

Effects of Using Captioned and Keyword-only Pictures on Grade 2 Learners' English Speaking Ability

Nittaya Manuy¹

Chonlada Laohawiriyanon²

Abstract

The purposes of this study were to examine the effects of using captioned and keyword-only pictures on Grade 2 learners' English speaking ability and to examine factors affecting their English speaking ability. The participants were recruited from 49 Grade 2 learners (25 females and 24 males) in a public primary school in Songkla, southern Thailand. They were selected by purposive sampling and randomly assigned into two experimental groups: captioned picture group and keyword-only picture group. The data were collected from pre-, post-, retention tests and semi-structured interview. Following CEFR oral assessment criteria, the results revealed that in the captioned pictures group, Grade 2 learners' English speaking ability significantly increased from poor level to fair level, specifically in 'range' and 'accuracy' ($Z = -2.236$, $p < .05$). For keyword-only picture group, it was found that the learners' speaking performance increased from poor level to fair and good levels at the significant difference 0.01 ($Z = -3.145$, $p < .01$). The three highly improved sub-skills were 'interaction', 'fluency' and 'range'. For further studies, the researcher should use a larger sample size and add a control group. It may well be interesting to compare the effects of using multi-modalities (pictures, keywords and sound) and bi-modalities (pictures and sound) on learners' English speaking ability.

Keywords: Captioned Pictures; Keyword-only Pictures; English Speaking Ability; Dual Coding Theory; Cognitive Load Theory

¹ M. A. student. Teaching English as an International Language Program, Faculty of Liberal Arts, Prince of Songkla University, Hat Yai Campus, Songkla, Thailand, LA_peung23@windowslive.com

² Asst. Prof. Dr., Department of Languages and Linguistics, Faculty of Liberal Arts, Prince of Songkla University, Hat Yai Campus, Songkla, Thailand, chonlada.l@psu.ac.th

ผลของการใช้รูปภาพที่มีคำบรรยายได้ภาพและรูปภาพที่มีเฉพาะคำสำคัญได้ภาพต่อความสามารถด้านการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 2

นิตยา มาน้อย³

ชลลดา เลาหวิริยานนท์⁴

บทคัดย่อ

วัตถุประสงค์ของงานวิจัยนี้คือเพื่อศึกษาผลของการใช้รูปภาพที่มีคำบรรยายได้ภาพและรูปภาพที่มีเฉพาะคำสำคัญได้ภาพต่อความสามารถด้านการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 2 และศึกษาปัจจัยเสริมและอุปสรรคของการใช้รูปภาพทั้งสองประเภทที่มีผลต่อความสามารถทางด้านการพูดกลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 2 ของโรงเรียนประถมศึกษาของรัฐแห่งหนึ่งในจังหวัดสงขลา ภาคใต้ของประเทศไทย จำนวน 49 คน (หญิง 25 คน และชาย 24 คน) ซึ่งได้มาจากการเลือกแบบเจาะจง และสุ่มให้กลุ่มหนึ่งเรียนภาษาอังกฤษโดยใช้รูปภาพที่มีคำบรรยายได้ภาพและอีกกลุ่มหนึ่งเรียนภาษาอังกฤษโดยใช้รูปภาพที่มีเฉพาะคำสำคัญได้ภาพ ข้อมูลได้จากผลการทดสอบก่อน-หลังเรียน ความคงทน และผลการสัมภาษณ์ถึงโครงสร้าง ใช้เกณฑ์การประเมินความสามารถทางการพูดของ CEFR ผลการวิจัยพบว่า กลุ่มที่เรียนภาษาอังกฤษโดยใช้รูปภาพที่มีคำบรรยายได้ภาพมีการพัฒนาทางด้านการพูดเพิ่มขึ้นอย่างมีนัยสำคัญ จากระดับอ่อนเป็นระดับปานกลาง ($Z = -2.236, p < .05$) โดยเฉพาะในด้านคำศัพท์และด้านความถูกต้องในการใช้ภาษา ในขณะที่กลุ่มทดลอง อีกกลุ่มหนึ่งมีความสามารถทางด้านการพูดเพิ่มขึ้นจากระดับอ่อนเป็นระดับปานกลางและระดับเก่งโดยมีทักษะด้านการมีปฏิสัมพันธ์ ด้านความคล่องแคล่วในการใช้ภาษา และด้านคำศัพท์เพิ่มขึ้นสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ($Z = -3.145, p < .01$) ในการทำวิจัยครั้งต่อไปผู้วิจัยควรเพิ่มจำนวนผู้เรียนในกลุ่มทดลองให้มากขึ้น และเพิ่มกลุ่มควบคุม หรือเปรียบเทียบผลของการใช้สื่อที่ให้ข้อมูลภาพ คำศัพท์ และเสียงกับสื่อที่ให้เฉพาะข้อมูลภาพ และเสียงต่อความสามารถทางด้านการพูดของผู้เรียนวัยเยาว์

