. Output tasks, such as speaking or writing activities, can benefit young learners as well. Students actively generate language through these tasks. The accompanying repetition and reinforcement can improve their lexical knowledge (Shintani, 2011). Moreover, output tasks prompt young learners to attend to accurate vocabulary usage and syntactic structures, thus fostering improved language precision and fluency. By incorporating both types of tasks into an instructional approach, young learners can receive a balanced language learning experience.



## 2.5. Rationale and research questions

A thorough examination of existing studies indicates that there has been minimal exploration into how young EFL learners acquire vocabulary through input and output tasks, especially in relation to WM. WM is a critical cognitive function for young language learners as it underpins their ability to develop language skills, retain new vocabulary, comprehend and apply grammar and syntax, and understand spoken language. The capacity of a child's WM is closely linked to their ability to concentrate and maintain attention on language tasks, which in turn affects their ability to learn and use new linguistic information effectively. Recognizing the role of WM can help educators create teaching strategies that align with the cognitive development stages of young learners, facilitating more effective and engaging language learning experiences.

Given the need to identify useful instructional approaches for young EFL learners and the role of WM in vocabulary learning, the present study explores word acquisition (and WM's impact) through two tasks: input-based and output-based. Results expand knowledge in this area. The findings can also guide educators and curriculum developers in applying vocabulary learning strategies for young beginner EFL learners. Two research questions were examined:

1. Do instructional approaches, specifically (a) input-based and (b) output-based, yield different results in vocabulary learning?
2. Is there a correlation between WM and participants' vocabulary learning gains across the examined instructional types?

## 3. Method

### 3.1. Research design

The present study was a quasi-experimental design. An input, an output, and a control group were included. Assessments were pre-, post-, and delayed vocabulary tests. Participants also took two WM tests related to phonological and complex WM. Both experimental groups were exposed to a linear syllabus. The target vocabulary was organized into eight sets, with each set comprising five nouns and three adjectives, introduced in individual sessions. To ensure an even distribution of time between exposure to these words and subsequent testing, students were prompted to review the target words introduced in earlier sessions during each class.

In the output-based group, the vocabulary learning process followed the PPP method, as outlined in the instructions. This approach featured a structured

sequence where learners were explicitly introduced to word form and meaning before engaging in production practice (Shintani, 2011). This guidance enabled them to establish a solid understanding before moving on to practice and production activities.

Input-based instruction aligned with the tenets of this instructional technique (Ellis et al., 2002). This technique prioritized learners' exposure to target words within the contexts of different linguistic elements, such as grammar or syntax. The intention was to provide participants with meaningful, contextualized encounters with the vocabulary items. This way, they could develop a more holistic understanding of words' usage and meaning.

### 3.2. Participants

The participants were soon-to-be first-graders at a primary school located in the southern region of China. Ninety-three Chinese children with minimal prior exposure to English were selected. The sample were 40 boys and 53 girls ( $M_{\text{age}} = 6.82$  years,  $SD = .76$ ). Participants were randomly and evenly allocated to three groups to ensure a thorough understanding of the interventions' effects. Throughout the study, participants did not receive any additional English instruction beyond what was provided as part of the intervention. We collaborated with five teachers who had taught at primary school for at least five years. One teacher served as the lead instructor, while the other four supported individual learners as needed.

### 3.3. Instruction

Students participated in two formal lessons per week, each lasting 40 minutes. These lessons took place over a 4-week period during summer vacation before formal primary school began. Participants mostly received explicit feedback on tasks. This feedback process was given as metalinguistic corrections, with target words' correct forms being provided along with a metalinguistic clue. This approach was chosen for several reasons. First, as Li (2013) argued, feedback mechanisms may significantly influence language analytic ability and WM. By strategically implementing explicit feedback to fully leverage its impact, we can explore its influence on vocabulary learning in young, beginning students. Second, integrating both explicit correction and metalinguistic feedback improved clarity, thereby boosting the likelihood of enhancing vocabulary acquisition through various tasks. The following sections describe each type of instruction.

### 3.3.1. Output-based instruction

The instructor explained the upcoming tasks' objectives to students before the start of each lesson. This step ensured that learners fully understood what they were expected to do. Each lesson followed a specific format, adhering to the three phases of the PPP method, a well-known approach to form instruction (DeKeyser, 1998). This method emphasizes the systematic introduction of language forms followed by guided practice. Students should ultimately apply the learned forms in their own production. For the purposes of this study, each lesson was built around one set of six tasks aligning with the three phases of the PPP method. These tasks were designed to facilitate the explicit introduction of word form and meaning before learners engaged in production practice. Learners were therefore familiar with the target words before attempting to produce them independently. The 6-task set was repeated throughout the eight lessons, with each lesson featuring different target words. This repetition enabled learners to reinforce their understanding while practicing the target vocabulary over time. The items were then more likely to be retained and integrated in participants' language repertoire. The teacher spoke in English and Chinese during lessons to facilitate communication. A bilingual approach was meant to give learners clarification as needed, ensuring that language barriers did not impede their understanding or language production. The teacher offered verbal and nonverbal guidance to assist with participants' language production. Verbal support entailed providing explicit feedback, including explanations, examples, and corrections when necessary. Nonverbal support encompassed the use of gestures, facial expressions, and visual aids to enhance comprehension and encourage learners' active participation. Appendix A contains detailed task descriptions, including instructions and procedures for each.

### 3.3.2. Input-based instruction

Six input tasks were developed based on work by Erlam and Ellis (2019) and Shintani (2016). These tasks were designed to facilitate meaningful interaction and were implemented within a single lesson. The tasks were then repeated over the course of four weeks, with each week featuring different target words to enhance variety and engagement. The task design adhered to Ellis's (2003) definition, which stresses the importance of prioritizing meaning, adding a gap (e.g., an information gap) to stimulate communication, and enabling learners to use linguistic and nonlinguistic resources. These tasks were intended to demonstrate more than correct language use; they were included to promote a deeper understanding of target words and their usage. Each task's goals and procedures were thoroughly

described to participants in Chinese to guarantee understanding. This early explanation afforded the learners a firm sense of task expectations to encourage purposeful execution. The teacher played a pivotal role in supporting the learners throughout: verbal and nonverbal cues were offered to help participants determine word meanings and to rectify errors without disrupting ongoing communication. This approach contributed to a nurturing environment where learners were motivated to participate and take risks in their language production. By structuring the tasks in this manner, we sought to maximize learners' language acquisition by promoting meaningful interaction, providing clear instructions, and offering comprehensive support.

