Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 14 (2). 2024. 311-338
https://doi.org/10.14746/sslIt.30183
http://pressto.amu.edu.pl/index.php/sslIt

The effectiveness of comprehension-based visual arts instruction and production-based flashcard instruction in young English language learners' vocabulary acquisition and retention

Maja Milosavljevic
University of Macau, China
https://orcid.org/0000-0002-7439-1177
mb74866@connect.um.edu.mo

Barry Lee Reynolds ✉
University of Macau, China
https://orcid.org/0000-0002-3984-2059
barryreynolds@um.edu.mo

Abstract

Research on young English language learners has been gaining popularity in recent years, but to this day remains underdeveloped. The present study aimed to add to this body of research by exploring the effects of visual arts activities implemented via comprehension and compared to more commonly used flashcard activities implemented via comprehension-based instruction on young English language learners' vocabulary acquisition and retention. The study specifically focused on very young learners who do not have access to the English language outside of the classroom. This study employed a within-subjects counterbalanced design with young English language learners (N = 47) with a mean age of three years and eight months. The results provided evidence that both instruction types are highly effective in the productive and receptive learning of vocabulary among young English language learners. Furthermore, the positive effects of visual arts activities implemented via comprehension-based instruction on vocabulary retention were found to be significantly higher than those of the production-based instruction.

Keywords: young L2 learners; comprehension-based instruction; production-based instruction; visual arts; vocabulary

1. Introduction

The number of children learning English as a second language (ESL) in an educational setting has been increasing exponentially (Butler, 2015, 2019; Rokita-Jaśkow & Ellis, 2019). In the East Asian context, English lessons are often provided in pre-primary institutions, and children are often being enrolled in private English classes at a very young age (Butler, 2015; Cameron, 2003; Ng & Rao, 2013; Rao, 2013). However, the existing research on young second language (L2) learners often focuses on children who are learning ESL in countries where English is spoken as a first language (e.g., Gordon, 2007). The lion's share of children learning English as an L2 do so in contexts with little exposure to the English language outside of the classroom (e.g., Butler, 2015) and yet the number of studies focusing on pre-primary school children under the age of six who do not have access to English in their social environment is relatively low (Butler, 2015, 2019). More specifically, a close look at 2,200 second language acquisition (SLA) studies conducted over a period of 25 years disclosed that less than one percent of them involved learners at this age (Pichette, 2002). The present study aimed to fill this research gap by focusing on the learning of English by very young learners under the age of six who have little to no exposure to the English language outside the classroom.

Common English teaching practices for very young learners in the East Asian context include approaches that focus on learners' production of the language by either speaking or writing (Ellis, 2003; Ng & Rao, 2013; Shintani et al., 2013). Teachers of young L2 learners most commonly use flashcard games that require these learners to produce the language. Although popular in the young learner classroom, this approach to language instruction might not be the best choice for all young language learners, especially for those at a very young age (Cameron, 2003). While visual arts activities are commonly used in first language (L1) pre-primary classrooms to teach language and literacy, implementing such activities in the young L2 learner classroom is undervalued. Thus, the present study further aimed to determine whether language production through flashcard activities or language comprehension through visual arts is more suitable for the young L2 English learner classroom. Since children learn language by finding meaning in the activities they are involved in (Cameron, 2001; Fleta, 2019; Rokita-Jaśkow & Ellis, 2019), we assumed that visual arts activities implemented via comprehension-based instruction would benefit young L2 learners' vocabulary acquisition and retention.

2. Literature review

2.1. Comprehension-based and production-based instruction

Comprehension-based instruction (CBI) focuses on promoting language acquisition through language input in the written and/or spoken form. It is largely based on the input hypothesis proposed by Krashen (1982, 1985, 1991) that claims L2 acquisition occurs when enough *comprehensible input* is provided. Learners are provided with comprehensible input when they are exposed to language that is above their current L2 level, but still comprehensible enough to be understood. Language acquisition happens when learners are able to understand the language they hear/read and without having been informed of any learning targets (Tragant et al., 2016). Krashen's comprehensible input level is often described as *i* + 1, with *i* being the learners' current L2 level, and +1 the next level of their L2 acquisition. Krashen believes that language acquisition can happen without any production, that is, without ever speaking the language. CBI, however, does not prohibit the learners from speaking; rather, participation in language production is voluntary and never forced (Krashen, 2004). Based on Krashen's (1982) input hypothesis, VanPatten proposed the Input Processing Theory. VanPatten (1996, 2004, 2007, 2015) agreed that comprehensible input is necessary for SLA; furthermore, his model of input processing explains that only a part of input becomes *intake.* Intake is the part of the input that has been processed by the learner and is required for making form-meaning connections that will remain accessible to the learners for output purposes. According to VanPatten's theory, the form-meaning connection is established when learners notice the input and are able to comprehend and process the message. One way of implementing CBI that focuses on input processing is through activities that provide enriched input while learners are focused on the activity (Rassaei, 2012). Here, *enriched input* refers to target items (e.g., words) being embedded in the input, allowing for a high exposure to these targets in the target language (Reinders & Ellis, 2009). By enriching the input, the learners may have a higher chance of receiving enriched intake, resulting in higher receptive *and* productive L2 acquisition (Benati, 2017). Providing plentiful opportunities for receiving good-quality input can produce native-like proficiency in very young L2 learners (DeKeyser, 2018). One way of implementing CBI, especially with young learners, is through visual arts activities that require comprehension of target words.

In contrast to CBI, production-based instruction (PBI) focuses on language output and is largely based on the output hypothesis (Swain, 1985). Swain (1985) proposed that, although comprehensible input is necessary, it is not sufficient for L2 acquisition, and that language output is just as important. By focusing on language output, learners have not only opportunities to practice the language,

but also notice the target items and modify their output, which in turn supports language acquisition (Swain, 1985; Swain & Lapkin, 1995). Engaging in language production also allows learners to *test* their language skills, receive feedback, draw from their knowledge to modify the output when needed, or seek out new solutions for the problems they encounter – all of which may result in a higher rate of L2 acquisition (Swain, 1985, 2005; Swain & Lapkin, 1995). PBI provides these opportunities for learners to produce the target language. One way of implementing PBI, especially with young learners, is through flashcard activities that require production of target words (e.g., Shintani, 2011a).

Shintani (2011b) provides a clear distinction between comprehension-based and production-based instruction, stating that "the fundamental difference lies in whether production is voluntary (in CBI) or required (in PBI)" (p. 22). In a meta-analysis of 35 research experiments published in 30 papers between 1991 and 2010, Shintani et al. (2013) examined the effects of the two instruction types on L2 receptive and productive grammar knowledge. The meta-analysis found that both instruction types were beneficial to receptive and productive grammar knowledge when compared to control groups and this effect was still present over time. When examining short-term results, that is, immediate post-tests, CBI provided opportunities for higher receptive knowledge than PBI, while no difference was found for the productive knowledge. When examining long-term results, that is, delayed post-tests, PBI provided opportunities for higher sustained productive grammar knowledge than CBI, while no difference was found for the receptive grammar knowledge. The overall results, however, showed no clear evidence that would support the superiority of one type of instruction over the other; meaning that both instruction types were found to be equally effective. However, whether CBI or PBI would be more effective for very young L2 learners' learning and retention of vocabulary knowledge has yet to be thoroughly investigated.