คำสำคัญ: รูปภาพที่มีคำบรรยายได้ภาพ รูปภาพที่มีเฉพาะคำสำคัญได้ภาพ ความสามารถทางด้านการพูดภาษาอังกฤษ ทฤษฎีรหัสคู่ ทฤษฎีภาระการทำงานของสมอง

³ นักศึกษาปริญญาโท สาขาการสอนภาษาอังกฤษเป็นภาษานานาชาติ คณะศิลปศาสตร์ มหาวิทยาลัยสงขลานครินทร์ LA_peung23@windowslive.com

⁴ ผู้ช่วยศาสตราจารย์ ดร.ภาควิชาภาษาและภาษาศาสตร์ คณะศิลปศาสตร์ มหาวิทยาลัยสงขลานครินทร์ chonlada.l@psu.ac.th

Introduction

English speaking ability is considered one of the four macro language skills necessary for effective communication in any language, especially when speaking to people from different language backgrounds (Madsen, Bowen & Hilferty, 1985). For young learners, aged 7-12 years, speaking is important for their language development. To do so, L1 learners experiment or play with words and sounds through meaning, through processes such as interacting with parents, teachers, or peers or taking part in story telling activities (Colon-Villa, 1997; Cook, 2000; Linse, 2005). When teaching English or a second language, which co-occurs with the development of their mother tongue to young learners, EFL teachers should make sure that learning one language should complement the other by employing suitable teaching methods designed specifically for young learners.

Brown (2001) has suggested seven principles for teaching speaking as follows: 1) activities should cover a wide spectrum ranging from accuracy to fluency, 2) teachers should create learners' intrinsic motivation to learn how to speak, 3) authentic language should be used in a meaningful context, 4) feedback and correction are important elements to foster language learning, 5) teachers should bear in mind a natural link between speaking and listening, 6) learners require opportunities to initiate oral communication, and 7) teachers should encourage learners to use speaking strategies during communication.

To successfully teach children to learn how to speak, one should keep in mind the characteristics of children which play a key role in teaching. According to Slatterly and Willis (2001) and Mackay (2006), there are three main characteristics of young learners. First, their attention spans are around 10-15 minutes. Second, they prefer physical activities such as running, jumping, and dancing. Finally,

while these learners like to be active, they are tired easily. Slatterly and Willis (2001) proposed that young learners can learn by doing and playing. They can learn languages from listening and being involved in activities or experiences in which they are using the languages. Finally, young learners benefit from repeating words, phrases, and sentences many times. With all of these factors in mind, it is suggested that young learners can learn languages from teachers, friends, and others through storytelling (Mackay, 2006).

Among various teaching methods, it is evident that pictures have an important role to play in teaching young children, especially vocabulary, because they can motivate and capture a learner's attention (Mansourzadeh, 2010; Wright, 1990). In the studies of Yoshii (2002) and Al-Ja Afari (2013), they have found that the use of pictures to teach vocabulary has positive effects on learners' attitude and vocabulary retention, while Rowe, Silverman and Mullan (2013) have found that the use of picture-word combinations enhanced four year-old learners' vocabulary knowledge. A study conducted with low proficiency learners reveals that pictorial aids assist learners to retain their knowledge (Yang & Chang, 2013).