### 3.3.3. Control group instruction

The control group participated in standard lessons conducted by the same instructor who taught the input and output groups. These lessons included three main areas: vocabulary acquisition through English songs, activities based on total physical response, and exercises in tracing and copying the alphabet. Care was taken to avoid teaching any of the target vocabulary words during these sessions.

Overall, to ensure that vocabulary learning outcomes are derived from either input or output tasks, it is important to align tasks with specific learning objectives and language acquisition principles. Firstly, it is crucial to clearly define the learning objectives for each task, specifying whether the focus is on recognition (input) or production (output) of vocabulary. This will help in tailoring the design of each task to emphasize either input or output. For input tasks, the focus should be on activities that involve listening, reading, and comprehension, while for output tasks, the emphasis should be on activities that involve speaking, writing, and active production of language.

Additionally, specific exercises should be developed to align with the learning objectives and the intended focus on input or output. For input tasks, the focus should be on comprehension, recognition, and understanding of vocabulary, while for output tasks, the focus should be on production, application, and active use of vocabulary. Furthermore, teachers provided different correction types, tailored to the specific focus of the task. For input tasks, feedback should be focused on comprehension and understanding, while for output tasks, feedback should address accuracy, fluency, and application of vocabulary in context. Nevertheless, given the inherent complexities of language learning tasks, it is difficult to strictly classify tasks as solely input or output. This presents a limitation in the research design. I attempted to address this limitation by carefully considering the specific

learning objectives associated with each task and by emphasizing the balance of input and output information. Additionally, exercises and teachers' correction were included to capture task features. By acknowledging this limitation and taking a comprehensive approach to task design and assessment, I aimed to mitigate the impact of this challenge on the research outcomes.

### 3.4. Target words

Each of the eight lessons featured eight target words for a total of 64 (see Appendix B). This study focused on nouns and adjectives, two vital word types for young learners' vocabulary acquisition (Gasser & Smith, 1998). The target words included 40 nouns and 24 adjectives. The criteria for choosing these words included: (1) clear, unambiguous meanings for Chinese EFL students, and (2) adjectives that frequently described the chosen nouns. Students had to match picture cards to nouns following specific instructions, with exposure to the adjectives occurring naturally throughout the tasks.

### 3.5. Assessment materials

Vocabulary knowledge consists of interconnected components, such as spelling, morphological understanding, word meaning, collocations, and grammatical knowledge (e.g., Ringbom, 1987). Although students may not develop comprehensive knowledge of a word after only a few encounters, multiple tests were employed here to appraise vocabulary knowledge in relation to the target words (Shintani, 2011; Takimoto, 2009). These assessments captured the degree to which students had mastered various aspects of word knowledge, including form, meaning, and use-related elements (Nation, 2022). This approach was adopted to provide a nuanced sense of learners' progress in vocabulary acquisition. By using several evaluations, we could measure multiple aspects of vocabulary knowledge (e.g., recognition and production) and account for possible variations in students' abilities. This method also painted a clearer picture of learners' vocabulary learning while enabling us to examine instructional approaches' effectiveness.

Four tasks were designed to measure learners' vocabulary knowledge, focusing on recognition and production. The recognition aspect was evaluated through a focus on word meaning. The production aspect was assessed through productive knowledge of word form. The evaluations were crafted to thoroughly assess the students' vocabulary knowledge along with their proficiency in understanding and using the language in focus. Two evaluators were invited to verify

the answers. Both came to 100% agreement on the recognition tests and to 96.6% and 97.1% agreement on the two production tests, respectively. Differences were resolved through joint discussion. Cronbach's alpha values for the multiple-choice listening task, listening-and-do task, discrete item production task, and gap-fill production task were .85, .81, .82, and .89, respectively. These values indicate high internal consistency and reliability. Students earned a point for each correct response, while incorrect answers did not receive any points. The teacher offered support by encouraging learners to respond when they struggled to answer a question. The teacher proceeded to the next item if learners could not provide a response.

- a) Multiple-choice listening test for word meaning recognition: Learners had to select the appropriate picture out of four after listening to a recorded word. Based on a pilot study, five seconds were allotted for each question.
- b) Listen-and-do task for word meaning recognition: Participants listened to a story and had to quickly identify the given situations (e.g., noun: zoo, waterfall, rainbow, and bubble; adjective: tall) mentioned in the story.
- c) Discrete item production task: This test used a set of flashcards, and participants had to name each. Answers were considered correct even if the pronunciation was slightly unclear (e.g., "reibow"\* for "rainbow") but were deemed incorrect if incomplete (e.g., "bow"\* instead of "rainbow"). Participants received no points if they failed to name the word on the flashcard or provided an incorrect answer.
- d) Gap-fill production task: Teachers provided different cards with one word per card. Learners had to place the card in the appropriate position. For instance, upon hearing the sentence "A tall boy was walking to a garden," students were required to place the card labeled "tall" at one end of a line and the card for "garden" at the other end.

### 3.6. WM

Phonological STM and complex WM were measured in this study. These two tests were performed as baseline assessments before the treatment.

#### 3.6.1. Phonological short-term memory (STM)

Phonological STM was assessed using a word repetition task adapted from a non-word repetition task (Gathercole et al., 2001). This test targeted only the storage component (Baddeley et al., 2019). Participants first listened to 22 sequences of

target words. They then attempted to repeat all words in each sequence. The sequences 4 to 7 one-syllable words. Different from Gathercole et al. (2001), the present study included real English words rather than nonwords. Previous studies have utilized nonwords to reduce the impact of the learners' existing language knowledge on their test results. In the present study, the participants were young, novice learners with no previous knowledge of L2 vocabulary; thus, employing English words would not result in a less precise assessment of phonological short-term memory (STM), unlike what Gathercole et al. (2001) indicated could occur with advanced learners. The stimuli were recorded by a person whose first language is English. The 22 sets of one-syllable words equaled a total of 126 items. Learners were awarded one point per correctly recalled item, with a maximum possible score of 126 points.

### 3.6.2. Complex WM

An O-span task was used to measure complex WM, focusing on the processing and storage components of STM (Turner & Engle, 1989; Unsworth et al., 2005). Throughout the task, participants engaged in mathematical operations while simultaneously memorizing a sequence of unconnected letters. They tackled each math operation and corresponding letter sequentially: announcing the letter, solving the mathematical equation, and then vocalizing the next letter. The math problems involved simple calculations such as " $1 + 1 = 2$ " to accommodate participants' skills. The operation-letter strings were in sets of 2-5 items (i.e., set sizes) that were introduced randomly. After completing a full set of tests, learners had to recall the letters that fit with the presented order. Participants clicked on the letter they were supposed to remember in this step. A criterion of 85% accuracy was set for the math operations to ensure learners did not deliberately focus on remembering letters while neglecting the processing components of STM (Unsworth et al., 2005). Participants completed three practice sets before the formal test. The scoring system depended on the accurate sequence recall of numbers, encompassing a total of 60 designated letters. This test was developed in JavaScript and administered individually via iPad.

