Research Article:

The Impact of Image-Driven Stimuli on Malaysian Young Learners' Receptive and Productive Vocabulary

Debbita Tan Ai Lin, Lee Bee Choo, Malini Ganapathy* and Shaidatul Akma Adi Kasuma

School of Languages, Literacies and Translation, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

*Corresponding author: malinik@usm.my

ABSTRACT

In Malaysia's present scenario, competency in English is not a surety despite it being part of the curriculum from the very start of primary education. In view of the importance of lexical knowledge in language development and the hypothesised superiority of pictorial stimuli, the present study tested the effectiveness of an image-driven intervention for receptive and productive vocabulary knowledge development among Malaysian Year 1 students. A quasi-experimental design was utilised, involving Experimental and Control groups (totalling 159 participants). Three tiers of testing were administered – pre-testing, post-testing to measure for recall, and delayed post-testing for retention. The Experimental group experienced five image-driven sessions designed for lexical development (one session per week, 30 minutes per session). The Control group was not accorded the same treatment. Approximately 90% of the participants belonged to the B40 (Bottom 40%) group – the group with the lowest income level in Malaysia. Despite the lack of statistical significance, our findings on the whole show development with respect to the participants' lexical knowledge after experiencing the image-driven intervention, with raw scores clearly indicating receptive and productive vocabulary improvement at both the recall and retention stages. The study's findings extend the current knowledge base, especially within the Malaysian domain which suffers from an unfortunate lack of similar studies focusing on lower primary learners. Also, the results appear promising (with only five treatment sessions) and pave the way for more research concerning the effectiveness of the intervention among young ESL/EFL learners.

Keywords: Images, lexical development, ESL/EFL, young learners, Malaysia

INTRODUCTION

Despite globalisation, an acute awareness of the import of English language mastery as well as more emphasis on English language education in Malaysian schools and tertiary institutions, English proficiency levels among Malaysian youths in the 21st century remain questionable (Kashinathan & Abdul Aziz, 2021; Tan et al., 2020). Today, Malaysian graduates are still commonly labelled as "unqualified" and "incompetent" in terms of proficiency in English, which has led to an unfortunate suboptimal global image.

To further complicate matters, there is immense pressure for the Malaysian education system, especially where English is concerned, to measure up to global, borderless standards and demands – to not only "improve the English language competency of our young but also to align local English standards with that of an internationally recognised one, the CEFR (Common European Framework of Reference for Languages)" (Tan et al., 2020, p. 178).

The Malaysian government has taken serious measures in efforts to overcome the issue. The reforms include the Malaysia English Language Roadmap (2015–2025), Malaysia Education Blueprint (2013–2025), and the Primary School Standards-Based Curriculum for English Language Education (SBELC) introduced in 2011. These were launched to produce students with excellent English proficiency levels. Campbell (2018) stressed that the government must invest in a reform or approach that focuses on improving English proficiency. As such, there exists a need for effective English language learning methods and more importantly, for effective methods that focus on lexical development as vocabulary knowledge is fundamental to the learning and mastery of any language. As put forth by Khamis and Ho Abdullah (2018, p. 26), the "tenet of a language description is in its words". In addition, previous findings have consistently documented that vocabulary knowledge holds a significant niche in language development. Second/foreign language learning scholars such as Neuman and Kaefer (2018), Tan (2016a), McKeown and Beck (2014), Alemi and Tayebi (2011), Biemiller (2010), Hunt and Beglar (2005), and Nation (1990) have, over time, consistently demonstrated the significance of vocabulary knowledge with respect to achieving language mastery.

Research Problem and Objectives

As evident in the present scenario in Malaysia, competency in English is not a surety despite it being included in the curriculum from the very start of primary education – Year 1 (or Standard 1) in both national and national-type schools. It has been observed that Malaysian students' English proficiency leaves much to be desired (Azman, 2016). It is not a paucity of investment in the promotion of the language that has caused the lack of proficiency, nor is it a paucity in remediation efforts.

Although opinions may differ, we propose that English language education reform, particularly at the lower primary school level, should target the "building blocks" of the language – vocabulary. In other words, we put forward that any initiative aimed at improving or elevating English proficiency skills should not only ideally commence at the beginning of mandated schooling (i.e., lower primary school level), but also make lexical or vocabulary knowledge development its central tenet because it is largely futile to attempt the enhancement of the four skills (reading, writing, listening, and speaking) without first establishing good receptive and productive vocabulary knowledge.

Unfortunately, the teaching of English vocabulary is not specified in the Ministry of Education's *Dokumen Standard Kurikulum dan Pentaksiran* (Curriculum and Assessment Standards Document). What is available is a word list for teachers to refer to, but there is no model or method that they can refer to for the teaching of English vocabulary per se. Teachers frequently resort to drilling or rote-learning (memorisation) due to the phenomenon of teaching for the purpose of testing alone; this method is known to be useful for knowledge recall, but it may not be necessarily effective for knowledge retention (long-term memory storage of knowledge learnt).

As mentioned earlier, the dire state of Malaysian students' English proficiency – and given the reliance of language proficiency on lexical knowledge – it is only practical that an effective approach to vocabulary development is identified in the best interest of our students and of our nation. As such, we proposed the study of using images for English vocabulary knowledge development among Year 1 students. A comprehensive search of the existing canon of literature revealed a sore lack of studies, within the Malaysian context, that concentrate on image-driven interventions for lexical development among young or emergent ESL learners.

In essence, escalating interest in ESL/EFL vocabulary knowledge development has paved the way to a relatively wide corpus of studies in this area. However, many questions remain unanswered especially with respect to research output on vocabulary instruction among young learners providing us with clear directions or allowing us to reach a consensus on the most effective methods or approaches for this cohort. More specifically, the number of available research related to the use of visual aids for language learning among young learners is still scarce, particularly within settings outside of the Western sphere.

The present study is also extensive, assessing both the receptive and productive dimensions of vocabulary knowledge. Additionally, it measures vocabulary knowledge *recall* (short-term memory storage of vocabulary knowledge gained) and vocabulary knowledge *retention* (long-term memory storage of vocabulary knowledge gained). Past studies in the domain of vocabulary instruction (irrespective of intervention type) are generally limited to the receptive dimension and to the recall stage.