Pictures can also be used to teach oral communication. According to Bowen (1982) learners can describe pictures to their partners. Teachers can use pictures to encourage discussion in the classroom and motivate the learners to ask questions creatively based on the pictures. Alternatively, teachers can create interactive conversation using pictures as a prompt. If learners are interested in the details in the pictures, they can ask the teacher questions. Dobson (1992) also stated that the pictures can be good conversation starters and can create different discussions on various topics in the classroom, such as nature, food, classroom, and so on. There are many methods that can support the learners and the teachers to interact with each other by using pictures.

Storytelling through picture books is one of the more interesting ways to teach speaking to young learners. Lever and Sénéchal (2011) suggested using picture books to develop the learners' speaking ability. Teachers can tell a story slowly. During storytelling, the teacher interacts with learners while he/she is telling the story by asking relevant questions using Wh-questions. For example, "Where is/are X?" and "What does X do?". Most importantly, a teacher should constantly provide feedback in the form of correct sentences to learners, whether learners answer correctly and incorrectly. This is to provide them with the correct model of language use. When the teacher finishes telling the story, the learners should retell the story again by using the connective words "and" or "then." Learners who are listening to a story while looking at pictures receive bi-modal inputs, which in turn strengthen their understanding.

Based on the Dual Coding Theory (DCT) proposed by Paivio (1971,1986), effective learning takes place when learners receive bi-or multi-modal inputs through different sensory systems (such as the visual, e.g., pictures or written texts; and the auditory, i.e., verbal) in a learner's working or short-term memory. According to Clark and Paivio (1991) and Mayer (2009), different modalities of input would complement each other rather than interfere. However, it is not always the case that bi/multi-modalities of messages would lead to better learning. Learners with different language proficiencies might have different reactions to such inputs. For instance, young learners who begin to read might benefit more from listening only while adult learners would learn more effectively through reading and listening simultaneously (Sticht & James, 1984; Taylor, 2005).

Another theory that accounts for learning is the Cognitive Load Theory (CLT). This theory influences instructional designs. The main

principle of CLT is that the cognitive capacity in a learner's working memory is limited. If learners are required to perform heavy cognitive tasks, their ability to learn will be lowered (Kalyuga, Chandler & Sweller, 1999). There are three different types of cognitive load: 1) intrinsic, 2) extraneous, and 3) germane. Intrinsic cognitive load has to do with the complexity of content, being vocabulary or grammar. For learning to occur, learners should have sufficient prior knowledge. If instructional materials are unimportant, it is said to cause extraneous cognitive load in learners. Another source of this type of cognitive load is that the materials or inputs are presented in bi-or multi-modes and the messages are not highly relevant. This results in learners facing a cognitive overload situation which negatively affects learners' working memory, and therefore, comprehension ability. If teaching materials are at the right level of learners' ability and relevant to learning objectives, then germane cognitive load created by this combination would automatically enhance learners' schemas (Mayer, 2005).

Given that modalities of input or instructional materials contribute differently to linguistic performance, the present study applies both DCT and CLT in comparing how two different multimodal inputs (captions, pictures and sounds or CP and keywords, pictures, and sounds or KP) affect Grade 2 learners' English speaking ability. This study is driven by two research questions.

1. Are there any differences between the effects of using captioned-pictures and keyword-only pictures on Grade 2 learners' English speaking ability? If so, how?

2. What are facilitating factors and barriers of using captioned-pictures and keyword-only pictures to teach young learners' speaking ability?

Research Methodology

A quasi-experimental research design with two experimental groups was adopted to collect both quantitative and qualitative data. Quantitative data were obtained from three speaking tests: pre, post and retention. Qualitative data were obtained from individual interviews with seven high and seven low achievers.