### 3.7. Procedure

Ethical approval was obtained from relevant institutional review boards. Prior to beginning the study, we gave guardians detailed information about the purpose, procedures, potential risks, and benefits of this research and outlined the expected time commitment. Guardians were allowed to ask questions and clarify any

concerns. They were also informed that learners' participation was voluntary. They could withdraw from the study at any time. Written consent was obtained from the students, guardians, and teachers to document their agreement and understanding. Parents viewed this study as a valuable opportunity for their children to learn English. The study spanned 12 weeks. The first two weeks involved pretest vocabulary assessments along with the WM test, which took 20 minutes. This was done because the treatment might affect students' performance on cognitive tests, for example, the instructional intervention could influence the results of cognitive assessments (Granena & Yilmaz, 2018). Although there was no fixed time for the four vocabulary tests, learners typically spent 40 minutes on them on average. Some learners could not provide feedback on certain items; in such cases, the teachers skipped those items to maintain the pace of the test. Tests were administered starting with vocabulary production followed by the receptive knowledge test. This sequencing was chosen because the production test may provide helpful hints for learners' vocabulary recognition by focusing on word forms (Schmitt & Redwood, 2011). Classroom instruction took place during the fifth to eighth weeks. After the treatment, all students took the post-treatment vocabulary tests, which lasted two weeks. Then, learners had a two-week break before taking the four vocabulary knowledge tests again in the eleventh and twelfth weeks. Learners received no instruction on the target words during this interval. Ten teachers volunteered for research purposes.

### 3.8. Data analysis

Data were analyzed to address each research question. Descriptive statistics were calculated first. Next, differences between the tests and among the three groups were analyzed using multivariate analysis of covariance (MANCOVA), taking into account pretest scores and WM when predicting immediate posttest scores and delayed posttest scores. MANCOVA allows multiple dependent variables to be evaluated simultaneously. Immediate posttest scores and delayed posttest scores constituted dependent variables in this case. By using MANCOVA, researchers can examine the combined effects of independent variables (e.g., treatment or intervention) and covariates (i.e., pretest scores and WM) on multiple dependent variables. This analytic approach provides a comprehensive understanding of how factors interact and contribute to the outcomes of interest. Multiple comparisons were based on post hoc Bonferroni adjustments (alpha level = .05). Our data did not include significant outliers that deviated from the usual pattern. Both dependent variables were approximately normally distributed. The assumption of homogeneity of regression slopes was examined; results indicated a  $p$  value greater than .05 for the four test sections, signifying that this assumption was met.



#### 4. Results

Table 1 displays descriptive statistics for the four vocabulary tests. Students in the three groups scored fairly low on the pretest. However, input and output groups demonstrated higher scores on the immediate and delayed posttest. The treatment groups' scores on the pretest rose in the posttest, suggesting that the treatment positively affected learners' vocabulary knowledge. A decrease in scores from the immediate posttest to the delayed posttest was noted, suggesting that the benefits of the group intervention did not persist over time. No significant variations appeared in terms of phonological and complex WM.

Table 1 Descriptive statistics

	Pretest		Immediate posttest		Delayed posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Multiple-choice listening test						
IB group	2.77	1.31	44.39	3.09	33.41	3.18
OB group	3.16	1.61	42.58	3.22	32.29	2.81
CG	2.87	1.20	10.32	2.91	4.32	1.72
A listen-and-do task						
IB group	1.39	1.25	34.94	1.78	25.96	2.72
OB group	1.16	1.29	36.74	2.91	26.25	3.08
CG	1.23	1.08	7.03	2.41	2.25	1.09
Discrete item production task						
IB group	.16	.37	29.87	1.64	19.61	3.74
OB group	.19	.40	31.64	2.83	21.03	3.01
CG	.23	.42	2.09	1.73	1.22	.84
Gap-fill production task						
IB group	.00	.00	24.67	2.05	12.77	3.91
OB group	.06	.25	26.00	3.06	14.51	3.17
CG	.06	.25	1.45	1.54	.67	.59
Phonological WM						
IB group	64.58	8.85				
OB group	65.54	5.16				
CG	66.93	7.22				
Complex WM						
IB group	31.96	4.61				
OB group	31.48	5.85				
CG	36.80	6.07				

Note. IB = Input-based; OB = Output-based; CG = Control group

Table 2 lists the multiple-choice listening test results. The group treatment significantly affected both the immediate posttest scores [ $F(2, 92) = 1091.335, p < .001$ , partial  $\eta^2 = .962$ ] and the delayed posttest scores [ $F(2, 92) = 1284.183, p < .001$ , partial  $\eta^2 = .967$ ]. The pretest score, used as a control variable, showed no significant correlation with either the immediate [ $F(1, 92) = .003, p = .959$ ] or delayed posttest scores

[ $F(1, 92) = .160, p = .690$ ]. Phonological working memory, another control variable, demonstrated a significant connection with both the immediate [ $F(1, 92) = 9.528, p < .05, \text{partial } \eta^2 = .099$ ] and delayed posttest outcomes [ $F(1, 92) = 27.658, p < .001, \text{partial } \eta^2 = .241$ ]. The last control variable, complex working memory, did not show a significant relationship with the scores of the immediate [ $F(1, 92) = .078, p = .780$ ] or delayed posttests [ $F(1, 92) = 1.753, p = .189$ ].

Table 2 Tests of between-subject effects (multiple-choice listening test)

		DF	F	p	partial $\eta^2$
Pretest	Immediate	1	0.003	.959	.000
	Delayed	1	0.160	.690	.002
Phonological WM	Immediate	1	9.528	.003	.099
	Delayed	1	27.658	.000	.241
Complex WM	Immediate	1	0.078	.780	.001
	Delayed	1	1.753	.189	.020
Group treatment	Immediate	2	1091.335	.000	.962
	Delayed	2	1284.183	.000	.967

A Bonferroni-adjusted post hoc examination of the multiple-choice listening immediate posttest indicated that the scores for both the input and output groups were significantly superior to those of the control group ( $p < .001$ ). There was also a notable difference in the posttest results between the input and output groups ( $p < .05$ ). Further analysis using the Bonferroni method for the delayed listening posttest revealed that these groups again outperformed the control group significantly ( $p < .001$ ), with no marked difference between the input and output groups' scores on this delayed assessment ( $p = .075$ ). Analysis of simple effects, after applying the Bonferroni adjustment, showed that for all three groups, immediate posttest scores were significantly higher than pretest scores ( $p < .001$ ), and these immediate posttest scores significantly exceeded scores on the delayed posttest ( $p < .001$ ).