2.2. Comprehension-based and production-based instruction with young L2 learners

Studies comparing CBI and PBI effectiveness in vocabulary acquisition, especially that of young L2 learners, are scarce. Shintani (2011a) contrasted vocabulary acquisition and retention of young EFL (English as a foreign language) learners in Japan resulting from comprehension- and production-based vocabulary instruction. 36 participants between the ages of six and eight were divided into two experimental (CBI and PBI) and one control group. CBI and PBI experimental groups received treatment twice a week for a total period of three weeks, while the control group followed a different curriculum and was not exposed to the target vocabulary. The treatment consisted of 30-minute lessons, including comprehension-based tasks

for the CBI group and production-based tasks for the PBI group. The assessment included four vocabulary tests administered at three points in time in the form of a pre-test, post-test, and delayed post-test. Shintani found that both instruction types were equally effective in receptive and productive vocabulary acquisition. Although based on language comprehension, CBI still provided plentiful opportunities for participants to spontaneously produce the language. At the same time, although based on language production, PBI required the participants to comprehend the language in order to participate in the activities. Shintani concluded that, contrary to a common belief, comprehension-based activities "can be successfully implemented in EFL classrooms for young beginners and are at least as effective as production activities where vocabulary learning is concerned" (Shintani, 2011a, p. 156).

In another study, Ma and Sin (2015) investigated the effectiveness of reading-based comprehension and production exercises in young ESL learners' vocabulary acquisition and retention. Twenty-five primary school students from Hong Kong, between the ages of eight and nine, were involved in their within-subjects design study. The participants received treatment once a week in the form of 35-minute lessons, for a total period of four weeks. The participants were involved in two conditions; the first condition included reading lessons supported by receptive language exercises, while the second condition included reading lessons supported by receptive *and* productive language exercises. The assessment was completed at one point in time in the week following each of the two interventions. The results showed the superiority of the latter condition, which included both receptive *and* productive exercises after book reading, suggesting that CBI and PBI should be fused when it comes to teaching young English language learners.

These two studies have investigated learners at the age of six or above, and the results have been conflicting. While it has been noted that PBI is more prevalent in young English language learners' classrooms (Ellis, 2003; Shintani et al., 2013), there is a lack of studies that contrast the two instruction types. There is a need to further explore the benefits of CBI and PBI, especially with learners under the age of six. The current paper aimed to investigate this issue with an intervention study – different visual art activities that focused on language input (CBI) and flashcard activities that focused on language output (PBI) were used with very young learners in a natural classroom setting.

2.3. Arts and second language acquisition

The body of research exploring the impact of visual arts (drawings, paintings, photographs/pictures, sculptures, crafts) on very young learners' vocabulary learning

is scarce. Although visual arts is a very common teaching tool in the young L1 learner classroom, there is a major gap when it comes to investigating its benefits in an L2 setting with very young EFL learners. However, using performing and visual arts to support L2 teaching is not new to ESL teachers; L2 lessons often include songs to support vocabulary learning, or short drama performances to practice the language (Albaladejo et al., 2018; Ludke, 2016). Performing and visual arts activities have been found to support L2 development (Andrade, 1990; Brouillette, 2012; Greenfader & Brouillette, 2013; Ludke, 2016; Shier, 1990). Ludke (2016), for example, investigated the effects of incorporating singing, songwriting, visual arts, and drama activities into regular curriculum for young adolescent Scottish L2 learners of French. He found that the arts activities benefited students' grammar and vocabulary learning, listening and comprehension skills, as well as their productive language and pronunciation. Several studies have investigated the impact of performing arts, such as songs, (e.g., Albaladejo et al., 2018; Coyle & Gracia, 2014; Leśniewska & Pichette, 2014), or drama (e.g., Rieg & Paquette, 2009) on young learners' L2 acquisition. These studies have all shown a positive effect of performing arts on children's L2 development; visual arts, however, have not received much attention (e.g., Andrade, 1990; Shier, 1990), especially in the early years' EFL classroom.

Some studies have reported that, when incorporated into L2 lessons, visual arts activities allow students to physically engage in learning; the focus steers away from the language being taught to the arts activity itself, creating a relaxed, low-anxiety atmosphere where L2 acquisition happens naturally (Bassano & Christison, 1982; Moore et al., 1993; Shier, 1990). Different forms of visual arts can aid students in making a lasting form-meaning connection when learning new vocabulary (Wright, 1989). It has been noted that incorporating student-created artwork, such as pictures and sculptures made by the students, increases their interest and motivation in L2 lessons and boosts their L2 confidence (Bassano & Christison, 1982; Brouillette, 2012; Moore et al., 1993; Shier, 1990; Wright, 1989). Engaging in arts activities promotes children's cognitive development and supports their L2 literacy and language skills (Farokhi & Hashemi, 2012). Creating art pieces nurtures students' creativity (Bassano & Christison, 1982; Farokhi & Hashemi, 2012) and strengthens their language skills (Bassano & Christison, 1982). As students are personally attached to the artwork they create, the conversations inside the classrooms become more personal, meaningful, and relevant to their lives (Andrade, 1990; Bassano & Christison, 1982; Farokhi & Hashemi, 2012; Shier, 1990). Visual arts foster language comprehension and provide resources for children to relate known and new information in a way that is meaningful to them (Viale, 2010). Engaging in visual arts activities is seen as a type of play by children, helping them stay engaged for an extended period of time (Shier, 1990).

Brouillette (2012) and Greenfader and Brouillette (2013) found a positive impact of theater, dance, and visual arts activities on English language learning by primary grade students who participated in a San Diego *Teaching Arts Project*. The project promoted the integration of art-based activities, such as drama projects, singing and dancing, drawing, and painting, into everyday classroom teaching. Primary grade students exposed to teaching through arts for two consecutive years scored higher on an English language development test than the control group students not involved in the project, showing a positive impact of arts on young learners' ESL development. However, this study was completed in the United States, and the participants' exposure to the language outside of the classroom was not measured. This leaves us with the question as to whether or not integrating visual arts into learning would benefit the learners in situations where English is seldom spoken outside of the classroom.

Although a positive influence of integrating visual arts into children's L2 learning has been suggested (e.g., Wright, 2001), to the best of our knowledge no empirical research has investigated its effect on the L2 acquisition of young L2 learners in countries where English is not used in everyday life outside of the classroom. The current study aimed to fill this research gap by examining the impact of visual arts activities on young EFL learners' vocabulary acquisition.

3. Methodology

The current study aimed to fill the research gap found from the literature review by focusing on the language learned by very young EFL learners (under the age of four); there currently is a lack of research dedicated to this age group, even though EFL learning often begins at an even younger age, especially in the East Asian context. Furthermore, the study addressed the need to further explore the benefits of CBI and PBI. The accomplishment of this research goal required a comparison between language learning resulting from production-based instruction and comprehension-based instruction. Visual arts activities were selected as CBI activities as they have been found to be effective in the very young learner L1 classroom but have yet to be assessed in the very young EFL learner classroom. Flashcard activities were selected as PBI activities as they are commonly used in the L2 classroom with very young learners. In the flashcard PBI activities, the very young learners were required to produce the language (e.g., Shintani, 2011a; Swain, 1985; Swain & Lapkin, 1995). In the visual arts CBI activities, the very young learners were required to comprehend the language (e.g., Shintani, 2011a). The current study aimed to answer the following questions:

1. Between comprehension- and production-based vocabulary instruction, which is more effective in increasing young learners' productive and receptive vocabulary knowledge?
2. Between comprehension- and production-based vocabulary instruction, which results in less decay of young learners' productive and receptive vocabulary knowledge?