Participants

The study recruited 49 EFL Grade 2 students (25 females and 24 males), aged 7-9 on average. Three students were from Myanmar and 46 were Thais. They were in a public primary school in Songkhla, southern Thailand. The researcher collected the data in the second semester of the 2015 academic year. The participants represented a homogeneous group based on the pre-test scores. They were chosen by purposive sampling and randomly assigned into two treatment groups: a group in which captioned pictures were used (26 participants) and a group in which keyword-only pictures were used (23 participants).

Instruments

1. Two types of pictures. The researchers selected pictures, following the Basic Education Core Curriculum B.E. 2551 (A.D.2008). The contents of pictures covered numbers, animals, classroom items, sports, occupations, time, clothing, fruit and vegetable, weather, actions, food, shapes and colors, body parts, days and seasons. All pictures were approved by three EFL experts prior to the experiment. They were piloted with students who had similar English proficiency in another primary school. The first type of picture included three modalities, i.e. picture, caption, and sound. The group that received this treatment was called CP. The second type of picture included three modalities, i.e. picture, keyword

only, and sound. This group was referred to as KP. Following the techniques used by Lever and Sénéchal (2011), during speaking activities, the researcher teacher described the picture using basic English structure containing 5-7 words per sentence (Lutz & Huitt, 2003). For example, “The dog is inside his house”. The teacher’s verbal description was exactly the same as that in the caption. Each description was repeated several times so that the participants could repeat it after the teacher. After that, to initiate interactions, the teacher asked Wh-questions such as “Where is the dog?” “What color is the dog?” or “How many dogs can you see in this picture?”. In short, the teacher used exactly the same teaching techniques to teach the two groups. The only difference was that the CP group saw the full captions while looking at the teacher and listening to the teacher uttering the same description as in the caption, whereas the KP group saw only the keyword such as “Dog” while looking at the teacher and listening to the teacher uttering the full description (The dog is inside his house). Finally, the participants were asked to retell the whole story.

The Examples of Captioned and Keyword-only Pictures



Captioned Picture



Keyword-only Picture

2. Speaking test. The same picture was used three times in the pre, post and retention tests. The pre and post tests were 15 weeks apart while the post and retention tests were only two

weeks apart. Each participant was required to describe a picture in two minutes. If he/she could not describe the picture, the teacher would elicit the answers by asking simple questions to help him/her, for example, What is this? What are they doing? All the tests were recorded for grading. The criteria used to examine speaking ability were based on the Common European Framework of Reference (CEFR). There were five sub-skills of oral assessment criteria grid: range, accuracy, fluency, interaction and coherence (University of Cambridge ESOL Examinations Research and Validation Group, 2009). Three raters who were EFL experts (One American and two Thais) were employed to obtain inter-rater reliability. Each rater assessed the participants' performance following the grid. The results obtained from the three raters were averaged.

3. Semi-structured interview. The purpose of the interview was to provide additional information to help interpret the quantitative data regarding factors contributing to the participants' speaking performance as well as problems they encountered while studying. Examples of interview questions were as follows:

1. What do you think about listening to the teacher's description and looking at the caption?
2. Did you look at the caption when you wanted to describe the picture?
3. Can you remember what the caption says?

Data collection procedure

This study consisted of five stages. The entire procedure took 15 weeks. The procedure was as follows. Firstly, the participants took the speaking test individually to assess their baseline knowledge concerning speaking ability. Secondly, the participants were randomly assigned into two groups (CP and KP). Except for the

different mode of captions, the sequence of teaching was all the same. Thirdly, each class was 50 minutes long and met twice a week over 15 weeks. Fourthly, the participants took the post test, followed by individual interviews. Finally, two weeks after the post test, the participants took the retention test.

Data analysis

To answer the first research question, the test results were tallied to arrive at the percentages of participants (poor, fair, and good levels). The percentages were then analyzed using Wilcoxon signed ranks test to examine the learners' English speaking ability in pre, post and retention tests for within group comparison and the differences of English speaking ability between groups were analyzed by employing the Mann Whitney U test. To answer the second research question, qualitative data were analyzed for themes and then classified.