Table 3 showcases the outcomes of the listen-and-do task. A pronounced effect of the group intervention was detected both in the immediate [ $F(2, 92) = 1334.971, p < .001, \text{partial } \eta^2 = .968$ ] and delayed [ $F(2, 92) = 1125.703, p < .001, \text{partial } \eta^2 = .963$ ] posttest scores. The pretest score, serving as a control variable, showed no significant correlation with either the immediate [ $F(1, 92) = 1.889, p = .173$ ] or delayed posttest scores [ $F(1, 92) = .105, p = .747$ ]. In contrast, phonological WM, another control variable, displayed a significant association with both the immediate [ $F(1, 92) = 7.061, p < .05, \text{partial } \eta^2 = .075$ ] and delayed posttest results [ $F(1, 92) = 27.658, p < .001, \text{partial } \eta^2 = .297$ ]. However, complex WM, also a control variable, was not significantly linked to scores on the immediate [ $F(1, 92) = .000, p = .995$ ] or delayed posttests [ $F(1, 92) = 1.723, p = .193$ ].

Table 3 Tests of between-subject effects (a listen-and-do task)

		DF	F	p	partial $\eta^2$
Pretest	Immediate	1	1.889	.173	.021
	Delayed	1	0.105	.747	.001
Phonological WM	Immediate	1	7.061	.009	.075
	Delayed	1	36.689	.000	.297
Complex WM	Immediate	1	0.000	.995	.000
	Delayed	1	1.723	.193	.019
Group treatment	Immediate	2	1334.970	.000	.968
	Delayed	2	1125.703	.000	.963

A further analysis using the Bonferroni method for the immediate posttest of the listen-and-do task showed that both the input and output groups outperformed the control group significantly ( $p < .001$ ), with no noticeable difference between the input and output groups themselves ( $p = .078$ ). Further Bonferroni-adjusted analysis for the delayed posttest of the same task confirmed that these treatment groups continued to score significantly above the control group ( $p < .001$ ), without significant differences between the input and output groups at this later stage ( $p = .071$ ). An analysis of simple effects, after applying Bonferroni adjustments, indicated that for all groups involved, there was a significant improvement from pretest to immediate posttest scores ( $p < .001$ ). Moreover, the scores on the immediate posttest were significantly higher than those on the delayed posttest ( $p < .001$ ).

Table 4 Tests of between-subject effects (discrete item production task)

		DF	F	p	partial $\eta^2$
Pretest	Immediate	1	0.139	.711	.002
	Delayed	1	0.993	.322	.011
Phonological WM	Immediate	1	20.560	.000	.191
	Delayed	1	77.356	.000	.471
Complex WM	Immediate	1	0.000	.998	.000
	Delayed	1	3.387	.069	.037
Group treatment	Immediate	2	1666.857	.000	.975
	Delayed	2	749.876	.000	.945

Table 4 details the outcomes for the discrete item production task. There was a notable impact of the intervention on the scores immediately after the task [ $F(2, 92) = 1666.857, p < .001, \text{partial } \eta^2 = .975$ ] and on those measured on a delayed posttest [ $F(2, 92) = 749.876, p < .001, \text{partial } \eta^2 = .945$ ]. The scores before the intervention (used as a control variable) did not show a significant connection with either the immediate [ $F(1, 92) = .139, p = .002$ ] or delayed posttest scores [ $F(1, 92) = .993, p = .011$ ]. The influence of phonological WM (another control variable) was significantly evident in both the immediate [ $F(1, 92) = 20.560, p < .001, \text{partial } \eta^2 = .191$ ] and delayed posttest scores [ $F(1, 92) = 77.356, p < .001, \text{partial } \eta^2 = .471$ ].

$\eta^2 = .471$ ]. However, complex WM (also a control variable) did not show a significant relationship with the scores right after the task [ $F(1, 92) = .000, p = .998$ ] or those measured later [ $F(1, 92) = 3.387, p = .069$ ].

A further Bonferroni post hoc examination for the immediate posttest of the discrete item production task showed that both the input and output groups significantly outperformed the control group ( $p < .001$ ). However, no significant score differences were observed between the input and output groups in the immediate posttest ( $p = .057$ ). Further analysis with Bonferroni adjustments for the delayed posttest of this task demonstrated that the input and output groups again significantly surpassed the control group in scores ( $p < .001$ ), with no significant variances found between the input and output groups in the delayed posttest ( $p = .068$ ). An analysis of simple effects, after applying Bonferroni adjustments, revealed that the scores of the input and output groups in the immediate posttest were significantly greater than their scores in the pretest ( $p < .001$ ), and these immediate posttest scores were also significantly above their scores in the delayed posttest ( $p < .001$ ). For the control group, there were no significant score changes observed between the pretest, the immediate posttest, and the delayed posttest ( $p = .065$ ).

Table 5 outlines the outcomes of the gap-fill production task. The analysis showed a pronounced influence of the group intervention on both the immediate [ $F(2, 92) = 1068.212, p < .001, \text{partial } \eta^2 = .961$ ] and delayed [ $F(2, 92) = 370.595, p < .001, \text{partial } \eta^2 = .895$ ] posttest outcomes. Pretest scores, used as a baseline measure, did not significantly correlate with either the immediate [ $F(1, 92) = 1.818, p = .181$ ] or delayed posttest results [ $F(1, 92) = .000, p = .994$ ]. Phonological WM, considered as another baseline measure, had a significant correlation with both the immediate [ $F(1, 92) = 16.159, p < .001, \text{partial } \eta^2 = .157$ ] and delayed posttest outcomes [ $F(1, 92) = 98.324, p < .001, \text{partial } \eta^2 = .531$ ]. However, complex working memory did not show a significant relationship with the scores of the immediate [ $F(1, 92) = 1.312, p = .255$ ] or delayed posttests [ $F(1, 92) = 2.400, p = .125$ ].

Table 5 Tests of between-subject effects (Gap-fill production task)

		DF	<i>F</i>	<i>p</i>	partial $\eta^2$
Pretest	Immediate	1	1.818	.181	.020
	Delayed	1	0.000	.994	.000
Phonological WM	Immediate	1	16.159	.000	.157
	Delayed	1	98.324	.000	.531
Complex WM	Immediate	1	1.312	.255	.015
	Delayed	1	2.400	.125	.027
Group treatment	Immediate	2	1068.212	.000	.961
	Delayed	2	370.595	.000	.895

An analysis using the Bonferroni method for the immediate posttest of the gap-fill production task demonstrated that scores from both the input and output groups were significantly superior to those of the control group ( $p < .001$ ). There was no significant difference in the immediate posttest scores between these two groups ( $p = .055$ ). Further Bonferroni-adjusted analysis for the delayed posttest of this task showed that the input and output groups continued to outperform the control group significantly ( $p < .001$ ), with no substantial differences between the input and output groups in the delayed posttest scores ( $p = .068$ ). An analysis of simple effects, after applying Bonferroni adjustments, showed that for the input and output groups, there was a significant increase in scores from the pretest to the immediate posttest ( $p < .001$ ), and these posttest scores were also significantly higher than the scores on the delayed posttest ( $p < .001$ ). The control group's scores did not exhibit significant changes across the pretest, immediate and delayed posttest ( $p = .061$ ).