The following research objectives are addressed in the present study:

1. To determine if Malaysian Year 1 students' receptive and productive vocabulary knowledge recall are enhanced with the image-driven intervention.

2. To determine if Malaysian Year 1 students' receptive and productive vocabulary knowledge retention are enhanced with the image-driven intervention.

RELATED LITERATURE

Vocabulary in Language Learning

The importance of vocabulary in language learning is largely uncontested, and is regarded as fundamental to overall language proficiency and development (Masoumi & Sadeghi, 2020; Nation, 2013). However, although it is clear that lexical knowledge is essential to mastering a language, the best path to achieving good vocabulary is still arguable because numerous factors are often at play. For instance, cognitive ability, psychological make-up, and intervening sociocultural factors such as familial support and learning opportunities. Also, according to Schmitt (2008, pp. 329–330), "it is perhaps not surprising that teachers and learners have often been unsure of the best way to pursue it, especially as textbooks and syllabuses have typically been negligent in providing clear descriptions and guidelines".

Past studies have demonstrated that specific language skills are heightened and reciprocally related to vocabulary size. Meara (1996, p. 37) observed: "All other things being equal, learners with big vocabularies are more proficient in a wide range of language skills than learners with smaller vocabularies, and there is evidence to support the view that vocabulary skills make a significant contribution to almost all aspects of L2 proficiency."

Receptive and Productive Vocabulary

Essentially, to know a word is to recognise the word's form, and entails knowledge of the word's meaning(s) as well as the ability to use it appropriately. Nation (1990) initially proposed four constructs – form, meaning, function and position, a more concise version of Richard's (cited in Shen, 2008) elaborated version comprising seven constructs. In a 2001 publication, *Learning Vocabulary in Another Language*, Nation espoused a componential stance and presented word knowledge as composed of the three core facets of form, meaning and use (including associations with other words).

Receptive vocabulary knowledge is also known as vocabulary breadth and according to Nation (1990), a prominent feature of this dimension is the ability to recognise word form and retrieve meaning in reading and listening. Meanwhile, productive vocabulary knowledge (vocabulary depth) is the ability to use words accurately in the target language to express a particular meaning via writing or speaking (Nation, 1990). Included in the

latter dimension is also word-association ability (Meara, 2009; Nation, 2001); Beck and McKeown (1991) considered knowledge of a word to exist on a continuum, from not knowing to rich knowledge of a word's meaning, including its relationship with other words.

Meara and Wolter (2004) and Meara and Fitzpatrick (2000) emphasised that productive vocabulary knowledge reflects an individual's grasp of word meanings or in short, his or her receptive vocabulary knowledge. They were also of the position that vocabulary depth is deemed achieved as long as the learner is able to successfully link one lexical item with others, demonstrating a rich network of knowledge.

In the present study, both vocabulary forms (receptive/breadth and productive/depth) are covered and are operationalised as follows:

3. Receptive vocabulary (breadth): Knowledge of word meanings.

4. Productive vocabulary (depth): Word-association ability.

It can be stated that productive vocabulary knowledge is a critical but data-deficient area in the field of second/foreign language learning, particularly in terms of data derived from research on Malaysian learners, and this study seeks to extend our knowledge base in the area.

Picture Superiority Effect

The greater conceptual processing that pictures or images have over words is often justified by the PSE (Picture Superiority Effect), augmented by encoding theories and the notion of Transfer-Appropriate Processing (TAP). Pictures can be construed as being more effective for knowledge uptake because they are encoded differently from words alone. The picture superiority hypothesis denotes that pictures are more advantageous for learning as they trigger both verbal and non-verbal codes (dual-coding) whereas words are single-coded. This is potentially beneficial for young learners (Lee et al., 2020) and falls in line with the recommendation that interventions involve deep mental processing to encourage better retention (Tan, 2016b; McQuirter-Scott, 2010; Nemati, 2010).

There are basically two forms of processing, deep and shallow, with the former also known as semantic processing and involving more meaningful analysis of information; Craik and Lockhart (1972) also observed that stimuli like images are compatible with existing cognitive structures and are thus likely to be processed to a deeper level more rapidly and be better retained in both short- and long-term memory.

The Picture Word Inductive Model (PWIM) (Calhoun, 1999), the intervention applied in the present study, operates on the structure of deep processing as its base, involving

elaborate encoding of images presented. Paivio's (1971) dual-coding theory complements the PWIM; it suggests that pictures activate the verbal code and the non-verbal code, and thus have an edge over mono-coded information (isolated words).

Simply put, the basic underlying premise of the theory concerns: (1) mental structures; and (2) mental processes. The verbal dimension consists of visual, auditory, articulatory, and modality-specific codes (e.g., word-like codes such as 'book'), while the non-verbal dimension includes modality-specific images (e.g., for shapes) and non-linguistic objects and events (Clark & Paivio, 1991; Paivio, 1971).

As for TAP, the PSE can be experienced via the interrelation between encoding and retrieval. Roediger and his colleagues (Roediger & McDermott, 1993; Weldon & Roediger, 1987) have developed TAP elaborations encompassing conceptual and perceptual processes. TAP is inclined towards emphasis on general forms of processing, conceptual or perceptual, and used to explain the relationship between stimulus (exposure), encoding (memory), and retrieval (performance). In essence, TAP is concerned with how information is initially encoded and how it is later retrieved. For instance, if items are primarily encoded conceptually during a semantic task, performance is expected to be greater in a test that relies on concepts related to the items for retrieval. It would be unusual if greater performance is instead achieved in a test that relies on perceptual features.

It is interesting to note, however, that there are competing views. Contrary to the position of picture superiority, Amrhein et al. (2002) found that words produced a faster response than pictures, and that pictures did not have the advantage of easier access to semantic memory. Strauber et al. (2020) observed that past studies have revealed that words embedded with images afforded no advantage over text alone. Ensor et al. (2019), in their work comprising four experiments to test the physical distinctiveness account of the PSE, found that although coloured pictures were recognised better than distinctive words, distinctive words were significantly better recognised than black and white pictures.