Findings

To compare the results of the pre, post and retention tests of the two groups, descriptive statistics were applied. Table 1 shows the percentages of participants in the two groups who achieved certain levels of English speaking performance in the pre, post and retention tests. The results reveal that 100% of the participants in both groups were all at poor level in the pre-test, suggesting the inability to use English to give details of people or concrete situations by using simple words.

For the post test, the results revealed that the CP group had shown some apparent development in two sub-skills, namely, range and accuracy. To be specific, 35% of the participants achieved fair level of 'range' and 27% of learners reached fair level of 'accuracy'.

As for the KP group, they showed a more remarkable improvement than the CP group. That is, 52% of them reached fair level and 9% reached good level of ‘interaction’. Thirty-nine percent achieved fair level and the 26% were at good level of ‘range’. For ‘fluency’, 39% of the participants reached fair level and 17% reached good level. The results suggest that the participants who looked at the keyword only picture while listening to the teacher describing the picture developed their vocabulary knowledge, could answer questions when asked, and were more fluent. However, they still could not use connective words, such as “and” or “then”, to connect the situations that they saw.

For the retention test, the overall results showed a decline in English speaking ability, especially in the KP group. To be specific, those who demonstrated their ability at good level in the post test only performed at fair level. It is suggested that the influence of the treatment is not long-term.

Table 1: Percentages of participants who achieved different level of English speaking ability in pre, post and retention tests

Skills	Levels	CP Group						KP Group					
		Pre-test		Post-test		Retention Test		Pre-test		Post-test		Retention Test	
		n	%	n	%	n	%	n	%	n	%	n	%
Overall	Poor	26	100	21	81	22	85	23	100	11	48	15	65
	Fair	0	0	5	19	4	15	0	0	6	26	8	35
	Good	0	0	0	0	0	0	0	0	6	26	0	0
Range	Poor	26	100	17	65	12	46	20	87	8	35	8	35
	Fair	0	0	9	35	14	54	3	13	9	39	15	65
	Good	0	0	0	0	0	0	0	0	6	26	0	0
Accuracy	Poor	26	100	19	73	20	77	23	100	12	52	15	65
	Fair	0	0	7	27	6	23	0	0	6	26	8	35
	Good	0	0	0	0	0	0	0	0	5	22	0	0

Skills	Levels	CP Group						KP Group					
		Pre-test		Post-test		Retention Test		Pre-test		Post-test		Retention Test	
		n	%	n	%	n	%	n	%	n	%	n	%
Fluency	Poor	26	100	21	81	23	88	23	100	10	43	17	74
	Fair	0	0	5	19	3	12	0	0	9	39	6	26
	Good	0	0	0	0	0	0	0	0	4	17	0	0
Interaction	Poor	26	100	21	81	21	81	23	100	9	39	12	52
	Fair	0	0	5	19	5	19	0	0	12	52	11	48
	Good	0	0	0	0	0	0	0	0	2	9	0	0
Coherence	Poor	26	100	21	81	24	92	23	100	13	57	15	65
	Fair	0	0	5	19	2	8	0	0	9	39	8	35
	Good	0	0	0	0	0	0	0	0	1	4	0	0

Table 2 below shows the results of Wilcoxon signed-rank test analysis to examine the English speaking of participants within groups.

Table 2: Wilcoxon signed-rank test summary of English speaking ability of each group

Skills	CP				KP			
	Test				Test			
	Pre and Post		Post and Retention		Pre and Post		Post and Retention	
	Z	Asymp. Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)
Overall	-2.236*	0.03	-.577	0.56	-3.145**	.00	-2.352*	.02
Range	-3.000**	0.00	-1.890	0.06	-3.286**	.00	-1.355	.18

Skills	CP				KP			
	Test				Test			
	Pre and Post		Post and Retention		Pre and Post		Post and Retention	
	Z	Asymp. Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)
Accuracy	-2.646**	0.01	-.447	0.65	-3.017**	.00	-1.890	.06
Fluency	-2.236*	0.03	-1.000	0.32	-3.314**	.00	-2.653**	.01
Interaction	-2.236*	0.03	.000	1.00	-3.557**	.00	-1.667	.10
Coherence	-2.236*	0.03	-1.342	0.18	-3.051**	.00	-1.342	.18