## 5. Discussion

The present study examined the degree to which phonological and complex WM relate to vocabulary learning from input and output tasks. MANCOVA was used to determine if these two types of WM were significantly associated with immediate and delayed posttest scores on four vocabulary knowledge tests. Findings showed that, first, input and output tasks positively influenced vocabulary learning as reflected in recognition and production. Second, complex WM was not a significant predictor, whereas phonological STM played a significant role in predicting the impacts of input and output tasks on learning and retaining new words. These results are interpreted in this section based on the nature of the interventions, prior findings, and related theories.

### 5.1. Effects of input- and output-based instruction on vocabulary learning

Our first research question pertained to how input and output tasks influence vocabulary learning. In this case, the input treatment surpassed the output treatment only for the multiple-choice recognition test. There were no significant differences between the two groups on the other three types of tests. The results support the benefits of input- and output-based instruction (vs. a control group) in vocabulary learning. In line with previous research (de la Fuente, 2002, 2006; Duong et al., 2021), input-based instruction leads to receptive and productive knowledge acquisition. Regarding individual tests, participants in the input group performed better on recognition than on production. Receptive knowledge was thus

gained before productive knowledge, as previously indicated by Shintani (2011). As argued by Teng (2018), one explanation for this outcome is that input-based instruction allows learners to negotiate the meanings of target words during task execution. This process may encourage learners to seek out the meanings of target words. While doing so, learners must infer target items' meanings to fulfill task requirements. Learners' recognition and production of these items may increase accordingly.

The results also supported the effectiveness of output-based instruction for vocabulary learning compared with the control group. Laufer (2006) proposed that word-focused exercises produced significantly better outcomes than exercises focusing solely on vocabulary learning from reading. Laufer and Rozovski-Roitblat (2015) introduced vocabulary-focused exercises following reading sessions and found that these activities led to better outcomes than mere reading comprehension, irrespective of the aspect of vocabulary knowledge being assessed. The process of acquiring lexical items demonstrates the advantage of output-based instruction. This technique appears crucial for teaching and learning vocabulary, especially in developing recognition and vocabulary production. However, the aforementioned studies primarily focused on adult learners, which is why these findings may not generalize to young learners. The said instructional approaches' effectiveness can be inconsistent among young learners. For instance, Ma and Sin (2015) favored a vocabulary learning condition that included reading, as well as receptive and productive exercises, over a condition with only reading and receptive exercises. Conversely, Shintani (2011) found that there were no notable differences in performance on production tests between a group receiving input and another group focused on output. These contrasting findings highlight the need for further investigation involving young learners. The effectiveness of instructional approaches may vary with individuals' age, cognitive development, and learning characteristics.

The results support SAT (DeKeyser, 2007), which suggests that acquiring a second or foreign language mirrors the process of learning a complex skill, emphasizing the importance of practice. It recommends that clear presentation of linguistic elements should precede both structured and open-ended practice to enhance learning. Additionally, the concepts of "noticing" the input (Schmitt, 1990), and "pushed output" (Swain, 1985), which encourage learners to actively recognize linguistic forms and clarify meaning through interaction, are also deemed beneficial. Input may translate into vocabulary acquisition. The output tasks might have compelled learners to cognitively process the target words. This step likely raised participants' awareness of knowledge gaps and directed their attention to input during interaction. Explicit word-focused exercises that target production can help students learn lexical items. The concept of task involvement load (Hulstijn & Laufer, 2001) explains this phenomenon: Tasks that demand higher levels of

cognitive engagement and active participation tend to yield more favorable vocabulary learning outcomes.

However, the advantage of input-based over output-based instruction only applied to the multiple-choice recognition test. There were no significant differences between input and output tasks for the other three test types in the present study. Shintani (2011) pointed out that learners in their input group did not significantly outperform those in the output group on production tests. These findings suggest caution in extrapolating from Anderson's (1993) cognitive theory on the adaptive control of thought when maintaining that skill development is predictable and that skills improve with practice. The role of practice should not be overlooked and warrants closer investigation in the field of vocabulary acquisition. Nonetheless, output-based instruction cannot replace the opportunities for word acquisition that input-based instruction provides. Ample input needs to be generated for practice in target word acquisition, particularly for young EFL learners who may quickly learn new words by contrasting them with familiar ones (Clark, 1993). Gathering data from young EFL learners in classroom-based conditions can offer further insights into input- and output-based instruction. Syllabus designers and teaching professionals can then make better-informed pedagogical choices.

We may also need to consider differences in terms of recognition and production of vocabulary knowledge. Based on SAT (DeKeyser, 2007), procedural knowledge is skill-specific; that is, when gained through practice, this type of knowledge is difficult to transfer from one skill to another (e.g., from recognition to production). This principle leads to a logical question regarding the challenges young EFL learners may face in retrieving and producing newly learned words: Should the experimental treatment truly have led to vocabulary acquisition among learners, a pertinent question arises as to why their scores on the delayed posttest were significantly lower than those on the immediate posttest. One potential explanation could be vocabulary attrition. For instance, without regular reinforcement or application of the newly acquired vocabulary, learners might gradually lose their grasp over these words (Teng, 2020, 2022). This phenomenon suggests that the initial learning spike observed in the immediate posttest scores might not have been sustained over time without continued exposure or practice.

## 5.2. WM and vocabulary learning from tasks

The relationship between WM and language learners' age is crucial to consider when examining vocabulary learning. Our results showed that WM, particularly PWM, influenced learners' recognition and retention of vocabulary knowledge. An unexpected finding was the non-significant predictive effect of complex WM

on input- and output-based instruction for young learners' vocabulary learning. Teng (2022) previously discovered that complex WM played a part in vocabulary learning from sentence writing, a type of output exercise. His study involved university students. The contradictory outcomes could be due to the substantial cognitive load that various tasks impose on young beginner learners' WM resources; vocabulary learning was relatively new to this group. Potentially, these young EFL learners faced obstacles in simultaneously processing language input, activating knowledge from tasks, and retrieving information from long-term memory. Additionally, they may have had limited executive WM capacity for storing and manipulating information while engaging in vocabulary learning tasks that involve input and output.