Concept Attainment

Learners are considered natural conceptualisers with the capacity to make discernments when presented with diverse items, incidents or experiences (Joyce et al., 2009). The PWIM is an approach intended to persuade a learner's innate potentiality and motivate the expansion of initial concepts formed. Concept attainment subsumes within its structure the notion of inductive thinking. Interventions or teaching-learning approaches that include concept attainment fundamentally induce learners to think, to form concepts, expand them, and apply them.

A visceral property of concept attainment is the testing of data. The process typically requires learners to focus on an attribute, form theories, and test them against existing knowledge until new concepts are formed. Concept attainment is a constructivist approach

drawn from the work of psychologist Jerome Bruner (Bruner et al., 1956), is student-centred (Sukardjo & Salam, 2020), and can be used at any level and subject area (Mayer, 2012).

The crux of our study's intervention lies in capitalising on learners' ability to think inductively, coupled with the use of images (of objects and actions), to elicit for words from their existing vocabularies in the process of learning new ones. At least three stages of the PWIM are devoted to concept attainment (Joyce et al., 2009): (1) identification of data; (2) data classification based on general elements; and (3) reclassification and hypothesis development. Figure 1 illustrates the premises that frame the present study.

Past Studies

In recent years, interest in the use of images for language learning has increased although empirical studies in this area remain scarce, particularly within the context of Asia. Among the available studies is one by Jiang (2015). The study involved 30 seventh graders in China and found the PWIM to be useful for improving narrative writing in English. Previously, Jiang (2014) studied the impact of the model on the English vocabulary development of fourth and seventh graders, as well as the possibility of cognitive overload when words are combined with pictures. Jiang found no significant improvement in terms of lexical development with some subjects reporting instances of cognitive overload.

Closer to home, Nurani and Rosyada (2017), within the context of ESL, investigated the effects of the PWIM on the communicative competence of 47 eleventh graders enrolled in a public school in Jakarta, Indonesia. Their research comprised experimental and control groups as well as post-testing and the use of questionnaires; results indicated significant effects in terms of enhanced comprehension, pronunciation, fluency, grammatical and vocabulary knowledge, and increased learning motivation.

In Malaysia, the availability of PWIM research is poor especially with respect to studies that demonstrate fidelity to rigour. Lee et al. (2019) introduced a modified version of the original PWIM and tested it on Malaysian Year 1 students in a rural setting; students were permitted to use their mother tongue when they did not know the words in English and the teacher was required to translate them into English immediately. A total of 60 subjects participated in the research, comprising two intact classes which formed the experimental and control groups. Results gleaned from independent samples t-tests revealed the former group's superior vocabulary scores.

Tan et al. (2020) investigated the impact of images among Year 1 Malaysian ESL learners in a quasi-experimental study involving 53 participants from an urban public school. Two tests were administered and predictions that the participants who experienced vocabulary learning through images would outperform their peers were largely, but not wholly,

supported by descriptive and inferential statistics. However, one of the tests – the Yes-No Test by Meara (1992) – suffered a ceiling effect, causing the absence of statistical significance and prompting future studies to consider the inclusion of more complex lexical items for testing as well as the employment of larger participant cohorts.

METHODOLOGY

Design

A quantitative approach was adopted and a quasi-experimental design formed the core of the study. Convenience sampling was employed and the intact classes used were assigned to the researchers by school authorities, as is common in educational research (Ary et al., 2009). The coin-toss method (Neuman, 2013) was used to determine the status of each intact class (Experimental status or Control status).

Quasi-experimental designs are considered worthwhile because they permit researchers to reach reasonable conclusions (Ary et al., 2009). It is also a design that allows researchers to collect evidence of treatment effects, especially in the context of educational interventions (Diem, 2002). Nested within this is the pretest-posttest-delayed posttest control-group approach (Johnson & Christensen, 2012; see Figure 1). Denzin (1978) observed the importance of within-method triangulation measures such as conducting post-testing and assigning control groups in order to increase the reliability of research findings.
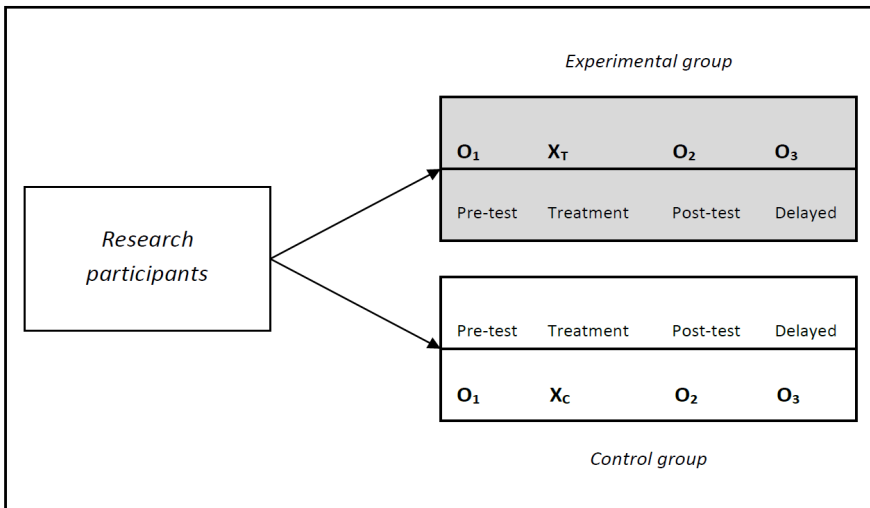

Figure 1. Pretest-posttest-delayed posttest control-group design (Adapted from Johnson & Christensen, 2012)

Instruments and Intervention

1. Images: Selected from Year 1 textbook, blown up to A1-sized full-colour posters .

2. Word list: Year 1 list from MOE's *Dokumen Standard Kurikulum dan Pentaksiran* (2016).

3. Tests: Receptive Vocabulary Yes-No Test (Meara, 1992); Productive Vocabulary Lex30 Test (Meara & Fitzpatrick, 2000).

According to vocabulary testing expert John Read (2007), the Yes-No Test by Paul Meara (1992) is the simplest testing format and is practical for use in terms of testing for receptive vocabulary knowledge. The test has been validated by various studies with returned values indicating a positive correlation with other similar tests (Mochida & Harrington, 2006).