The study employed a quantitative approach and a within-subjects counterbalanced design was chosen to allow all the participants to experience both types of instruction. Prior to the onset of data collection, a proposal of this research including the methods section was approved by the Research Ethics Committee at the University of Macau under no. SSHRE18-00036-FED.

3.1. Participants

This study took place at a private kindergarten in Guangdong Province, located in Southern Mainland China. At the time this research was conducted, the school had thirty classes separated into different age groups from two- to six-year-olds, and the number of its students amounted to around seven hundred in total. The school's curriculum was based on a half-day English program, meaning that each class spent half the day with a foreign teacher, whose responsibilities included daily English teaching and taking charge of organizing all the activities during their time in class, such as outdoor time, free-play time, snack time, etc. Each foreign teacher taught two classes, one in the morning and the other in the afternoon. The number of the participants at the beginning of the study totaled 53 students. However, since this study focused on EFL teaching and learning, four students were excluded due to their first language being English. Another two students were eliminated as they were absent for over fifty percent of the lessons. The final number of the participants for the current study after excluding invalid data was 47 (females = 16, males = 31). All the participants were in their first year of kindergarten and with no previous exposure to English, with a mean age of three years and eight months. The two intact classes (class 1 and class 2) taught by the first author (N = 34) were chosen for the intervention and each class was randomly divided into two groups, forming two experimental groups: Group A (N = 17) and Group B (N = 17). Group A and Group B of class 1 had English lessons in the mornings, while Group A and Group B of class 2 had English lessons in the afternoons. Group C (N = 13) acted as a control group and consisted of the students from a randomly selected third class who were of the same age as the experimental group and were in their first year of kindergarten.

The recruitment of a control group was necessary to not only determine whether all the experimental group learners started the study with the same level of English knowledge, but also to confirm if any potential effectiveness shown for PBI or CBI was due to the instruction and no other external exposure to the words targeted for assessment in this research. The control group followed the school's regular curriculum and was taught topics different from those covered in the experimental groups throughout the period of this study. All the students participating in the formal study shared the same first language – Mandarin Chinese. The participants were taught in English only; their first language was not used during English lessons.

3.2. Target vocabulary

Requirements for choosing the target vocabulary included the participants' interests, their familiarity with the meaning of the items in their first language, and lexical considerations such as part of speech and syllable count. The participants' interest was determined through interviews with the students' teachers and two topics were chosen before the start of the study: *Safari Animals* and *Fruits and Vegetables*. Each topic consisted of ten target words and five distractors. The distractors were used to foster the participants' motivation throughout the testing procedure (Shintani, 2011a). Since the distractors were familiar to the participants, when providing correct answers for those words their motivation was boosted. For measuring the students' familiarity with the vocabulary, they were asked to name the items on cards in their first language and then in English during the pre-test. Since nouns are found to be easier to learn for children than verbs or adjectives due to their higher imageability rate (Ellis & Beaton, 1993; McDonough et al., 2011) (i.e., due to it being easier for children to visualize objects represented by nouns), only nouns were chosen to be the target vocabulary for this study. Additionally, learning verbs and adjectives in the L2 might be above the cognitive ability of three-year-olds (Albaladejo et al., 2018). Another reason for choosing only nouns was practical; the assessment for students of such a young age may be difficult if other parts of speech are included since it would be challenging to represent such words with images (e.g., adverbs like *always*, verbs like *show*, etc.). Target words were determined according to the pre-test results: in case a potential target word did not meet one or more of the requirements, the word was used as a distractor instead. For example, the word *ostrich* met the requirement of the students' familiarity with the item (meaning that at least 90% of them were familiar with the item as determined through their ability to name the item in their first language) and at least 90% of the students did not know the English word for the item; the word

was therefore chosen to be one of the target words. On the other hand, words like *buffalo* that students were not familiar with, or *elephant* which was known in English by more than 10% of the students, were chosen to be distractors. To keep the tests consistent, all ten target words and five distractors were included in the pre-, post-, and delayed post-test for each topic (see Table 1). Since longer words have been found to be more challenging to learn (Laufer, 1990; Reynolds, 2016), only nouns between two and three syllables were chosen for this study.

Table 1 Target words: Safari Animals (Weeks 1 and 2) and Fruits and Vegetables (Weeks 3 and 4)

Word type	Vocabulary
Target words (Week 1)	zebra (2), hippo (2), camel (2), crocodile (3), kangaroo (3)
Target words (Week 2)	ostrich (2), cheetah (2), rhino (2), hyena (3), gorilla (3)
Distractors	giraffe (2), monkey (2), tiger (2), elephant (3), buffalo (3)
Target words (Week 3)	onion (2), mushroom (2), kiwi (2), tomato (3), broccoli (3)
Target words (Week 4)	pepper (2), cabbage (2), jujube (2), pomelo (3), cucumber (3)
Distractors	orange (2), ginger (2), apple (2), banana (3), potato (3)

Note. The number in parentheses represents the syllable count

3.3. Intervention and data collection

All CBI and PBI lessons began with a brief introduction of the target words: the teacher-researcher showed flashcards with the target word pictures on them and described the target words by using simple explanations the participants were familiar with. After the introduction, CBI lessons moved to visual arts activities focused on language comprehension, while the PBI activities employed flashcards that promoted language production. The students were not required to produce the language in CBI lessons as opposed to PBI lessons. During the CBI activities, the teacher-researcher provided enriched input for the learners by emphasizing the target words, but without asking the participants to repeat the language. During the PBI activities, the participants were instructed by the teacher-researcher to repeat the target words in order to complete the activities. A detailed description of the activities can be found in Milosavljevic and Reynolds (2023). Each lesson lasted for 20 minutes, and each topic was taught for two weeks, making the intervention last a total of four weeks. The participants had English lessons four times a week, making the total number of lessons 32 for each group. The study timeline is shown in Table 2.

Table 2 The timeline for the 14-week period of the study

Time Period	Group		
	A (Experimental)	B (Experimental)	C (Control)
		Pretest (Safari Animals)	
Week 1	CBI (Safari Animals)	PBI (Safari Animals)	School's regular
Week 2	PBI (Safari Animals)	CBI (Safari Animals)	curriculum
End of Week 2		Immediate post-test (Safari Animals) Pre-test (Fruits and Vegetables)	
Week 3	CBI (Fruits & Vegetables)	PBI (Fruits & Vegetables)	School's regular
Week 4	PBI (Fruits & Vegetables)	CBI (Fruits & Vegetables)	curriculum
End of Week 4		Immediate post-test (Fruits & Vegetables)	
Weeks 5-12		School's regular curriculum	
End of Week 12		Delayed post-test (Safari Animals)	
End of Week 14		Delayed post-test (Fruits & Vegetables)	

Data were collected at three points in time. Pre-tests for each topic were completed in the week before the intervention, post-tests were completed immediately after the intervention, and delayed post-tests took place 10 weeks after the intervention. The participants of all three groups followed the school's regular curriculum during the 10-week period before the delayed post-test; the topics used for the current study were not taught to any of the groups during this time.