These findings also contradict previous studies (Cheung, 1996; Yang et al., 2017) where complex executive WM was closely correlated with vocabulary learning performance. Consistent with Engel and Gathercole (2012), complex WM did not play a significant role in vocabulary learning. This discrepancy might be related to the nature of task completion. Task completion demands mainly involved performing exercises based on teachers' instructions and error correction. Learners might not be able to process metalinguistic knowledge and detect errors in information during task completion, resulting in the non-significant effects of complex WM on task vocabulary learning. Reducing the cognitive burden on learners might explain why executive working memory did not significantly predict outcomes in task-based vocabulary learning. In particular, those who underwent output-based instruction received error correction, a form of deductive teaching. This technique involves first introducing a rule to the learners, who are then expected to apply this rule in practical exercises. This deductive approach is frequently contrasted with inductive teaching, where learners observe examples and are tasked with inferring the underlying rule from these examples (Cerezo et al., 2016). Therefore, output-based instruction may have lessened the cognitive load on learners, with practice in production diminishing the influence of complex working memory.

A noteworthy finding is that phonological STM predicted the efficacy of input and output instruction in vocabulary learning. Learners with stronger phonological STM, or the ability to memorize semantically empty sound sequences, were more adept at grasping phonological rules and memorizing vocabulary. By contrast, those with lower phonological STM struggled to discern underlying patterns for vocabulary acquisition. Children with limited WM capacity might only store minimal information, which poses challenges to analyzing the input structure. Their recognition and production of vocabulary knowledge may decline in kind. These possibilities echo earlier work on WM and vocabulary acquisition (e.g.,



Martin & Ellis, 2012). As Ellis (1996) explained, phonological STM helps consolidate phonological features. This process may aid learners in maintaining pertinent information from language input and modulating their cognitive processing. Learners might then recognize linguistic features more easily and gradually automatize vocabulary knowledge.

As Engel and Gathercole (2012) pointed out, there is a need to differentiate between the effects of executive WM and PWM on vocabulary learning. The findings partially supported Révész (2012), who argued that phonological STM contributes to implicit knowledge development; complex WM is vital for explicit knowledge development. The present study suggests that vocabulary acquisition, dependent on the sequential sound patterns of words and their random linkages to meanings, necessitates the use of phonological WM for processing input and facilitating students' learning of new vocabulary. Unlike the high school students in Révész's (2012) study, the young learners in our sample might have been more proficient in retrieving recently processed chunks from STM to understand and produce vocabulary knowledge. Nevertheless, phonological STM does not necessarily have a more substantial impact than complex WM on EFL vocabulary learning. It would be an oversimplification to assert that phonological STM is more critical than complex WM: Varying task conditions may demand different types of working memory for vocabulary learning. This study's results must be considered with caution as measures related to phonological STM still have ambiguous effects on various aspects of vocabulary acquisition.

## 6. Conclusion

Overall, the present study showcased that both input tasks and output tasks had a significant positive impact on vocabulary acquisition and retention among EFL young learners. This highlights the importance of engaging learners in a variety of activities that require both the intake and production of language to enhance vocabulary learning.

Furthermore, the research revealed that complex WM – which involves simultaneous storage and processing of information – did not play a significant role in predicting the success of these tasks in facilitating vocabulary learning and retention. This suggests that the ability to handle multiple pieces of information at once is not a critical factor influencing vocabulary acquisition in the context of these tasks. In contrast, phonological working memory, which pertains to the ability to retain and manipulate sounds within short-term memory, exerted a pronounced influence on vocabulary learning and retention, especially when learners were engaged in both input and output tasks. This underscores the importance

of phonological processing abilities in the context of language learning, indicating that the capacity to work with sounds is more directly related to vocabulary acquisition than the ability to juggle multiple cognitive operations simultaneously. These findings collectively underline the multifaceted nature of WM's role in task-based vocabulary learning for young EFL learners.

There are several limitations to be acknowledged in the present study. First, because we focused on young EFL learners, findings may not generalize to multiple age groups or language proficiency levels. Second, the experimental treatments might not represent the range of possible input and output tasks; thus, the results may not apply to other instructional approaches. Third, participants were given limited time to complete several tests. This constraint could have affected their performance. Although I sought to use tasks that aligned with either input or output, the multifaceted activities for each type of instruction may have affected participants' vocabulary learning outcomes. Fourth, despite our reliance on immediate and delayed posttest scores to assess vocabulary learning, the frequency of exposure to target words in lessons or assessments was not controlled. The evaluations included target word occurrences. Input from the evaluations could have also affected participants' word learning. Finally, we did not examine how teaching approaches influenced different aspects of vocabulary learning, such as implicit and explicit knowledge development. Scholars can use regression analysis to explore the interactions among learners' aptitude (e.g., cognitive ability) and various instructional types. Conducting an aptitude-treatment interaction study, as outlined by DeKeyser (2012), would be particularly valuable. Such research is crucial given that instructional effectiveness is heavily contingent on learners' characteristics.

Based on the study outcomes, language teachers working with young EFL students can consider the following suggestions. First, blending input and output tasks will promote these learners' vocabulary acquisition. Educators ought to offer tasks that provide students with chances to engage with new vocabulary within relevant contexts, enabling them to actively apply and incorporate these words into their own speaking and writing. Second, the results highlighted the role of phonological WM in vocabulary learning and retention. Teachers need to be aware of students with low phonological WM capacity and strive to support their learning. Instructors can use auditory exercises, rhymes, chants, and mnemonic devices to strengthen phonological memory and enhance vocabulary retention. Although we did not identify complex WM as a significant predictor, additional research is warranted to explore its impact on vocabulary learning under different task conditions. Language teachers can contribute to this line of inquiry by integrating task formats that challenge diverse aspects of WM, such as

organizing information, manipulating multiple linguistic elements, and managing cognitive load. Results will clarify the relationship between complex WM and vocabulary learning. Furthermore, educators should be mindful of the potential limitations of relying on input and output tasks for vocabulary acquisition. It is essential to develop a comprehensive approach that includes several vocabulary learning techniques. Teachers can use visual aids, hands-on activities, games, multimedia resources, and interactive technology-based tools to cater to diverse learning styles and foster vocabulary acquisition. Finally, these findings emphasize the need for targeted interventions geared towards learners with varying cognitive abilities. Teachers should assess students' strengths and weaknesses, including WM components, to provide interventions that address personal needs. Learners with lower cognitive capacities may need additional support, scaffolding, repetition, and explicit instruction, whereas those with higher cognitive abilities may benefit from more challenging activities.