The Lex30 Test by Meara and Fitzpatrick (2000) has been repeatedly validated as a reliable instrument to measure ESL learners' productive vocabulary knowledge across all ages, including young learners (Gonzalez & Piriz, 2016). Fitzpatrick and Clenton (2010) highlighted that the test produces reliable scores, comparable to those of similar tests. All the instruments were pilot-tested prior to the study.

As for the PWIM, each cycle of the intervention comprised the steps (based on Calhoun, 1999) as shown in Figure 2.

(1) Teacher puts up selected picture on the board and gets students to sit in groups.
(2) Teacher goes through the steps of the intervention with students:

- Points at an item in the picture and asks students to identify it;
- Labels by drawing a line from the identified object or area;
- Students spell and read the word aloud;
- When the teacher has gone through the items in the picture, students review the picture aloud (read and spell);
- Teacher helps students in identifying concepts that are related to each item;
- Teacher leads students in creating a title for the picture;
- Teacher guides students in generating and reviewing simple sentences using the picture;
- Only a few words are induced (enough to generate simple sentences) as examples for the students, and the students then continue in groups;
- Teacher walks around to guide, supervise, and extend help when necessary. Teacher ensures that every student gets a chance to participate and that no one is isolated.

Figure 2. Steps of the intervention

Participants, Phases and Procedures

A total of 159 learners were registered to participate in the study and 25 English target words were taught over a five-week duration. Three enlarged pictures, selected from the mainstream textbook, were employed in the intervention. The participants were tested for recall (receptive and productive vocabulary) immediately after the final intervention session and three weeks after that, tested for retention (receptive and productive vocabulary). The primary phases and procedures are as listed in Table 1.

Table 1. Phases and procedures

Phase	Activity	Timeline/Duration	Participants	
			Exp	Ctrl
1	Pre-testing	Week 1	80	69
2	Intervention	5 weeks (1 session per week at 30 minutes per session)		
3	Post-testing	Week 7	83	73
4	Delayed post-testing	Week 10	79	68

It is to be noted that participation varied in terms of test-taking; not all the participants turned up/sat for the pre-, post- and delayed post-tests, a rather common occurrence in research involving children from poor households, hence the difference in numbers as reflected in Table 1. These absences were accounted for during statistical analysis.

The research participants were Malaysian Year 1 students (mean age: 7.6 years). 144 were Malays while Indians numbered 15. Approximately 90% of the participants belonged to the B40 (Bottom 40%) group – the group with the lowest income level in the country.

Data Analysis

Both descriptive and inferential statistics were employed to lend greater depth to the results. Data garnered from the administration of the tests was analysed to derive central tendency values (mean scores) as well as standard deviation values (variance of distribution in relation to the mean). One-way analysis of variance (ANOVA) procedures were also run to check for statistical significance. The key contrast between ANOVA and a *t*-test is that the latter can only compare two means at a time while ANOVA has no such restrictions and is known to reduce the incidence of Type 1 errors (false positives).

The one-way ANOVA is essentially a comparison of the means of two or more groups in order to gauge whether variables are dependent on or independent of each other, which leads to either acceptance or rejection of a null hypothesis based on a pre-determined

significance level. A *p*-value that is smaller than a pre-determined threshold value (normally .05) indicates that the observed effect is unlikely to have happened by chance (Privitera, 2012).

It is worthy of note that the *p*-value should not be regarded as absolute proof of a treatment's efficacy. Opinions vary regarding reliance on the use of *p*-values in hypothesis testing; applications of significance testing can be useful to a certain extent in providing researchers with an idea of treatment effects, but to treat the values returned via significance testing as sure evidence that an intervention works or as sure evidence of its generalisability may run the risk of overstatement.

For instance, Kline (2015) opined that in applying hypothesis testing researchers are simply testing if mediation exists, rather than determining if a mediation effect is sufficiently large and precise to be appreciable in a particular context. Similarly, Trafimow and Marks (2016, p. 1) stated that the *p*-value does not "provide a strong basis for drawing conclusions" while Fricker and his colleagues (2019), in their work concerning statistical analyses used in human studies, acknowledged that although there are limitations to statistical inference, they ascribed to the philosophy of British scientist Francis Galton: "it remains important to emphasize Galton's key caveats that statistical methods *are not brutalized* and that they are *warily interpreted*" (Fricker et al., 2019, p. 382).

Integrity

The researchers, in cooperation with school authorities, closely monitored the research to ensure the smooth-running of study protocols. Consent was also obtained and anonymity is preserved to fully protect the participants' privacy.

RESULTS

Receptive Vocabulary Knowledge (Breadth): Recall and Retention

Both descriptive and inferential statistics were employed to lend greater depth to the findings. The *p*-value was set at $p < 0.05$.

As shown in Table 2, the mean scores of the pre-test for both the Control and Experimental groups do not differ significantly, denoting baseline similarity at 14.25 (*SD* = 5.842) and 13.27 (*SD* = 5.262) respectively. Post-test data indicate that the Experimental group experienced higher vocabulary breadth recall than the Control group; the Experimental group's mean score at 17.23 (*SD* = 6.549) is superior to the Control group's achievement of 16.44 (*SD* = 6.303). As for the delayed post-test, which measures vocabulary breadth retention, the mean scores achieved were 19.05 (*SD* = 6.641) (Experimental group) and 18.03 (*SD* = 6.202) (Control group), indicating no knowledge decay and interestingly, possible delayed learning effects.

Table 2. Pre-test, post-test and delayed post-test

Test	Group	N	Mean	SD
Pre-test	Control	69	14.25	5.842
	Experimental	79	13.27	5.262
Post-test	Control	73	16.44	6.303
	Experimental	83	17.23	6.549
Delayed post-test	Control	70	18.03	6.202
	Experimental	76	19.05	6.641

As shown in Table 3, the Sig. (*p*) value obtained was 0.231, 0.680 and 0.422 at the pre-test, post-test and delayed post-test levels respectively. As the resulting *p*-values were more than 0.05 ($p > 0.05$), the null hypothesis of no difference between the groups' variances was accepted. This indicates that the assumption of homogeneity of variances was met.