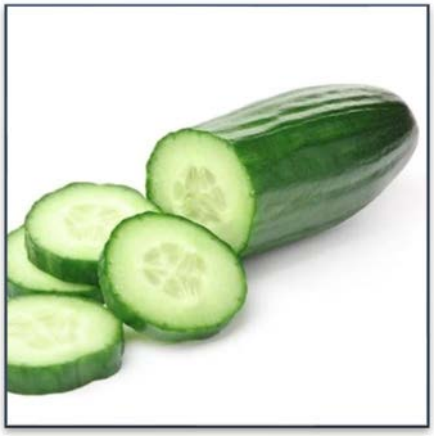

Figure 1 Productive knowledge test example. The figure illustrates a productive knowledge test flashcard for the target word *cucumber*

For assessing productive vocabulary knowledge, a "Name the item" test was used. The teacher-researcher showed a card to the participants with a picture of the target word item on it, asking them, "What's this?". The participants were given five seconds to answer the question; when the participants did not provide the answer, the researcher moved on to the next picture. Each correct

answer for a target item carried one point, with a maximum of ten points total for the test. Each test included ten target words and five distractors; however, results for distractors were excluded from the analysis. Slight mispronunciations, such as using one wrong phoneme (e.g., *proccoli* instead of *broccoli*), were accepted as correct answers, as well as providing the plural instead of the singular form (e.g., *peppers* instead of *pepper*). When the participants mispronounced or failed to produce a syllable (e.g., *kengaa* instead of *kangaroo*), the answer was marked as wrong, and no points were given. If a participant's answer was incorrect, the researcher moved on to the next item. The participants completed the test individually (see Figure 1).

A multiple-choice listening test was chosen for assessing the receptive knowledge of the learners. It included ten target words and five distractors for each of the topics. In this test, the teacher-researcher showed a card with six pictures representing target words and distractors. The use of cards with six pictures for the multiple-choice listening test (similar to Shintani, 2011a) instead of the more commonly used four-picture cards decreased the possibilities of the participants providing the correct answer by chance. The participants were asked to point to the image of the correct target word spoken by the teacher-researcher, for example, "Where's the hippo?". All six images on each of the cards were placed randomly, and an image of the same target word never appeared in the same place on the card as the preceding or the following card, to avoid the participants' usage of elimination techniques or memory to answer the questions. The participants were not required to produce language during this test. Five seconds were given to answer the question; if the participants did not answer it, or gave an incorrect answer, the teacher-researcher moved on to the next picture. The decision to allow five seconds before moving on to the next question was based on Shintani (2011a) that employed a similar design. Additionally, for very young learners, waiting for a long time when not knowing the answer could be stressful and their self-confidence may be affected. The learners may also lose interest quickly as their attention span is relatively short (Cameron, 2001; Fenyvesi, 2020; Reynolds & Teng, 2021); this could in turn affect their motivation to complete the test. Each correct answer carried one point with a maximum of ten points total for the test (results for distractors were excluded from the analysis). For an example item from the receptive vocabulary test, see Figure 2.

The amount of exposure to target words can profoundly influence a learner's acquisition and retention of new vocabulary (Albaladejo et al., 2018; Clark, 1993; Shintani, 2011a). Thus, to ensure the participants' learning was not influenced by any external factors, a control group was used; the control group completed the same pre-, post-, and delayed post-tests as the experimental group. To ensure all the participants received the same amount of in-class exposure, audio recordings

for all the CBI and PBI lessons were collected. The same versions of the two assessments were used for the pre-, post-, and delayed post-tests.

Figure 2 Receptive knowledge test example item. The figure illustrates the receptive knowledge test card for the target word *cucumber*

3.4. Data analysis

Independent samples *t*-tests were used to compare the control and experimental groups' pre-test results to determine whether all the participants started the study with the same knowledge of the target words. Then, due to the difference found, post-test gain scores of the experimental and control group were computed. The post-test gain scores were compared via independent samples *t*-tests to show that the learning of the experimental group was due to the intervention and not out-of-class language exposure. Gain scores for the delayed post-tests were also analyzed in a similar manner. Furthermore, audio recordings of all the lessons were used for counting the tokens of target words produced by the teacher and the participants to measure the exposure to target words inside the classroom. The results of these counts were analyzed to measure whether both instruction types provided equal exposure of the target words to the participants. Finally, to answer the first research question, post-test gain scores were compared using a paired-samples *t*-test. To answer the second research question, delayed post-test gain scores were compared using a paired-samples *t*-test.

4. Results

4.1. External exposure to target words

Independent samples *t*-tests were used to compare the control and experimental groups' pre-test results to determine whether all the participants started the study with the same knowledge. For the productive knowledge test, a Levene's test suggested unequal homogeneity of variance, $F_{(45)} = 35.63$, $p = .000$. Owing to this violated assumption, a *t* statistic not assuming homogeneity of variance was computed, and no significant difference was found between the experimental group ($M = .06$, $SD = .24$, $N = 34$) and its control counterpart ($M = 1.08$, $SD = 2.06$, $N = 13$), $t(12.12) = -1.78$, $p = .101$. For the receptive knowledge test, a Levene's test suggested unequal homogeneity of variance, $F_{(45)} = 24.27$, $p = .000$. Owing to this violated assumption, a *t* statistic not assuming homogeneity of variance was computed, and a significant difference was found between the experimental ($M = .26$, $SD = .90$, $N = 34$) and control group ($M = 1.78$, $SD = 2.24$, $N = 13$), $t(13.5) = -2.35$, $p = .035$. These results indicated that the two groups had the same productive knowledge of the target words at the beginning of the study. However, the control group showed higher receptive knowledge than the experimental group. This difference was controlled by comparing computed gain scores.

An independent samples *t*-test was used to compare the post-test gain scores, which consist of the difference between the pre-test and the post-test scores, of the experimental group and the control group. The maximum score was 20 (i.e., 10 words for each of the two topics, as all the participants received instruction under both conditions). For the productive knowledge test, a Levene's test suggested unequal homogeneity of variance, $F_{(45)} = 12.38$, $p < .01$. Owing to this violated assumption, a *t* statistic not assuming homogeneity of variance was computed, and a significant difference was found between the experimental ($M = 7.00$, $SD = 3.43$, $N = 34$) and the control group ($M = .08$, $SD = .49$, $N = 13$), $t(36.38) = 11.47$, $p < .01$. For the receptive knowledge test, a Levene's test suggested unequal homogeneity of variance, $F_{(45)} = 17.75$, $p < .01$. Owing to this violated assumption, a *t* statistic not assuming homogeneity of variance was computed, and a significant difference was found between the experimental ($M = 13.32$, $SD = 3.51$, $N = 34$) and control group ($M = -.23$, $SD = 1.17$, $N = 13$), $t(44.59) = 19.86$, $p < .01$. These results indicated the gain scores of the experimental group were significantly higher than those of the control group in both productive and receptive knowledge tests, suggesting that the participants of the study had no statistically significant external exposure to the target words outside of the intervention. In addition, the receptive knowledge test results showed a decrease in vocabulary knowledge of the control group participants.

Delayed post-test gain scores, which consist of the difference between the pre-test and the delayed post-test scores, were also investigated to further ensure learning effects were due to the intervention. For the productive knowledge test, a significant difference was found between the experimental (M = 3.15, SD = 3.04, N = 34) and control group (M = 1.15, SD = 1.95, N = 13), $t(45) = 2.19$, $p = .03$. For the receptive knowledge test, a significant difference was found between the experimental (M = 8.97, SD = 3.91, N = 34) and control group (M = 1.46, SD = 2.33, N = 13), $t(45) = 6.47$, $p < .01$. The results are shown in Figure 3. The results indicated the delayed post-test gain scores of the experimental group were significantly higher than those of the control group in both productive and receptive knowledge tests, suggesting that the participants of the study had no statistically significant exposure to the target words outside of the intervention.

