Pertanika J. Soc. Sci. & Hum. 28 (S2): 297 - 319 (2020)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Recall and Retention of Vocabulary Depth of Young Learners via PWIM

Bee Choo Lee[1]*, Ambigapathy Pandian[2] and Souba Rethinasamy[2]

[1]School of Languages, Literacies and Translation, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[2]Faculty of Language and Communication, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

ABSTRACT

This quasi-experimental study explored the impact pictures had on the vocabulary depth knowledge of 7-year-olds in rural Malaysia. The Picture Word Inductive Model (PWIM) is designed to elicit words from students' existing listening and speaking vocabularies while capitalising on a learner's ability to think inductively. Modified and adopting a single-factor one-cycle design with PWIM as the independent variable, the intervention employed a pretest-posttest-delayed posttest control-group design. The subjects were comprised of 2 intact classes ($n = 60$), 1 class formed the Experimental group ($n = 30$) and another formed the Control group ($n = 30$). The primary testing instrument, Read's Word Associates Test (WAT), was specifically adapted to measure subjects' vocabulary depth knowledge for recall and retention. Analysed results revealed superior recall and retention by the Experimental group, denoting the effectiveness of PWIM in enhancing vocabulary depth knowledge recall and retention among Malaysian young learners. Moreover, given vocabulary depth's claim as the stronger predictor of language proficiency, the functionalities of such data in providing insights on effective measures to boost English language development, particularly among young learners, cannot be disputed.

Keywords: PWIM, recall and retention, vocabulary depth, young learners

ARTICLE INFO

Article history:
Received: 10 December 2019
Accepted: 30 January 2020
Published: 18 March 2020

E-mail addresses:
beechoo_lee@usm.my (Bee Choo Lee)
pambigapathy@unimas.my (Ambigapathy Pandian)
rsouba@unimas.my (Souba Rethinasamy)
* Corresponding author

INTRODUCTION

Acknowledging the significance of mastering English language, both for academic as well as career purposes, the Ministry of Education in Malaysia highlighted the need for students to excel in the language via various English language education reforms such as the Primary School Standards-

ISSN: 0128-7702
e-ISSN 2231-8534

© Universiti Putra Malaysia Press

Based Curriculum for English language education (SBELC), introduced in 2011, the Malaysia Education Blueprint (2013-2025) (MEB, n.d.) as well as the Malaysia English Language Roadmap (2013-2025) (MELR, n.d.) and the latest being the Subject Grade Point Average (Gred Purata Mata Pelajaran [GPMP]), a set target for English language to ascertain that schools in every state perform well in the subject. Yet, our students still fail to attain a reasonable level of competency and unemployment among our graduates remains high with the main reason given as poor communication skills due to lack of English proficiency ("Poor command of English", 2018).

The present study pinpoints vocabulary acquisition on account of the robust correlation between vocabulary knowledge and language mastery, and all-inclusive academic excellence as well. Whilst researchers such as Laufer (1997), Laufer and Sim (1985) and Nation (1994) stood unanimous in their opinions that language learning would occur once learners had acquired an adequate vocabulary size, learners and teachers alike view vocabulary as a vital requisite for first and second language learning (Duppenthaler, 2007). Specifically, the study looks at vocabulary depth recall and retention given its status as a stronger predictor of language competency compared to its other dimension, breadth (Hadley & Dickson, 2018; Hadley et al., 2016; Hoffman et al., 2014).

Employing pictures, the present study's intervention is via PWIM, a teaching model that leverages on inquiry-oriented arts strategy. Designed and continuously researched upon, and refined by Joyce et al. (2015, 2009), the PWIM is deemed sufficiently versatile to be engaged with learners of any age group, either individually or in groups. Calhoun's (1999) earliest effort in experimenting with PWIM to promote literacy was with preschoolers and middle school learners but she eventually reached out to include adult non-readers. Her study documented significant positive findings in terms of vocabulary acquisition, reading skills and comprehension among 22 first graders via action research (Joyce et al., 2001).