Table 3. Levene's Test for homogeneity of variances

Homogeneity of variances				
Test	Levene	df	df2	Sig.
Pre-test	1.447	1	146	0.231
Post-test	0.171	1	154	0.680
Delayed post-test	0.647	1	144	0.422

At the pre-test level, Table 4 shows that the Sig. (*p*) value obtained was more than 0.05 ($p > 0.05$) at $p = 0.284$, indicating no statistically significant difference between group means. Table 4 also indicates that the *p*-value obtained was more than 0.05 ($p > 0.05$) at $p = 0.445$ at the post-test level, indicating no statistically significant difference between group means at the recall level for vocabulary breadth. Similarly, no statistically significant difference between group means was found at the delayed post-test level (retention) with a value of $p = 0.338$.

Table 4. Statistical significance

Test		Sum of squares	df	Mean square	F	Sig.
Pre-test	Between group	35.413	1	35.413	1.154	0.284
	Within groups	4480.229	146	30.687		
Post-test	Between groups	24.274	1	24.274	0.856	0.445
	Within groups	6376.623	154	41.407		
Delayed post-test	Between groups	38.213	1	38.213	0.923	0.338
	Within groups	5961.732	144	41.401		

The *p*-value obtained at the pre-test level, which specifies no statistically significant difference between group means, indicates that the testing instrument is reliable in terms of "reliability as stability over similar samples" (Cohen et al., 2018), thus presenting it as a feasible option for use among similar cohorts in future studies.

Productive Vocabulary Knowledge (Depth): Recall and Retention

Descriptive and inferential statistics were also employed to lend greater depth to the findings on productive vocabulary knowledge development. As with the receptive dimension, the *p*-value was set at $p < 0.05$.

Table 5. Pre-test, post-test and delayed post-test

Test	Group	N	Mean	SD
Pre-test	Control	69	6.91	5.249
	Experimental	80	7.29	5.766
Post-test	Control	72	7.56	6.373
	Experimental	81	9.11	7.301
Delayed post-test	Control	66	10.85	7.267
	Experimental	79	11.10	7.512

As shown in Table 5, the mean scores of the pretest for both the Control and Experimental groups do not differ significantly, denoting baseline similarity at 6.91 (*SD* = 5.249) and 7.29 (*SD* = 5.766), respectively. Posttest data indicate that the Experimental group experienced higher vocabulary depth recall than the Control group; the Experimental group's mean score at 9.11 (*SD* = 7.301) is superior to the Control group's achievement of 7.56 (*SD* = 6.373). As for the delayed posttest, which measures vocabulary depth retention, the mean scores achieved were 11.10 (*SD* = 7.512) (Experimental group) and 10.85 (*SD* = 7.267) (Control group), indicating no knowledge decay and possible delayed learning effects.

Table 6. Levene's Test for homogeneity of variances

Homogeneity of variances				
Test	Levene	df	df2	Sig.
Pre-test	0.439	1	147	0.509
Post-test	1.625	1	151	0.204
Delayed post-test	0.326	1	143	0.569

As shown in Table 6, the Sig. (*p*) value obtained was 0.509, 0.204 and 0.569 at the pretest, posttest and delayed posttest levels respectively. As the resulting *p*-values were more than 0.05 (*p* > 0.05), the null hypothesis of no difference between the groups' variances was accepted. This indicates that the assumption of homogeneity of variances was met.

Table 7. Statistical significance

Test		Sum of squares	df	Mean square	F	Sig.
Pre-test	Between group	5.915	1	5.195	0.170	0.681
	Within groups	4499.866	147	30.611		
Post-test	Between groups	92.235	1	92.235	1.94	0.165
	Within groups	7147.778	151	47.336		
Delayed post-test	Between groups	2.298	1	2.298	0.042	0.838
	Within groups	7833.675	143	54.781		

At the pretest level, Table 7 shows that the Sig. (p) value obtained was more than 0.05 ($p > 0.05$) at $p = 0.681$, indicating no statistically significant difference between group means. Table 7 also indicates that the p-value obtained was more than 0.05 ($p > 0.05$) at $p = 0.165$ at the posttest level, indicating no statistically significant difference between group means at the recall level for vocabulary depth. Similarly, no statistically significant difference between group means was found at the delayed posttest level (retention) with a value of $p = 0.838$.

Additionally, the p-value obtained at the pretest level, which specifies no statistically significant difference between group means, indicates that the testing instrument is reliable in terms of 'reliability as stability over similar samples' (Cohen et al., 2018), thus presenting it as a feasible option for use among similar cohorts in future research.

DISCUSSION

The researchers aimed to ascertain if Malaysian Year 1 students' receptive (breadth) and productive (depth) vocabulary knowledge recall and retention are enhanced with image-driven stimuli. The overall findings show development with respect to the participants' receptive and productive vocabulary knowledge after experiencing image-driven intervention. More specifically, development is observed at both the recall and retention levels, in line with previous studies suggesting the effectiveness of the PWIM in improving vocabulary knowledge.

We elect, however, to abide by caution and note that although the raw scores clearly indicate receptive and productive vocabulary improvement at the recall and retention stages, statistical analysis did not return values indicating significance (i.e., $p < 0.05$). Incorporating more treatment sessions in future research will likely afford us a better picture in terms of statistical significance.

Both the Control and Experimental groups began with comparable mean scores. However, posttest data shows that the Experimental group achieved higher vocabulary breadth recall than the Control group. Meanwhile, data collected from delayed posttesting shows that the mean scores for vocabulary breadth retention achieved by the same cohort of participants were markedly higher for the Experimental group in comparison to the Control group. Interestingly, the retention mean scores were also higher than the recall mean scores for both groups, demonstrating not only the absence of knowledge decay but also suggesting the possibility of delayed learning effects. The latter is particularly worthy of attention as the intervention appears to have long-term potential in terms of positive learning effects.

In terms of productive vocabulary, the Experimental group also outperformed the Control group at the recall stage. Both groups commenced with similar baseline scores. This augments the proposal that with the PWIM, students are able to experience productive vocabulary development. Meanwhile, delayed posttest data reveals only minor disparities in the mean scores obtained by the Experimental and Control groups. On the whole, this indicates the absence of knowledge decay as well as the possibility of delayed learning effects – a promising avenue for future research to look into the long-term potential of the PWIM, within the context of continued productive vocabulary knowledge development.