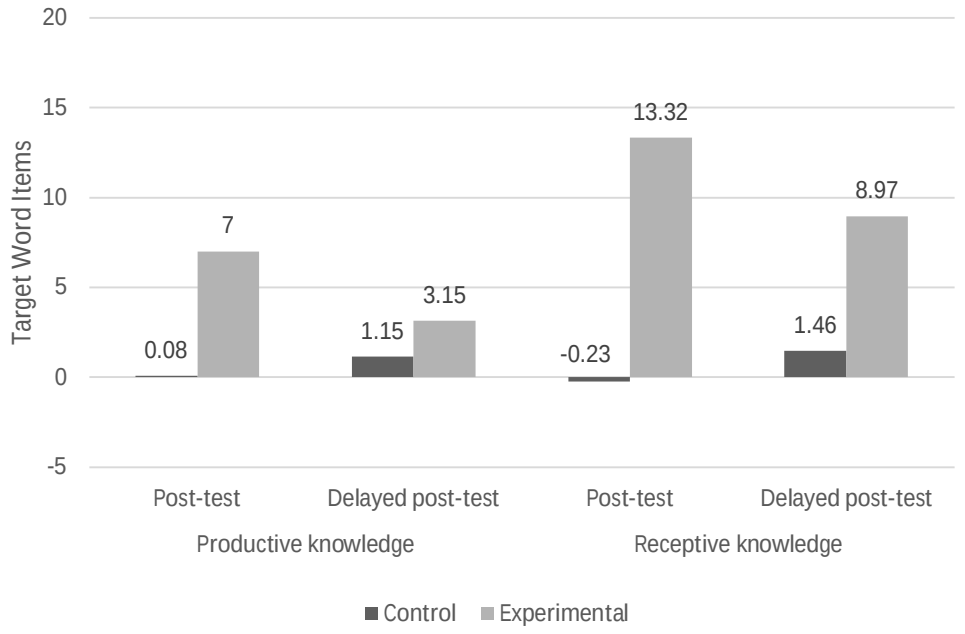

Figure 3 Control and experimental groups' post-test and delayed post-test gain scores

4.2. In-class exposure to target words

Exposure to the target words during class can significantly influence vocabulary learning and recall (Albaladejo et al., 2018), with more exposure leading to the higher likelihood of learning. Audio recordings of all the lessons were used for counting the tokens of target words produced by the teacher-researcher and the participants. The

results of the counts were analyzed to measure whether both instruction types provided equal exposure to the target words for the two experimental conditions.

The first author listened to all the 20-minute English lessons and counted the total number of target word tokens for each lesson. Only the tokens of the target words taught in a particular lesson were counted in the total number of tokens for that lesson. For example, when listening to the first week's lessons, only the tokens of the five target words of the first week were counted (*zebra*, *hippo*, *camel*, *crocodile*, *kangaroo*). An independent-samples *t*-test was used to compare the exposure to the target words between the PBI and CBI experimental conditions. The *t*-test results showed no significant difference between the PBI (M = 231.13, SD = 16.68) and CBI target words' exposure (M = 225.41, SD = 14.61), $t(62) = 1.15$, $p = .15$, indicating that the exposure to the target words was similar in both experimental conditions. Furthermore, an independent-samples *t*-test was used to compare the output of the target words produced by the participants in the CBI and PBI lessons. The *t*-test results showed a significant difference between the PBI (M = 131.34, SD = 18.32) and CBI (M = 72.69, SD = 18.78), $t(62) = -12.65$, $p < .01$, indicating that the participants produced more target words in the PBI lessons.

4.3. Production-based vs. comprehension-based instruction

Since a within-subjects design was used to collect data in the two experimental conditions, the experimental group participants received both types of instruction. The receptive and productive word knowledge assessment data was collected as follows: PBI productive, PBI receptive, CBI productive, and CBI receptive. Each participant completed four pre-tests, four post-tests, and four delayed post-tests. The maximum score for each test was 10, with the minimum being 0.

To answer the first research question, post-test gain scores were compared using a paired-samples *t*-test. For the productive knowledge test, no significant difference was found between PBI (M = 3.65, SD = 2.10, N = 34) and CBI scores (M = 3.35, SD = 1.95, N = 34), $t(33) = .79$, $p = .43$. Further, the computed Cohen's effect size ($d = .13$) was negligible as it failed to meet Cohen's (1988) convention for a small effect ($d = .20$). For the receptive knowledge test, no significant difference was found between PBI (M = 6.91, SD = 2.37, N = 34), and CBI scores (M = 6.41, SD = 1.88, N = 34), $t(33) = 1.20$, $p = .24$. Further, the computed Cohen's effect size ($d = .19$) was negligible as it failed to meet Cohen's (1988) convention for a small effect ($d = .20$). These results indicated that both instruction types were equally effective in increasing young learners' productive and receptive vocabulary knowledge.

To answer the second research question, delayed post-test gain scores were compared using a paired-samples *t*-test. For productive knowledge, a significant

difference was found between PBI (M = 1.09, SD = 1.42, N = 34), and CBI scores (M = 2.06, SD = 2.03, N = 34) $t(33) = 3.24$, $p < .01$. Further, the computed Cohen's effect size (d = .56) suggested moderate practical significance and it exceeded Cohen's (1988) convention for a medium effect (d = .50). For receptive knowledge, a significant difference was found between PBI (M = 3.56, SD = 2.49, N = 34), and CBI scores (M = 5.41, SD = 1.99, N = 34), $t(33) = 4.84$, $p < .01$. Further, the computed Cohen's effect size (d = .83) suggested large practical significance and it exceeded Cohen's (1988) convention for a large effect (d = .80). These results indicated that the CBI produced significantly higher scores in the productive and receptive delayed post-tests than the PBI. The results are shown in Figure 4.

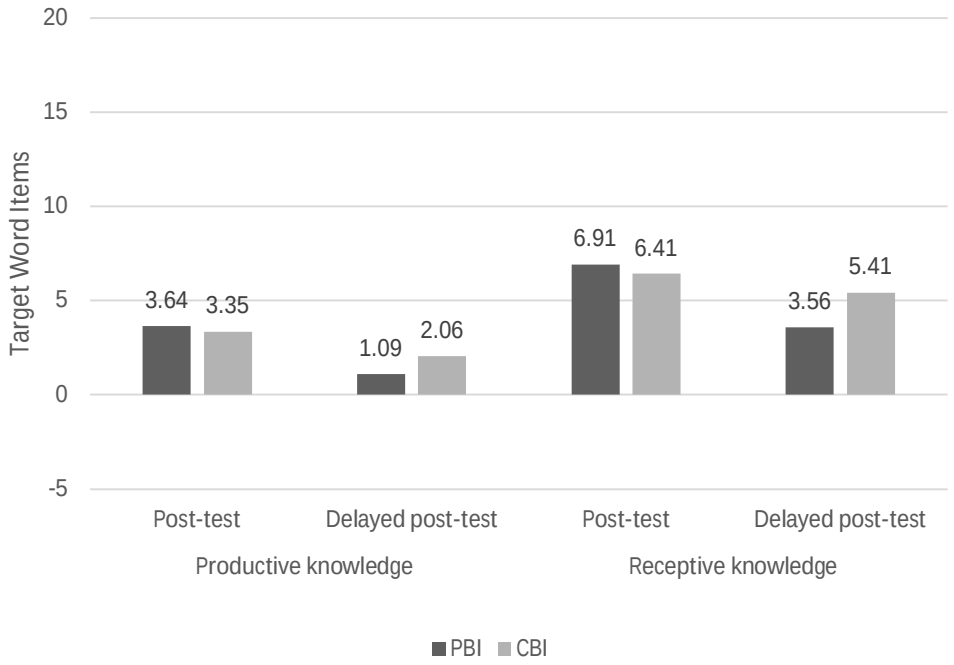

Figure 4 PBI and CBI post-test and delayed post-test gain scores

5. Discussion

5.1. Productive and receptive vocabulary learning

The analysis of post-test gain scores showed that both CBI and PBI have resulted in gains in productive vocabulary knowledge. These results indicated that both instruction types were effective in increasing young L2 learners' productive vocabulary knowledge. Furthermore, no significant difference was found between