The numerous recent studies conducted by local researchers such as Adibah Halilah et al. (2014), Lee et al. (2017), Susanto (2017) and Tan (2016a) on vocabulary acquisition indicate rejuvenated interest in second language vocabulary acquisition, as well as attempts at addressing the lack of proficiency in English language given the significance of vocabulary knowledge to overall language learning and development, and academic

Purpose of Study

Past and present Malaysian education reform initiatives have irrevocably reverted to English language education in both primary and secondary schools, with attention on the issue of underachievement in terms of English language proficiency among Malaysian learners, be it young or old. As observed by Misbah et al. (2017) in their study on factors contributing to young Malaysian learners' difficulties in English

language learning, lacking vocabulary knowledge among learners is a major causal dynamic. Consistent research findings have also pointed to the possibility that a gap in vocabulary knowledge among learners, including those in preschool, will eventually lead to compromised language proficiency in their later years (Biemiller & Slonim, 2001; Hart & Risley, 2003; Stanovich et al., 1998; Wagner et al., 1997). Therefore, it is of critical importance to address the issue of insufficient vocabulary knowledge without further delay (Biemiller, 2003; Hadley & Dickinson, 2018) and that includes attending to the vocabulary needs of children as young as preschool (McKeown & Beck, 2014).

Witnessing fellow researchers in solidarity on the need to increase learners' vocabulary knowledge for language development is certainly motivating. However, a fundamental question needs to be addressed; should measurements of vocabulary knowledge be all-inclusive or should its different dimension be taken into account - breadth, the quantity or depth, the quality? While Nation (2006) elucidated vocabulary depth as to how well a word was comprehended or applied by language learners, vocabulary breadth was simply referred to as the number of words - in numerical - known to a language learner. Such a dilemma is justifiable since young learners' vocabulary assessment is largely breadth based as the main indicator of vocabulary knowledge (Hadley et al., 2016). A report by National Early Literacy Panel (NELP, 2008) that defined vocabulary as encompassing only vocabulary breadth showed young learners' (kindergarten) vocabulary breadth size to be a weak predictor (r = 0.25) of reading comprehension. Vocabulary depth test recorded a significantly stronger predictor (r = 0.45) in later tests.

Hence, the decision to explore the Picture Word Inductive Model (PWIM) as an intervention for effective vocabulary acquisition and to measure vocabulary depth in this study. To this end, attention is on how pictures, viewed as a key component in cognitive processing (Joyce et al., 2015) that serves to assist young learners in acquiring, recalling and retaining the meaning of a new word over time, impact the recall and retention of vocabulary depth knowledge of our young learners via Calhoun's (1999) PWIM. This study attempts to provide answers to the following questions (RQ) and null hypotheses (H_o):

RQ1: Is there a difference in the vocabulary depth knowledge recall between the Experimental group and the Control group?

$H_{o1:}$ There is no significant difference in the vocabulary depth knowledge recall between the Experimental group and the Control group.

RQ2: Is there a difference in the vocabulary depth knowledge retention between the Experimental group and the Control group?

$H_{o2:}$ There is no significant difference in the vocabulary depth

knowledge retention between the Experimental group and the Control group.

Literature Review

Vocabulary Depth. Anderson and Freebody's (1981) definition of vocabulary depth as to how well words are known is infamous for its straightforwardness and easy comprehension. Vocabulary has been long viewed as a prerequisite and a worthy determinant of language proficiency and that it embraces two dimensions, breadth and depth as two individual but closely interconnected constructs. Testimonies to this are studies carried out by Lee (2013), Qian (2002, 1998), Read (1989, 1988), Tan (2016b), Tannenbaum et al. (2006) and Wesche and Paribakht (1996) with findings specific to vocabulary breadth and depth.

More studies on vocabulary testing (Cain & Oakhill, 2014; Dickinson et al., 2010; Proctor et al., 2012; Roth et al., 2002) merely corroborated the NELP's report mentioned earlier where the majority of vocabulary research on young learners demonstrated over-reliance on standardized assessments intended for breadth. In short, though helpful, as well as necessary in supplying an indicative vocabulary threshold, vocabulary size certainly falls short as a primary indicator for a comprehensive study.