The growth in mean scores for knowledge recall and retention, for both receptive and productive vocabulary, offers support for the basis that pictures are optimal in knowledge development due to comprehensive encoding in the brain. The findings of the present study suggest pictures' utility and superiority in learning, as well as the significance of activating dual-coding in the brain instead of relying solely on single-coding; Paivio's dual-coding theory highlights that pictures activate both the verbal and non-verbal codes – the edge that this process offers is consistent with the results of the current study.

Additionally, the results also demonstrate that interventions that compel learners to engage in deep mental processing are more likely to promote knowledge retention. Evidently, pictorial stimuli are compatible with our cognitive structures and are hence likely to be processed deeper more rapidly, and be better stored in our short- and long-term memory. As for the notion of concept attainment, a constructivist approach characterised by persuasion of a learner's innate potentiality and the expansion of initial concepts formed, the findings of the study imply its efficacy as the crux of the study's intervention lies in capitalising on the participants' ability to think inductively, coupled with the use of images (of objects and actions), to elicit for words from their existing vocabularies in the process of learning new ones.

Despite earlier research, such as by Amrhein et al. (2002), claiming that learning via images is less effective due to students requiring more time to respond, the findings obtained from this study concur with more recent ones. For example, Ensor et al. (2019) advanced that coloured pictures are significantly easier for students to comprehend than words. Conversely, Strauber et al. (2020) put forward that past studies have shown that incorporating images delivers no advantage over using text alone. However, in their own enquiry, Strauber and

his colleagues found positive effects as stated in the following: "Children in the picture-embedded word condition performed significantly higher than those in the word-alone condition on an immediate post-training test and later retention tests." (Strauber et al., 2020, p. 1).

Closer to home, our findings are also to differing degrees supportive of those produced via previous research. For instance, studies within the Malaysian context by Tan et al. (2020) and Lee et al. (2019), and another within the Indonesian context by Nurani and Rosyada (2017). In the first study which involved a total of 53 participants, Tan et al. (2020) assessed the impact of images on Malaysian Year 1 students' English vocabulary development. Results revealed superior receptive and productive vocabulary scores at the recall and retention stages, in favour of the Experimental group (n = 27). Statistical significance was also achieved but only for the productive dimension. When compared to the present study, both data sets exhibit a general pattern of similarity that supports the employment of image-driven stimuli in English language classrooms. It is to be noted also that only the present study garnered findings that suggest the possibility of delayed learning effects. This may be due to the use of a larger sample; larger samples often provide more varied and richer data, unless the subjects are extremely similar.

Lee et al. (2019) also did not report the possibility of delayed learning effects. In their study involving 60 Malaysian Year 1 participants in total, the Experimental group (n = 30) significantly outperformed the controls at the recall and retention levels. The study covered only the receptive dimension and concluded that vocabulary acquisition was consistently positive when the learners were exposed to picture-text stimuli via the PWIM approach. This largely runs in tandem with the findings of the present study in that image-driven treatments are helpful for receptive vocabulary knowledge development. Nurani and Rosyada (2017) studied the effects of the PWIM on the English communicative competence of 47 eleventh graders enrolled in a public school in Jakarta, Indonesia. They observed significant effects in terms of enhanced comprehension, pronunciation, fluency, grammatical and vocabulary knowledge, and increased learning motivation. The study can be considered comprehensive because it covered a range of language-related aspects, including the affective facet. The same cannot be said for the present study as it is only focused on vocabulary knowledge. However, it is imperative to note the difficulty in determining the Indonesian study's level of research rigour. For instance, information regarding testing appears lacking although the study covers multiple aspects relating to learner performance.

Although the growth in the current study's mean scores does not translate to statistical significance, the results are still indicative of development, no knowledge decay, and possible delayed learning effects. These are positive discoveries that permit us to encourage the use of image-driven interventions for the purpose of improving vocabulary breadth and depth. Furthermore, the PWIM is an adaptive teaching-learning approach that enables learners to acquire knowledge via different parameters, thus allowing for greater acquisition

probabilities. It has also been postulated that adaptive learning systems can minimise knowledge decay, something which is invaluable in language education.

CONCLUSION

This study's findings extend the existing knowledge base on vocabulary development and as the results garnered are promising, pave the way for more studies concerning the effectiveness of the PWIM for vocabulary growth among language learners. Even so, the study is not without limitations and these afford directions for future lines of enquiry. First, data was gathered from a convenience sample of 159 participants, of which approximately 90% belonged to the B40 group. Furthermore, the participants were predominantly Malay. Only about 9.5% were Indians and the Chinese were not represented at all even though together, these three ethnicities make up the majority of Malaysia's population. The country's B40 group also includes all three ethnicities. For a more rigorous and representative examination, future research should utilise samples that comprise only participants with very similar economic backgrounds and ensure that at least the major ethnic groups are adequately represented. Another direction worth taking is to conduct comparative studies involving ethnically diverse samples.

Secondly, the study's experimental were taught 25 target words over a relatively short duration of five treatment weeks. It is recommended that future research attempt a longer duration and more target words. Testing was also a challenge as participation varied in terms of test-taking. Not all the participants turned up/sat for the tests. These absences were accounted for during statistical analysis, but learning losses due to absence from exposure to treatment were very likely not recouped. Children living in poverty usually encounter different barriers to accessing educational opportunities and non-attendance remains a persistent problem. Many disadvantaged children begin primary school with below par readiness. This includes poor socioeconomic, physical, and psychological well-being. Providing essential items that serve as incentives, in tandem with rendering psychological support when possible, may result in better research outcomes. However, incentivising participants can be costly with large samples and if this avenue is pursued, securing sufficient sponsorship should be a priority.

Lastly, productive vocabulary knowledge is operationalised as word-association ability in the present study and as such, the Lex30 Test by Meara and Fitzpatrick (2000) was considered suitable for use. While this approach is valid and has produced encouraging findings, researchers should not limit themselves to these parameters. Productive vocabulary knowledge includes as well the dimension of "retrieval and use" (i.e., the ability to use words appropriately in speech and/or writing). One example of a relevant, comprehensive test is the Vocabulary Knowledge Scale (Wesche & Paribakht, 1996). In this regard, researchers are urged to capitalise on the PWIM's embodiment of concept attainment and extend the model's reach to include other language skills (such as public speaking and narrative writing) as well as other languages to further enrich the literature on second/ foreign language teaching and learning.