Note: * $p < .05$ and ** $p < .01$

As shown in Table 2, a similar pattern of improvements in English speaking ability within groups existed. That is, significant improvements in the post test results of both groups were found overall ($Z = -2.236$, $p < .05$ and -3.145 , $p < .01$ respectively), while there was an overall significant difference in the KP group, indicating that the participants who looked at the keyword-only pictures while listening to the teacher could retain their English speaking ability in the long run.

Table 3 demonstrates the results of the Mann-Whitney U test analysis to examine the differences in English speaking of participants between groups

Table 3: Mann-Whitney U test summary of English speaking ability between the two groups

Tests	Skills	CP	KP	Z	Asymp. Sig. (2-tailed)	Effect Size
		Mean Rank	Mean Rank			
Pre-test	Overall	25	25	0	1.00	0
	Range	23.5	26.7	-1.881	0.06	-0.56
	Accuracy	25	25	0	1.00	0
	Fluency	25	0	1.00	0	
	Interaction	25	25	0	1.00	0
	Coherence	25	25	0	1.00	0
Post-test	Overall	20.63	29.93	-2.701**	0.01	-0.84
	Range	20.44	30.15	-2.628**	0.01	-0.81
	Accuracy	21.92	28.48	-1.879	0.06	-0.56
	Fluency	20.33	30.28	-2.862**	0	-0.9
	Interaction	20.02	30.63	-3.039**	0	-0.96
	Coherence	22.12	28.26	-1.874	0.06	-0.56
Retention Test	Overall	22.77	27.52	-1.56	0.12	-0.46
	Range	23.69	26.48	-0.8	0.42	-0.23
	Accuracy	23.65	26.52	-0.896	0.37	-0.26
	Fluency	23.33	26.89	-1.299	0.19	-0.38
	Interaction	21.71	28.72	-2.108*	0.03	-0.63
	Coherence	21.88	28.52	-2.324*	0.02	-0.7

Note: *p < .05 and ** p<.01

As shown in Table 3, a different pattern of improvements in English speaking performance between groups in pre-test and post-test, showing overall significant improvement of the KP group in post test results (z = -2.701, p < .01). Further, the effect size value (-0.84)

suggested high practical significance, meaning that the influence of using keyword-only pictures on learners' speaking ability was 84%. Based on the results of the post and retention tests, KP group outperformed CP group regarding interaction and coherence ($z = -2.108$, $p < .05$ and $z = -2.324$, $p < .05$). It showed that learners in KP group still retained their speaking ability at A1 level.

Facilitating factors and barriers to the use of captioned-pictures and keyword-only pictures for young learners' speaking ability

The results of the interviews reveal that HA differed from the LA in CP and KP groups in relation to the modality effects caused by the treatments employed in the present study. The HA in both groups reported that they relied on three modality language inputs, i.e., the CP group relied on pictures, captions and sound while the KP group relied on pictures, keywords and sound. If some words in the captions or keywords were difficult, they relied on their auditory sense, i.e., listening to the teacher, as reflected in an excerpt below.

"I looked at...pictures and caption together....looked at easy words....If there were some difficult words, I listened to the teacher. I could read, pronounce and speak out".

HA 1

"I looked at both pictures and captions because I could read all words...repeat after the teacher... I understood and I could speak out".

HA 2

In contrast, LA relied on only two modalities (pictures and teacher's description) because they could not read the captions, as shown in the next excerpt.

“I looked at pictures and keywords but I could not read. I listened to the teacher and repeated after her”.

LA 1

They also reported that the fact that the teacher repeated the utterance several times helped their remembering, thus they could speak quite comfortably.