the two types of instruction (p = .43), and Cohen's effect size was found to be negligible (d = .19), which brings us to a conclusion that CBI and PBI were *equally* effective for the productive vocabulary learning of young learners. The findings provided support for previous research conducted with older learners (e.g., Ellis & He, 1999) that found CBI led not only to receptive but also productive vocabulary learning. Although CBI did not require the learners to produce the language during the lessons, it had resulted in significant gains in productive vocabulary knowledge. The results are also in line with Shintani (2011a) and support her finding: although focused on vocabulary comprehension, CBI can also positively influence young learners' productive vocabulary learning. Similar to the current study, Shintani (2011a) found that both instruction types can benefit the productive vocabulary learning of young English language learners.

The analysis of post-test gain scores showed that both CBI and PBI resulted in gains in receptive vocabulary knowledge. These results indicated that both instruction types were effective in increasing young learners' receptive vocabulary knowledge. Furthermore, no significant difference was found between the two types of instruction (p = .24), and Cohen's effect size was found to be negligible (d = .19), which brings us to a conclusion that CBI and PBI were *equally* effective for receptive vocabulary learning of young L2 learners. These results are also in line with Shintani's study (2011a) and support her finding: Although focused on language production, PBI can lead to both receptive *and* productive L2 acquisition. Similar to the current study, Shintani (2011a) found that both instruction types can benefit receptive vocabulary learning of young English language learners.

As we can see from the results, two opposing types of instruction, CBI and PBI, have produced comparable levels of achievement in young learners' productive and receptive vocabulary learning. It has been suggested that the level of explicitness in the instruction can influence vocabulary acquisition (e.g., Hulstijn, 1992; Yeung et al., 2019), which could provide an explanation for this finding. The results could have been influenced by the identical introduction to the activities that the two instruction types provided during the intervention. Each CBI and PBI activity began with an introduction of the target words; during the introduction, flashcards representing target word items were presented to the participants, and the same pattern of introducing the items was used regardless of the instruction type.

Another explanation for this finding could be the amount of exposure to the target words in both instruction types. Previous research has found that exposure to the target words can significantly influence L2 vocabulary acquisition (Albaladejo et al., 2018; Clark, 1993; Krashen, 1985; Shintani, 2011a). Although the activities used for the two instruction types in the current study were different in nature, the in-class exposure to target words provided by the teacher was

found to be similar; this could have led to a similar productive and receptive vocabulary acquisition results.

5.2. Productive and receptive vocabulary retention

The analysis of delayed post-test gain scores indicated that both instruction types were effective in retaining young learners' productive vocabulary knowledge. However, CBI results were found to be significantly higher than those of PBI ($p < .01$), suggesting that CBI led to more productive vocabulary knowledge retention. Furthermore, the medium effect size ($d = .53$) implies that CBI had a stronger effect on young learners' productive vocabulary knowledge retention than PBI.

These findings contradict Ma and Sin's (2015) study, in which the combination of comprehension- and production-based exercises resulted in a higher productive vocabulary retention rate than comprehension-based exercises alone. Similarly, the results of the current study do not provide support for Shintani's (2011a) findings. Shintani (2011a) found no significant difference in the delayed post-test scores of CBI and PBI on productive vocabulary retention. These contradicting results could be explained by the different activities and the age of the participants in the two mentioned studies and the current study. Although focused on young learners, participants in Ma and Sin's (2015) study were eight to nine years of age and reading exercises were implemented through CBI and PBI during the intervention. Participants in Shintani's (2011a) study were six to eight years of age and task-based activities were implemented for the CBI, while flashcard activities were used for the PBI in the present study. The present study used visual arts activities with much younger learners – aged three to four; these activities are different in nature from reading or task-based activities. Hence, it is possible that the nature of the visual art activities was responsible for the higher productive vocabulary retention rate in the current study.

Visual arts have been found to instill a sense of pride in children, creating a meaningful connection between them and the artwork they create (Andrade, 1990; Bassano & Christison, 1982; Farokhi & Hashemi, 2012; Shier, 1990). This special connection could have led to a stronger ability to retain the knowledge gained via CBI lessons. Although the participants produced more target words during PBI lessons, it is possible that the language was not fully processed through the flashcard activities, leading to weaker form-meaning connections when compared to CBI.

Some studies that have compared different types of CBI with the use of art (e.g., music vs. storybooks) have only assessed learners' receptive vocabulary retention (e.g., Albaladejo et al., 2018; Leśniewska & Pichette, 2014). This has made it impossible to compare the findings of the current research into productive

vocabulary retention to those studies. It is important for L2 learners to not only understand but also be able to productively use the vocabulary they have learned by either speaking or writing (Nation & Webb, 2011); thus, more studies assessing multiple aspects of vocabulary knowledge are needed.

The analysis of delayed post-test gain scores indicated that both instruction types were effective in helping the participants retain the target vocabulary. However, CBI results were found to be significantly higher than those of PBI ($p < .01$), suggesting that CBI led to higher receptive vocabulary knowledge retention. Furthermore, the large effect size ($d = .80$) implied that the CBI had a stronger effect on receptive vocabulary knowledge retention than the PBI. These findings are in line with Ma and Sin's (2015) study, which found that the comprehension-based exercises resulted in a higher receptive knowledge retention rate than the combined comprehension- and production-based exercises. Similarly, the findings support Shintani's study results (2011a); Shintani found that the CBI participants outperformed PBI participants in the delayed post-test, resulting in a higher receptive vocabulary retention rate. Leśniewska and Pichette (2014) and Albaladejo et al. (2018) also found positive effects of CBI through the use of art, music, and storybooks on young learners' receptive vocabulary retention. The current study has added to the existing research by providing evidence for the positive influence of visual arts activities implemented via CBI on receptive vocabulary retention.

In their first language, children have the ability to assign meaning to novel words quickly, referred to as fast mapping (Bloom, 2000; Clark, 1993). In their L2, however, it can take more than 60 encounters with a new L2 word for very young learners to have a 50% chance of recall (Leśniewska & Pichette, 2014). Having this in mind, it is possible that the visual arts activities have influenced the form-meaning connection by making the learning process more meaningful to children, which, in turn, may have allowed for the connection between form and meaning to last longer. Children take pride in the artwork they make, and the pieces of art become important to them (Andrade, 1990; Bassano & Christison, 1982; Farokhi & Hashemi, 2012; Shier, 1990). This could have further led to a deeper connection between the target words and their meanings, allowing the participants to remember them for a longer period of time and resulting in a larger long-term vocabulary retention rate. It is important to note that the participants in the current study did not have access to their artwork after the lessons as all the art pieces were kept by the researcher. This was done in order to ensure that exposure to target words was limited to the classroom so that it could be comparable to that of the PBI lessons.