In conclusion, the key findings of the present study offer useful insights on the effectiveness of the PWIM for vocabulary recall and retention. The highly structural nature of the intervention makes it a particularly suitable option for use with young learners. Not only does the intervention encourage elaborate processing and systematically prompt inductive thinking in learners, but it exploits their prior knowledge in the process of contructing new corpus amid an interactive learning environment, an environment in which every learner has a chance to participate. Although admittedly there is a lack of statistical significance, the raw test scores derived from the current study still demonstrate gains and more interestingly, possible delayed learning effects. The lack of statistical significance could be due to insufficient frequency of exposure to the intervention. On the whole, there is no compelling reason as to why the intervention should not be deemed feasible for use as part of classroom practice and for research purposes.

ACKNOWLEDGEMENT

REFERENCES

Alemi, M., & Tayebi, A. (2011). The influence of incidental and intentional vocabulary acquisition and vocabulary strategy use on learning L2 vocabularies. *Journal of Language Teaching and Research, 2*(1), 81–98. https://doi.org/10.4304/jltr.2.1.81-98

Amrhein, P. C., McDaniel, M. A., & Waddill, P. (2002). Revisiting the picture-superiority effect in symbolic comparisons: Do pictures provide privileged access? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 843–857. https://doi.org/10.1037/0278-7393.28.5.843

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. K. (2009). *Introduction to research in education* (8th ed.). California: Cengage.

Azman, H. (2016). Implementation and challenges of English language education reform in Malaysian primary schools. *3L: The Southeast Asian Journal of English Language Studies, 22*(3), 65–78. https://doi.org/10.17576/3L-2016-2203-05

Beck, I. L., & McKeown, M. G. (1991). Conditions of vocabulary acquisition. In P. D. Pearson (Ed.), *The handbook of reading research* (pp. 789–814). New York: Longman.

Biemiller, A. (2010). Vocabulary development and implications for reading problems. In A. McGill-Franzen, & R. Allington (Eds.), *Handbook of reading disabilities research* (pp. 208–218). New York: Routledge.

Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening, 14*, 14–31. https://doi.org/10.1080/10904018.2000.10499033

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. London: Chapman & Hall. https://doi.org/10.2307/1292061

Calhoun, E. F. (1999). *Teaching beginning reading and writing with the Picture Word Inductive Model*. Alexandria: Association for Supervision and Curriculum Development (ASCD).

Campbell, J. (2018). English language and an inclusive Malaysia. *3L: The Southeast Asian Journal of English Language Studies, 24*(3), 206–219. https://doi.org/10.17576/3L-2018-2403-15

Clark, J. M., & Paivio, A. D. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149–210. https://doi.org/10.1007/BF01320076

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). London: Routledge. https://doi.org/10.4324/9781315456539

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Defeyter, M. A., Russo, R., & McPartlin, P. L. (2009). The picture superiority effect in recognition memory: A developmental study using the response signal procedure. *Cognitive Development, 24*(3), 265–273. https://doi.org/10.1016/j.cogdev.2009.05.002

Denzin, N. K. (1978). *The research act* (2nd ed.). New York: McGraw-Hill.

Diem, K. G. (2002). Using research methods to evaluate your extension program. *Journal of Extension, 40*(6). https://archives.joe.org/joe/2002december/a1.php

Ensor, T. M., Surprenant, A. M., & Neath, I. (2019). Increasing word distinctiveness eliminates the picture superiority effect in recognition: Evidence for the physical-distinctiveness account. *Memory & Cognition, 47*, 182–193. https://doi.org/10.3758/s13421-018-0858-9

Eyckmans, J. (2004). Measuring receptive vocabulary size: Reliability and validity of the yes/no vocabulary test for French-speaking learners of Dutch. Volume 91 of LOT: LOT (Landelijke Onderzoekschool Taalkunde), Netherlands Graduate School of Linguistics.

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing, 27*(4), 537–554. https://doi.org/10.1177/0265532209354771

Fricker, R. D., Burke, K., Han, X., & Woodall, W. H. (2019). Assessing the statistical analyses used in basic and applied social psychology after their *p*-value ban. *The American Statistician, 73*(S1), 374–384. https://doi.org/10.1080/00031305.2018.1537892

Gonzalez, R. A., & Piriz, A. M. P. (2016). Measuring the productive vocabulary of secondary school CLIL students: Is Lex30 a valid test for low-level school learners? *Vigo International Journal of Applied Linguistics, 13*, 31–54.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17*(1), 23–59.

Jiang, X. (2014). Vocabulary learning through use of the Picture Word Inductive Model for young English learners in China: A mixed methods examination using cognitive load theory. Unpublished doctoral dissertation, University of Florida.

Jiang, X. (2015). Exploring seventh graders' perceptions of the Picture Word Inductive Model in their English narrative writing in China. In S. M. Nielsen (Ed.), *Proceedings of the 14th Annual South Florida Education Research Conference* (pp. 76–84), Miami, Florida.

Johnson, B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). California: Sage.

Joyce, B., Weil, M., & Calhoun, E.F. (2009). *Models of teaching* (8th ed.). Boston: Pearson Education.

Kashinathan, S., & Abdul Aziz, A. (2021). ESL learners' challenges in speaking English in Malaysian classroom. *International Journal of Academic Research in Progressive Education & Development, 10*(2), 983–991. https://doi.org/10.6007/IJARPED/v10-i2/10355

Khamis, N., & Ho Abdullah, I. (2018). What do different word lists reveal about the lexical features of a specialised language? *3L: The Southeast Asian Journal of English Language Studies, 24*(3), 26–42. https://doi.org/10.17576/3L-2018-2403-03

Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology, 37*(4), 202–213. https://doi.org/10.1080/01973533.2015.1049349

Lee, B. C., Pandian, A., & Rethinasamy, S. (2020). Recall and retention of vocabulary depth of young learners via PWIM. *Pertanika Journal of Social Sciences & Humanities, 28*(S2), 297–319.