## References

- Anderson, J. R. (1993). *Rules of the mind*. Erlbaum.
- Atkins, P. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, 19(4), 537-552.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature, Reviews Neuroscience*, 4, 829-839.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic Press.
- Baddeley, A. D., Hitch, G. J. & Allen, R. J. (2019). From short-term store to multi-component working memory: The role of the modal model. *Memory & Cognition*, 47, 575-588.
- Bergström, D., Norberg, C., & Nordlund, M. (2022). "Words are picked up along the way" – Swedish EFL teachers' conceptualizations of vocabulary knowledge and learning. *Language Awareness*, 31(4), 393-409.
- Cerezo, L., Caras, A., & Leow, R. P. (2016). The effectiveness of guided induction versus deductive instruction on the development of complex Spanish Gústar Structures. *Studies in Second Language Acquisition*, 38(2), 265-291.
- Cheung, H. (1996). Nonword span as a unique predictor of second language vocabulary learning. *Developmental Psychology*, 32(5), 867-873.
- Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge University Press.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., & Wilhelm, O. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769-86. <https://doi.org/10.3758/bf03196772>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19(4), 450-466.
- DeKeyser, R. (1998). Beyond focus on form: Cognitive perspective on learning and practical second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge University Press.
- DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97-113). Lawrence Erlbaum.
- DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, 62(Suppl 2), 189-200. <https://doi.org/10.1111/j.1467-9922.2012.00712.x>
- de la Fuente, M. J. (2002). Negotiation and oral acquisition of L2 vocabulary: The role of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, 24, 81-112.

- de la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research, 10*, 263-295.
- Duong, P.-T., Montero Perez, M., Desmet, P., & Peters, E. (2021). Differential effects of input-based and output-based tasks on L2 vocabulary learning. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée, 24*(3), 120-144.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*, 91-126.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Ellis, R. (2012). *Language teaching research and pedagogy*. Wiley-Blackwell.
- Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus-on-form. *System, 30*, 419-432.
- Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning, 44*, 449-491.
- Engel, M., & Gathercole, E. (2012). Executive and phonological processes in second language acquisition. *Journal of Educational Psychology, 104*, 974-986.
- Erlam, R., & Ellis, R. (2018). Task-based language teaching for beginner-level learners of L2 French: An exploratory study. *Canadian Modern Language Review, 74*, 1-26.
- Erlam, R., & Ellis, R. (2019). Input-based tasks for beginner-level learners: An approximate replication and extension of Erlam & Ellis (2018). *Language Teaching, 52*(4), 490-511.
- Gary, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1-13.
- Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes, 13*, 269-306.
- Gathercole, S. (2006) Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543. <https://doi.org/10.1017/S0142716406060383>
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200-213. [https://doi.org/10.1016/0749-596X\(89\)90044-2](https://doi.org/10.1016/0749-596X(89)90044-2)
- Gathercole, S., & Baddeley, A. D. (1993). *Working memory and language*. Lawrence Erlbaum.
- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 54A*(1), 1-30. <https://doi.org/10.1080/02724980042000002>

- Granena, G., & Yilmaz, Y. (2018). Aptitude-treatment interaction in L2 learning: A research synthesis. *Studies in English Education, 4*, 803-830.
- Henriksen, B. (2008). Declarative lexical knowledge. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Vocabulary and writing in a first and second language* (pp. 22-66). Palgrave Macmillan.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.
- Johnson, K., & Johnson, H. (Eds.). (1999). Interaction hypothesis. *Encyclopedic dictionary of applied linguistics: A handbook for language teaching* (p. 174). Blackwell Publishers.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- Krashen, S. (1998). Comprehensible output? *System, 26*(2), 175-182.
- Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review, 63*, 149-166.
- Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research, 19*(6), 687-711.
- Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *Modern Language Journal, 97*(3), 634-654.
- Li, S., Hiver, P., & Papi, M. (Eds.). (2022). *Routledge handbook of second language acquisition and individual differences*. Routledge.
- Liu, Y. T., Nassaji, H., & Tseng, W. T. (2021). Effects of internal and external attentional manipulations and working memory on second language vocabulary learning. *Language Teaching Research*. <https://doi.org/10.1177/13621688211030130>.
- Ma, Q., & Sin, C. (2015). Teaching young learners English vocabulary with reading-based exercises in a real classroom situation. *Porta Linguarum, 23*, 125-138.
- Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*, 179-413.
- McDonald, J. A., & Reynolds, B.L. (2023). Learning semantic and thematic vocabulary clusters through embedded instruction: Effects on very young English learners' vocabulary acquisition and retention. *Applied Linguistics Review, 14*(5), 1129-1156. <https://doi.org/10.1515/applirev-2020-0102>
- Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Lawrence Erlbaum.

- Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.
- Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*, 46-59.
- Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381-402.
- Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning, 62*, 93-132.
- Ringbom, H. (1987). *The role of the first language in foreign language learning*. Multilingual Matters.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 17-45.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honor of Sylviane Granger* (pp. 173-209). John Benjamins.
- 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*, 137-158.
- Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. John Benjamin.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.
- Takimoto, M. (2009). The effects of input-based tasks on the development of learners' pragmatic proficiency. *Applied Linguistics, 30*(1), 1-25. <https://doi.org/10.1093/applin/amm049>
- Teng, F. (2018). The effect of focus on form and focus on forms instruction on the acquisition of phrasal verbs by Chinese students. *Asian EFL Journal, 20* (2), 136-164.
- Teng, F. (2020). Retention of new words learned incidentally from reading: Word exposure frequency, L1 marginal glosses, and their combination. *Language Teaching Research, 24*(6), 785-812.
- Teng, F. (2022). The roles of second-language proficiency level and working memory on vocabulary learning from word-focused exercises. *RELC Journal*. <https://doi.org/10.1177/00336882221102228>

- Teng, F. (2023). Effectiveness of captioned videos for incidental vocabulary learning and retention: The role of working memory. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2023.2173613>
- Teng, F., & Zhang, D. (2023). The associations between working memory and the effects of multimedia input on L2 vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1021-1049.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2), 127-154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498-505.
- Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge University Press.
- VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-32). Lawrence Erlbaum.
- Vogan, V. M., Morgan, B. R., Powell, T. L., Smith, M. L., & Taylor, M. J. (2016). The neurodevelopmental differences of increasing verbal working memory demand in children and adults. *Developmental Cognitive Neuroscience*, 17, 19-27.
- Webb, S. A., & Chang, A. C.-S. (2015). Second language vocabulary learning through extensive reading: How does frequency and distribution of occurrence affect learning? *Language Teaching Research*, 19(6), 667-686.
- Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System*, 70, 38-49.
- Yeung, S. S., Ng, M. L., Qiao, S., & Tsang, A. (2019). Effects of explicit L2 vocabulary instruction on developing kindergarten children's target and general vocabulary and phonological awareness. *Reading and Writing*, 33(3), 671-689.