Farokhi and Hashemi (2012) argued that visual arts can benefit language learning by providing more visual elements and cues for creating mental images of the vocabulary items in the minds of the learners, thus allowing for deeper

processing of vocabulary items. This could be another explanation for CBI resulting in a higher level of productive and receptive vocabulary retention than PBI. While PBI activities in the current study included flashcards with real-life pictures of target word items, CBI activities included additional imagery elements as a part of the visual arts program. For example, in addition to flashcards used in the introduction part of the activity, one of the CBI activities included a black-and-white print-out of target word items that participants were asked to paint. Thus, it can be argued that the added exposure to the visual representation of the target words could have led to higher retention of those words.

These findings shed new light on the CBI implemented through visual arts activities with young L2 learners, especially those younger than the age of four whose exposure to the English language is limited to the classroom. Visual arts activities are highly suited for the developmental needs of young learners (Wright, 1989, 2001), and the current study provided important insights into how such activities can be implemented in young L2 learners' classroom and result in high rates of newly acquired vocabulary knowledge and retention of this knowledge.

6. Conclusion

The primary aim of this study was to investigate how CBI and PBI affect young English language learners' vocabulary acquisition and retention. As the majority of the research on CBI and PBI has focused on older learners or does not directly compare the two instruction types when focused on younger learners, this study provided an important addition to the literature. It is noteworthy that the existing body of research with young learners often reports retention rate levels measured one to four weeks after the intervention (e.g., Albaladejo et al., 2018; Leśniewska & Pichette, 2014; Ma & Sin, 2015; Shintani, 2011a). Given that vocabulary retention is the ultimate goal of vocabulary learning (Laufer, 1997, 2001), the current study gave an important insight into the benefits of CBI and visual arts activities for young L2 learners.

To our knowledge, the research described in this paper is the first to make the comparison between the effects of visual arts activities carried out via CBI and PBI with L2 learners under the age of six. Although having different theoretical underpinnings, the two types of instruction showed a similar level of effectiveness in young L2 learners' receptive and productive vocabulary learning. However, CBI resulted in significantly higher rates of receptive and productive vocabulary retention than PBI. The results can be attributed to the nature of the activities used; the findings of this study suggested that hands-on visual arts activities could be better suited for young learners than the more commonly used flashcard activities.

Certain limitations need to be mentioned in relation to the current study. First, this study included a relatively small number of participants, which limits the generalization of the findings. However, conducting research with a large number of young learners requires various methodological aspects to be considered, including collecting informed consent forms from the guardians and assuring that the children themselves are willing to participate, as well as carefully reviewing the ethical considerations (Einarsdóttir, 2007). Although a larger sample was desired, the mentioned aspects made including a larger population challenging. Instead, this study implemented a within-subjects design to account for the low number of participants and to control for other intervening learner variables (Larson-Hall, 2016). Second, all the participants came from a similar background – they were all enrolled in one pre-primary institution well-known for its high early-years education standards. Future studies could attempt to conduct similar research with a more diverse sample by including learners with different socio-economic status and backgrounds; this could possibly be accomplished by drawing a sample from different schools. Third, the frequency and type of teacher feedback may differ in CBI and PBI lessons and potentially can have an effect on learning outcomes (Shintani, 2011c). In the current study, feedback was not analyzed; however, the only type of corrective feedback used during the intervention was recast, that is, reformulating errors produced by the participants, which may not have a strong effect on learning compared to other types of corrective feedback (Li, 2013; Lyster & Saito, 2010). Nonetheless, future studies should attempt to address the frequency and type of feedback as a factor when comparing CBI and PBI in the young learner classroom. It should also be mentioned that for reasons of ecological validity and practicality, this study targeted the learning of nouns between two and three syllables (which were a part of the school's curriculum). Having in mind that when learning a second language learners' recall drops by a half for each additional syllable in a word (Campaña Rubio & Ecke, 2001; Pichette, 2002), future studies could either further control this word variable or investigate for its effect in connection with CBI and PBI. Future research could also consider a different scoring system for word knowledge assessments. This study followed Shintani's (2011a) example by accepting slight mispronunciations and plural forms instead of singular forms as correct answers but not accepting missing syllables. It would be worth investigating, for example, the learning and retention of words when each correct syllable is awarded 1 point.

The implementation of visual arts activities via CBI had produced a deeper form-meaning connection in the participants, enabling them to retain the vocabulary knowledge long after the instruction. In the current study, all the carefully designed visual arts activities resulted in artwork made by the participants, which was collected and kept by the teacher-researcher after each CBI lesson. It

would be interesting for future research to explore the learning results when such artwork was kept by the learners instead. Future studies could attempt to compare the vocabulary retention in learners who keep their artwork and share it with their family and friends versus those whose access to the artwork is limited to lesson time only. One assumption would be that the form-meaning connection would have grown even stronger if the participants had prolonged access to the art they had made. Another aspect that future studies could explore is the motivation for learning in young L2 learners involved in visual arts activities. As engaging in visual arts is seen as a form of play participated by young learners, it is possible that this could also affect their motivation to learn (Shier, 1990). It would be interesting to explore the relationship between the activity type, learners' motivation to learn, and vocabulary acquisition.

The current study presented evidence that the use of visual arts activities implemented via CBI can positively affect children's L2 vocabulary acquisition. These findings are encouraging and can be implemented in young L2 learners' curriculum. Curriculum developers should consider including more hands-on activities, such as those used in the current study, to create a more meaningful learning atmosphere for young learners. This study also provided evidence that the learners can acquire productive vocabulary knowledge without being required to produce the language during the lessons. Thus, curriculum developers could try to focus on designing meaningful activities for the learners and steer away from the more commonly used drilling and repetition activities (see Ng & Rao, 2013).

This research was conducted in a real classroom by an experienced kindergarten English teacher with participants who had no previous knowledge of the English language. It is important to note that all the activities had been carefully designed: students' interests were considered, as well as their familiarity with the items represented by the target words. Future teaching practices should consider paying attention to these components, while at the same time ensuring enough comprehensible input is provided during the lessons. Although it has been argued that the implementation of visual arts into the pre-primary curriculum can benefit children's L1 and L2 development (e.g., McArdle & Wright, 2014; Wright, 1989, 2001, 2007), this is not often reflected in teaching practices with young L2 learners; drilling and repetition activities are still widely present in young learners' classroom (Ng & Rao, 2013). That being said, the current study provided evidence that visual arts activities implemented via CBI can be used to facilitate high rates of receptive as well as productive vocabulary learning and retention and should therefore be a vital part of young L2 learner classroom.

References

Albaladejo, S. A., Coyle, Y., & de Larios, J. R. (2018). Songs, stories, and vocabulary acquisition in preschool learners of English as a foreign language. *System, 76*, 116-128.

Andrade, C. (1990). Integrating the arts in the foreign/second language curriculum: Fusing the affective and the cognitive. *Foreign Language Annals, 23*(4), 315-316.

Bassano, S., & Christison, M. A. (1982). *Drawing out*. Alemany Press.

Benati, A. (2017). The role of input and output tasks in grammar instruction: Theoretical, empirical and pedagogical considerations. *Studies in Second Language Learning and Teaching, 7*(3), 377-396. https://doi.org/10.14746/ssllt.2017.7.3.2

Bloom, P. (2000). *How children learn the meanings of words*. MIT Press.

Brouillette, L. (2012). Supporting the language development of limited English proficient students through arts integration in the primary grades. *Arts Education Policy Review, 113*(2), 68-74.

Butler, Y. G. (2015). English language education among young learners in East Asia: A review of current research (2004-2014). *Language Teaching, 48*(3), 303-342.

Butler, Y. G. (2019). Teaching vocabulary to young second- or foreign-language learners. *Language Teaching for Young Learners, 1*(1), 4-33.