Lee, B. C., Pandian, A., Rethinasamy, S., & Tan, D. A. L. (2019). Effects of PWIM in the ESL classroom: Vocabulary knowledge development among primary Malaysian learners. *3L: The Southeast Asian Journal of English Language Studies, 25*(4), 179–197. https://doi.org/10.17576/3L-2019-2504-11

Masoumi, G. A., & Sadeghi, K. (2020). Impact of test format on vocabulary test performance of EFL learners: The role of gender. *Language Testing in Asia, 10*(2), 1–13. https://doi.org/10.1186/s40468-020-00099-x

Masrai, A. (2019). Vocabulary and reading comprehension revisited: Evidence for high-, mid- and low-frequency vocabulary knowledge. *SAGE Open, 9*(2), 1–13. https://doi.org/10.1177/2158244019845182

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes, 31*, 44–57. https://doi.org/10.1016/j.jeap.2017.12.006

Mayer, J. R. (2012). Effects of using the concept attainment model with inductive reasoning with high school biology students. Unpublished Master's thesis, Montana State University.

McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly, 29*(4), 520–530. https://doi.org/10.1016/j.ecresq.2014.06.002

McQuirter-Scott, R. (2010). *What works? Research into practice.* Ontario: The Literacy and Numeracy Secretariat.

Meara, P. (1992). *EFL vocabulary tests.* Swansea: Centre for Applied Language Studies, University College Swansea.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), P*erformance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition.* Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.24

Meara, P., & Fitzpatrick, T. (2000). Lex 30: An improved method of assessing productive vocabulary in an L2. *System, 28*(1), 19–30. https://doi.org/10.1016/S0346-251X(99)00058-5

Meara, P., & Wolter, B. (2004). V_Links: Beyond vocabulary depth. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Angles on the English-speaking world 4* (pp. 85–96). Copenhagen: Museum Tusculanum Press.

Ministry of Education (MOE), Malaysia (2016). Dokumen standard kurikulum dan pentaksiran. Retrieved from https://www.moe.gov.my/en/moe-staff/education-services-officer/dskp

Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing, 23*(1), 73–98. https://doi.org/10.1191/0265532206lt321oa

Nation, P. (1990). *Teaching and learning vocabulary.* New York: Newbury House.

Nation, P. (1997). Bring today's vocabulary research into tomorrow's classrooms. In G. M. Jacobs (Ed.), *Language classrooms of tomorrow: Issues and responses* (pp. 170–182). Singapore: SEAMEO Regional Language Centre.

Nation, P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nemati, A. (2010). Enhancing long-term retention by memory vocabulary learning strategies. *The Journal of Asia TEFL, 7*(1), 171–195. https://doi.org/10.4314/marang.v20i1.56821

Neuman, W. L. (2013). *Social research methods: Qualitative and quantitative approaches* (7th ed.). Essex: Pearson Education.

Neuman, S. B., & Kaefer, T. (2018). Developing low-income children's vocabulary and content knowledge through a shared book reading program. *Contemporary Educational Psychology, 52*, 15–24. https://doi.org/10.1016/j.cedpsych.2017.12.001

Nurani, S., & Rosyada, A. (2017). Implemented PWIM in developing students' communicative competence: SMK Islam Wijaya Kusuma Jakarta. *Lingua Cultura, 11*(1), 25–30. https://doi.org/10.21512/lc.v11i1.1608

Paivio, A. D. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Privitera, G. J. (2012). *Statistics for the behavioral sciences.* California: Sage.

Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies, 7*(2), 105–125.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In H. Spinnler, & F. Boller (Eds.), *Handbook of neuropsychology* (pp. 63–131). Amsterdam: Elsevier.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Sharma, G. (2017). Pros and cons of different sampling techniques. *International Journal of Applied Research, 3*(7), 749–752.

Shen, Z. (2008). The roles of breadth and depth of vocabulary knowledge in EFL reading performance. *Asian Social Science, 4*(12), 135–137. https://doi.org/10.5539/ass.v4n12p135

Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal, 18*(1), 87–98. https://doi.org/10.1207/s15326934crj1801_10

Strauber, C. B., Sorcar, P., Howlett, C., & Goldman, S. (2020). Using a picture-embedded method to support acquisition of sight words. *Learning and Instruction, 65*, 1–9. https://doi.org/10.1016/j.learninstruc.2019.101248

Sukardjo, M., & Salam, M. (2020). Effect of concept attainment models and self-directed learning (SDL) on mathematics learning outcomes. *International Journal of Instruction, 13*(3), 275–292. https://doi.org/10.29333/iji.2020.13319a

Tan, D. A. L. (2016a). Investigating guided extensive reading and vocabulary knowledge performance among remedial ESL learners in a public university in Malaysia. Unpublished doctoral dissertation, Universiti Sains Malaysia, Malaysia.

Tan, D. A. L. (2016b). Trends in second/foreign language teaching and learning: The position assigned to the learning of lexis over the years. *Advances in Language and Literary Studies, 7*(6), 84–90. https://doi.org/10.7575/aiac.alls.v.7n.6p.84

Tan, D. A. L., Lee, B. C., Adi Kasuma, S. A., & Ganapathy, M. (2020). Picture superiority effect: Using images for vocabulary enhancement among year one Malaysian ESL learners. *Pertanika Journal of Social Sciences & Humanities, 28*(S2), 177–196.

Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal, 112*(3), 497–518. https://doi.org/10.1086/663301

Trafimow, D., & Marks, M. (2016). Editorial. *Basic and Applied Social Psychology, 38*(1), 1–2. https://doi.org/10.1080/01973533.2016.1141030

Uchihara, T., & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research, 24*(4), 540–556. https://doi.org/10.1177/1362168818799371

Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 15*(4), 269–280. https://doi.org/10.3758/BF03197030

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*(1), 13–40. https://doi.org/10.3138/cmlr.53.1.13

Yuksel, H. G., & Kavanoz, S. (2010). Does vocabulary knowledge distinguish among proficiency levels of Turkish university students? *International Journal of Learning, 17*(5), 513–521. https://doi.org/10.18848/1447-9494/CGP/v17i05/47071