## APPENDIX A

### Tasks for instruction

#### *Output-based tasks*

Definitions: Output-based tasks involve activities where learners are required to actively produce or demonstrate their understanding of the target language. These tasks typically involve verbal repetition, matching, and physical interaction with the language being learned. The final stage of these tasks often includes a written component to reinforce learning. These tasks aim to engage learners in active participation and application of their language skills.

#### Task 1: Listen and repeat

- The teacher presented eight flashcards, each containing a target word, to the learners.
- The flashcards and words were presented once, with each flashcard shown to the learners individually.
- The learners repeated the word after the teacher in unison.
- This activity was repeated once, and the final stage involved writing down the words.

#### Task 2: Watch and repeat

- The teacher played an animation video that displayed the eight words as captions.
- The video was played twice for the learners.
- After watching the video, the teacher pointed to the words on the flashcards from Task 1, and the learners repeated the words after the teacher.
- The final stage of this task involved writing down the words.

#### Task 3: Match words with flashcards

- Each student received a set of eight flashcards without any words.
- The teacher presented each target word through a PowerPoint, and the students had to match the word with the corresponding flashcard.
- Feedback was provided, and the students were guided to read aloud the word. The final stage was to write down the words.

#### Task 4: Building towers

- This cooperative learning activity involved three to four students in a team practicing words.
- The teacher prepared sight words and attached them to plastic cups.
- Students constructed towers using cups, reading aloud each word during the process. Upon completion, they repeated each sight word while dismantling and storing the cups. The concluding step involved writing the words down.

#### Task 5: Kim's game

- Flashcards are scattered face down in a random arrangement on the table.

- Groups of three to four students gather around the table.
- Each participant, in turn, flips a card, and the group collectively vocalizes the term associated with the image displayed.
- Should a participant fail either to identify or to misidentify the item, the teacher offers corrective feedback through recasting.
- The activity concludes once every participant has flipped, identified, and documented the words for all the cards.

#### Task 6: Word Wall

- The Word Wall acts as a foundational reference point for the sight words that students acquire throughout the lesson.
- The teacher pastes the flashcards on the wall.
- Three to four students form a group to write the words for the flashcards.
- If an answer is wrong, the teacher provides a recast for correction after the activity is finished.

*The key futures of the output-based tasks are as follows:*

#### Task 1: Listen and repeat

- Presentation of eight flashcards with target words
- Verbal repetition by learners
- Writing down the words as a final stage

#### Task 2: Watch and repeat

- Video presentation of words as captions
- Verbal repetition by learners
- Writing down the words as a final stage

#### Task 3: Match words with flashcards

- Matching target words with flashcards
- Reading aloud with feedback
- Writing down the words as a final stage

#### Task 4: Building towers

- Cooperative learning activity
- Physical interaction with sight words on plastic cups
- Verbal repetition while building towers
- Writing down the words as a final stage

#### Task 5: Kim's game

- Random spread of flashcards on the table
- Group participation and verbal identification of items

- Feedback using recasts
- Writing down the words as a final stage

#### Task 6: Word Wall

- Use of Word Wall acts as a foundational reference point for sight words
- Group writing of words for flashcards
- Feedback provided for corrections after the activity is finished

#### *Input-based tasks*

Definition: Input-based tasks involve activities where learners receive information or stimuli from the teacher or a source, and they are required to process, respond to, or act upon that information in various ways. These tasks often involve listening, observing, and following instructions to complete the given activities.

#### Task 1: Bingo – 2

- A task involving a one-way gap in information
- The teacher vocalizes the items, and participants mark off the corresponding entries on their bingo cards as they hear them

#### Task 2: Flyswat (clothing)

- The teacher displays images of various clothing items
- The class is split into nine groups, with each group having three to four members
- In a turn-based manner, students represent their team, aiming to be the quickest to swat the image that matches the word the teacher announces

#### Task 3: Card-finding game

- Learners listen to the teacher and identify the cards that correspond to the target words
- Two boxes are placed beside them for sorting correct and incorrect answers
- Students count the number of cards in each box, and the student with the most cards in the correct box is the winner

#### Task 4: Pair cards-finding game

- Similar to Task 3
- Each student has a set of flashcards
- Students find pairs of cards that correspond to the teacher's requirements
- Dividing correct and incorrect responses into two separate containers, with the student accumulating the highest number of correct cards emerging as the victor.

#### Task 5: Building a castle

- Three to four students form a team

- Every team receives a castle illustration along with cut-out images of animals or different characters.
- When the teacher announces specific prompts, students arrange the animals or characters in precise locations within the castle following the provided directions.

Task 6: Where is it?

- Three to four students form a team
- Each team has a map and listens to directions
- Students draw routes on their map based on the directions given, requesting clarification if needed

*The key features of the input-based tasks are as follows:*

Task 1: Bingo – 2

- One-way information gap task
- Listening to the teacher who reads aloud the items
- Crossing out items on a bingo sheet as the teacher reads out stimuli

Task 2: Flyswat (clothing)

- Projected picture of clothing
- Class divided into nine teams
- Students alternate in striking the image that matches the prompt voiced by the teacher.

Task 3: Card-finding game

- Listening to the teacher and finding the cards that correspond to the target items
- Sorting correct and incorrect answers in two boxes
- Counting the number of cards in each box to determine the winner

Task 4: Pair cards-finding game

- Finding pairs of cards that correspond to the teacher's statements
- Sorting correct and incorrect answers in two boxes
- Determining the winner based on the number of correct cards

Task 5: Building a castle

- Team-based activity with a picture of a castle and cut-out animals or characters
- Placing animals in the castle based on the teacher's stimuli

Task 6: Where is it?

- Team-based activity with a map and listening to directions
- Drawing routes on the map based on the given directions, with the option to request clarification if needed

APPENDIX B

A list of target words

Noun: bookshelf, rat, cat, face, eyebrow, earring, elephant, frog, prince, princess, king, queen, desk, chair, backpack, doorman, wolf, mountain, river, water, tree, forest, snow, rain, castle, animal, zoo, building, rooftop, room, waterfall, sunshine, rainbow, color, bubble, meadow, flower, lake, boat, bedsheet (40 words)

Adjectives: beautiful, short, tall, large, long, kind, strong, wide, horrible, green, white, heavy, spectacular, shiny, deep, blue, quiet, clean, muddy, powerful, sharp, comfortable, happy, colorful (24 words)