As for the KP group, HA preferred to be exposed to all three modalities simultaneously; they reported that they relied on listening to the teacher only when they found some vocabulary difficult. If they knew certain vocabulary, they relied only on pictures and described the picture immediately. By contrast, LA reported that they looked at the pictures and listened to the teacher and ignored the keywords, meaning that they preferred only two modalities. Repetitions were also found to be important to this group of participants to be able to provide an oral description of the pictures in English. One problem that HA from both experimental groups had in common was the length of either captions or keywords. It might be difficult for them to remember and speak. For example, the CP group found that “The boy is taking photos.” was too long to remember, and the KP group commented that “taking photos” was problematic.

Discussion

The aims of this study were to examine the effects of using captions and keywords-only pictures on teaching speaking to Grade 2 learners and to study factors affecting their speaking performance. The post test results revealed that Grade 2 learners’ English speaking ability in both groups significantly increased from poor level to fair level but the observed development did not last through to the

retention test. The two sub-skills that the learners improved were 'range' and 'accuracy'.

Interestingly, a significant increase in the speaking performance of the learners in keyword-only picture group was found, improving from poor level to fair and good levels. The three improved sub-skills were interaction, range, and fluency.

The above finding indicated that 'range' was the common sub-skill that both groups of learners could improve. The result was congruent with previous studies which discovered the positive effect of using pictures on young learners' vocabulary learning (Al-Jafari, 2013; Mansourzadeh, 2014; Rasheed and Mohammed, 2007; Rowe, Silverman, & Mullan, 2013; Yoshii and Flaitz, 2002). One plausible explanation for this could be that young learners in this study had had little exposure to English (Chang and Read, 2007; Goh, 1999; Shang, 2008). Therefore, it was not unusual for them to have limited speaking ability and could handle only vocabulary at the beginning of their English education (Hayati and Mohmedi, 2011).

Findings from the KP group revealed that the learners were better able to remember vocabulary from the three modalities (keywords, pictures, and sounds). When comparing the complexity of keywords and captions, it can be seen that the caption is far more complex than single words. Moreover, vocabulary presented in the experiment was concrete and related closely to their background knowledge. Such a finding can be supported by the interview data in which the low English proficiency participants reported that they also looked at pictures and listened to sounds because they could not read. The finding is also in accordance with previous studies which posited that young learners who could not read books were good at listening (Sticht and James, 1984; Taylor, 2005). Viewed from CLT, the KP group had a lighter intrinsic cognitive load. According

to Kalyuga, Chandler & Sweller (1999), and Mayer (2005), the lighter the intrinsic cognitive load, the easier for working memory, which in turn automatically results in learning or creating learners' schemas (Sweller, van Merriënboer and Paas, 1998; Pollock et al, 2002). From the DCT point of view, it can be argued that different modalities of input would complement each other rather than interfere (Paivio, 1986; Clark and Paivio, 1991; Danan, 2004; Mayer, 2009).

In conclusion, it could be said that the speaking ability of the CP group significantly developed from poor to fair. The two sub-skills that they developed the most were 'range' and 'accuracy', indicating that they had better vocabulary knowledge and were more able to use to construct simple English to describe pictures when compared to the pre-test results. In contrast, the KP group's oral production ability significantly increased from poor to good, with 'interaction', 'range', and 'fluency' developing the most. This suggests that they could interact with the teacher more fluently and also had more knowledge in vocabulary. As far as Dual Coding Theory and Cognitive Load Theory is concerned, it could be concluded that for young learners with low English proficiency, multi-modal input with picture, keyword and sound would yield a better learning outcome than when picture, caption and sound are used because they have not yet mastered their reading ability.

Pedagogical recommendations and implications for further studies

Based on findings of the current study, it is recommended that the teacher might use only keywords below the pictures because young learners who are 7-9 years old can learn well through pictures, keywords, and sound. They can remember and speak out easily. Repetition is also meaningful for the beginners who started learning a language. The teacher should repeat words, phrases, or sentences many times while he/she is teaching through the uses of

pictures in the classroom because learners will remember and speak English well. The teacher should also create speaking games by using pictures to promote English speaking ability in classroom. For example, the teacher might use pictures for a whispering game because learners can see pictures and listen to sound from a teacher. After that, they can practice English speaking when they whisper sentences to their classmates several times. Looking at pictures and listening to sound can support them in remembering and speaking out more easily. They also promote new vocabulary learning.