Cameron, L. (2001). *Teaching languages to young learners*. Cambridge University Press.

Cameron, L. (2003). Challenges for ELT from the expansion in teaching children. *ELT Journal, 57*(2), 105-112.

Campaña Rubio, E. B., & Ecke, P. (2001). Un estudio experimental sobre la adquisición y recuperación (parcial) de palabras en una lengua extranjera. In G. López Cruz & M. del Carmen Morúa Leyva (Eds.), *Memorias del V Encuentro Internacional de Lingüística en el Noroeste* (pp. 63-84). Editorial Unison.

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Coyle, Y., & Gracia, R. G. (2014). Using songs to enhance L2 vocabulary acquisition in preschool children. *ELT Journal, 68*(3), 276-285.

DeKeyser, R. (2018). Input is not a panacea. *International Journal of Bilingualism, 24*(1), 79-81. https://doi.org/10.1177/1367006918768371

Einarsdóttir, J. (2007). Research with children: Methodological and ethical challenges. *European Early Childhood Education Research Journal, 15*(2), 197-211.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *The Quarterly Journal of Experimental Psychology, 46A*(3), 533-558.

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

Farokhi, M., & Hashemi, M. (2012). The impact/s of using art in English language learning classes. *Procedia - Social and Behavioral Sciences, 31*, 923-926.

Fenyvesi, K. (2020). English learning motivation of young learners in Danish primary schools. *Language Teaching Research, 24*(5), 690-713.

Fleta, T. (2019). From research on child L2 acquisition of English to classroom practice. In J. Rokita-Jaśkow & M. Ellis (Eds.), *Early instructed second language acquisition: Pathways to competence* (pp. 57-79). Multilingual Matters.

Gordon, T. (2007). *Teaching young children a second language*. Praeger Publisher.

Greenfader, C. M., & Brouillette, L. (2013). Boosting language skills of English learners through dramatization and movement. *The Reading Teacher, 67*(3), 171-180.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). Macmillan.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press Inc.

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

Krashen, S. (1991). The Input Hypothesis: An update. In J. E. Alatis (Ed.), *Georgetown University round table on languages and linguistics 1991* (pp. 409-431). Georgetown University Press.

Krashen, S. (2004). Applying the Comprehension Hypothesis: Some suggestions. *The International Journal of Foreign Language Teaching, 1*(1), 21-29.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R*. Routledge. https://doi.org/10.4324/9781315775661

Laufer, B. (1990). Why are some words more difficult than others? – Some intralexical factors that affect the learning of words. *IRAL - International Review of Applied Linguistics in Language Teaching, 28*(4), 293-308. https://doi.org/10.1515/iral.1990.28.4.293

Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge University Press.

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

Leśniewska, J., & Pichette, F. (2014). Songs vs. stories: Impact of input sources on ESL vocabulary acquisition by preliterate children. *International Journal of Bilingual Education and Bilingualism, 19*(1), 18-34.

Li, S. (2013). Oral corrective feedback. *ELT Journal, 68*(2), 196-198. https://doi.org/10.1093/elt/cct076

Ludke, K. M. (2016). Singing and arts activities in support of foreign language learning: An exploratory study. *Innovation in Language Learning and Teaching, 12*(4), 371-386.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA. *Studies in Second Language Acquisition, 32*(2), 265-302. https://doi.org/10.1017/s0272263109990520

Ma, Q., & Sin, C. (2015). Teaching young learners English vocabulary with reading-based exercises in a real classroom situation. *Porta Linguarum, 23*, 125-138.

McArdle, F., & Wright, S. (2014). First literacies: Art, creativity, play, constructive meaning-making. In G. M. Barton (Ed.), *Literacy in the arts: Retheorising learning and teaching* (pp. 21-37). Springer International Publishing. https://doi.org/10.1007/978-3-319-04846-8_2

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science, 14*(2), 181-189.

Milosavljevic, M., & Reynolds, B.L. (2023). Comprehension-based instruction activities and production-based instruction activites. Instructional/Intervention/Teaching/Training materials from "Investigating the effectiveness of comprehension-based and production-based instruction on young English language learners' vocabulary acquisition and retention," [Text/Materials], IRIS Database, University of York, UK. https://doi.org/10.48316/aF5dc-ZE74V

Moore, C. R., Koller, J. A., & Arago, M. K. (1993). The role of art in language learning. *MinneTESOL Journal, 11*, 1-20.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.

Ng, M. L., & Rao, N. (2013). Teaching English in Hong Kong kindergartens: A survey of practices. *International Journal of Literacies, 19*(3), 25-47.

Pichette, F. (2002). Second-language vocabulary learning and the Additivity Hypothesis. *Canadian Journal of Applied Linguistics, 5*(1), 117-130.

Rao, Z. (2013). Teaching English as a foreign language in China: Looking back and forward. *English Today, 29*(03), 34-39.

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.

Reinders, H., & Ellis, R. (2009). The effects of two types of enhanced 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, testing and teaching* (pp. 281-302). Multilingual Matters.

Reynolds, B. L. (2016). Investigating the effects of target word properties on the incidental acquisition of vocabulary through reading. *TESL-EJ: Teaching English as a Second or Foreign Language, 20*(3), 1-31.

Reynolds, B. L., & Teng, M. F. (2021). Editorial: Introduction to the special issue on teaching English reading and writing to young learners. *Studies in Second Language Learning and Teaching, 11*(3), 325-330. https://doi.org/10.14746/ssllt.2021.11.3.1

Rieg, S. A., & Paquette, K. R. (2009). Using drama and movement to enhance English language learners' literacy development. *Journal of Instructional Psychology, 36*(2), 148-154.

Rokita-Jaśkow, J., & Ellis, M. (2019). Introduction. In J. Rokita-Jaskow & M. Ellis (Eds.), *Early instructed second language acquisition: Pathways to competence* (pp. 1-6). Multilingual Matters.

Shier, J. H. (1990). Integrating the arts in the foreign/second language curriculum: Fusing the affective and the cognitive. *Foreign Language Annals, 23*(4), 301-314.

Shintani, N. (2011a). 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.

Shintani, N. (2011b). *A comparison of the effects of comprehension-based and production-based instruction on the acquisition of vocabulary and grammar by young Japanese learners of English* The University of Auckland.

Shintani, N. (2011c). Task-based language teaching versus traditional production-based instruction: Do they result in different classroom processes? *University of Sydney Papers in TESOL, 6*, 97-120.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning, 63*(2), 296-329.

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

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook on research in second language teaching and learning* (pp. 495-508). Routledge.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371-391.

Tragant, E., Muñoz, C., & Spada, N. (2016). Maximizing young learners' input: An intervention program. *Canadian Modern Language Review, 72*(2), 234-257. https://doi.org/10.3138/cmlr.2942

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Ablex.

VanPatten, B. (2004). *Processing instruction: Theory, research, and commentary*. Lawrence Erlbaum Associates.

VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 115-135). Lawrence Erlbaum.

VanPatten, B. (2015). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (2nd ed.) (pp. 113-135). Routledge.

Viale, F. I. (2010). Weaving paintings and language in the EFL kindergarten classroom. XXXV FAAPI Conference – EFL & Art: Learning English with all our senses, Cordoba, Argentina.

Wright, A. (1989). *Pictures for language learning*. Cambridge University Press.

Wright, A. (2001). *Art and crafts with children*. Oxford University Press.

Wright, S. (2007). Graphic-narrative play: Young children's authoring through drawing and telling. *International Journal of Education & the Arts, 8*(8), 1-28.

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.