For further studies, the researcher should use a larger sample size for better statistical analysis. Also, including a control group might yield a stronger finding. It may well be interesting to compare the effects of using multi-modalities (pictures, keywords and sounds) and bi-modalities (pictures and sounds) on learners' English speaking ability.

References

- Al-Ja Afari, I. S. (2013). *Using pictures in vocabulary in Grades 5 and 6 classrooms*, Sharqiya South Region.
- Bowen, B. (1982). *LOOK Here! Visual aids in language teaching*. Macmillan Publishers Ltd. London.
- Brown, H. D, (2001). *Teaching by principals*. Pearson Education, New York.
- Chang, C. S. & Read, J. (2007). *Support for foreign language listeners: Its effectiveness and limitations*. *RELC*. 38(3), 375 - 395.
- Clark, J. M. & Paivio, A. (1991). *Dual coding theory and education*. *Educational Psychology Review*, 3(3), 149 -170.
- Cook, G. (2000). *Language play, language learning*. Oxford: Oxford University Press.
- Colon-Vila, L. (1997). *Storytelling in an ESL classroom*. *Teaching Pre K-8*, 27(5), 58 - 59.
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Translators' Journal*, 49, 67 - 77.
- Dobson, M. J. (1992). *Effective techniques for English conversation group*. Washington: United States Information Agency.
- Goh, C. (1999). How much do learners know about the factors that influence their listening comprehension ?, Hong Kong. *Journal of Applied Linguistics*, 4(1), 17 - 42.
- Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British. Journal of Educational Technology*, 42(1), 181-192.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351 - 371.

- Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction.
Journal of Experimental Child Psychology, 108(1), 1 - 24.
- Linse, Caroline. T. (2005). *Practical English language teaching: Young Learners*. McGraw Hill, New York.
- Lutz, S., & Huitt, W. (2003). *Information processing and memory: Theory and applications*. Educational Psychology Interactive. Valdosta, GA: Valdosta State University.
- Mackay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- Madsen, H. J., Donald, B., & Ann, H. (1985). *TESOL techniques and procedures*. Cambridge, MA: Newbury House Publishers.
- Mansourzadeh, N. (2014). A comparative study of teaching vocabulary through pictures and audio-visual aids to young Iranian EFL learners. *Journal of Elementary Education*, 24 (1), 47- 59.
- Mayer, R. E. (2005). *Cognitive theory of multimedia learning*. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E. (2009). *Multimedia learning*. New York: Cambridge University Press.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
- Paivio, A. (1986). *Mental representations*. New York: Oxford University Press.
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, 12(1), 61- 86.
- Rasheed, W. & Mohammed, H. (2007). *The Impact of using pictures in vocabulary achievement for fifth primary pupil*. College of Basic Education.

- Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology*, 38(2), 109 -117.
- Shang, H. F. (2008). Listening strategy use and linguistic patterns in listening comprehension by EFL learners. *The Intl. Journal of Listening*. 22(1), 29 - 45.
- Slattery, M. & Willis, J. (2001). *English for primary teachers*. Oxford: Oxford University Press.
- Sticht, T. & James, J. (1984). *Listening and reading*. P.D. Perirson (Ed) Handbook of Reading Research. New York: Longman. 293 - 3 17.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10 (3), 251 - 296.
- Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, 38 (3), 422 - 427.
- University of Cambridge ESOL Examinations Research and Validation Group. (2009). *Example of speaking performance at CEFR levels A 2 to C 2*. Cambridge: Cambridge University Press.
- Wright, A. (1990). *Pictures for language learning*. Cambridge: Cambridge University Press.
- Yang, Y. T. C. & Chang, C. H. (2013). Empowering students through digital game authorship: *Enhancing concentration, critical thinking, and academic achievement*. *Computer & Education*, 68, 334 - 344.
- Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO journal*, 33 - 58.