Nation's (1990) four constructs-form, meaning, function and position-are tapered from Richard's (as cited in Shen, 2008) version that contains seven facets of word knowledge, a framework largely responsible for the subsequent varied yet complementary frameworks. Nation subsequently expanded the list to include receptive and productive vocabulary. In tandem, Qian's (2002) framework comprises vocabulary size, depth of vocabulary knowledge, lexical organization, and automaticity of receptive–productive knowledge. As stressed by Beglar (2010) and Harmer (2007), words should be studied in context, no less. Given the multi-facets of depth, it seems only right to measure learners for vocabulary depth.

There appears to be a general consensus that word knowledge progresses along a continuum (Waring, 2002) that spans levels and dimensions. Depth knowledge is seen as being positioned at various lexicon levels along a continuum. Here, the underlying notion is that one travels along the continuum of knowledge as one progressively accumulates more knowledge of a particular word. Rather than being limited to a single word, this basic concept extends to cover a wider dimension of vocabulary that encompasses receptively versus productive, passive against active, recognition and understanding as opposed to recall and vocabulary employment respectively.

It is relatively safe to enunciate that vocabulary depth is a critical and yet a largely under-researched dimension in the area of language learning; a destitute state indeed. In the present study, Read's (1998) WAT has been modified specifically to measure subjects' vocabulary depth knowledge for recall and retention.

Picture Superiority Effect (PSE). The greater conceptual processing pictures possess over words is often justified via the Picture Superiority Effect (PSE) with the support of encoding theories and transfer-appropriate processing (TAP). PSE is often explicated through Paivio's (1971) widely quoted dual-coding theory, the sensory-semantic theory (Nelson et al., 1977) and TAP by Weldon and Roediger (1987). Seemingly, it is the way pictures are decoded differently from words that account for picture superiority hypothesis. Pictures have the definite advantage of eliciting both verbal and image code whereas single-coded words are encoded verbally. Thus, the edge pictures enjoy over words in the encoding theory. Particularly when it concerns young learners, this notion found support with researchers such as Ally and Budson (2007), Alvarez and Oliva (2008), Anderson (2009) and Lee (2013) who collectively asserted on the superiority of a dual coded representation over a single coded one.

Nelson and his fellow colleagues (Nelson et al., 1977, 1976) proposed the Sensory-Semantic Theory, also an encoding theory of picture superiority that suggested a difference could be found in the order of access to phonemic information for both sensory and semantic codes. For pictures, semantic processing occurs prior to the availability of name codes since phonemic access is not direct. Alternatively, phonemic access for words is direct and therefore does not necessitate prior semantic processing. However, elevated levels of either schematic or conceptual similarity in pictures that lead to the distorted sequential arrangement can reverse or even eliminate PSE.

As for TAP, PSE can be experienced via the interrelation between encoding and retrieval (Weldon et al., 1989; Weldon & Roediger, 1987). Particularly on tasks pertaining to recall and recognition, the superior performance of pictures will be more palpable on those that entail conceptual retrieval (Weldon & Roediger, 1987). Pictures are often seen as being more conceivable to access meaning during encoding.

PWIM's basic structure is nearest to Paivio's dual-code model as it relies on two inter-reliant modes of memory codes, verbal and nonverbal, to the first process and then stores information. Apparently, there exists two disparate but interrelated modes of representational units (*imagen* and *logogens)* that refer to image illustrations and verbal entities depicting linguistic materials such as vocabulary items and sentences respectively. Hence, one can construe that retrieval and retention of monocoded information are more challenging as compared to dual-coded ones. Paivio (1971) justified by explaining that images were processed much faster through synchronous processing as opposed to the sequential access of the information in the verbal system. Paivio and Begg (1981) exemplified Paivio's dual-coding theory as in Table 1.

The referential connections between a verbal and a nonverbal mental representation, as well as the concurrent storage in short-term memory, are made possible by the close presentation of predominantly two

Table 1
The two systems that serve the memory

Sensory modality	Symbolic systems	
	Verbal	Nonverbal
Visual	Printed words	Picture or objects
Auditory	Speech sounds	Environmental sounds
Tactual	Braille	Fellable objects
Kinaesthetic	Motor feedback from writing	Motor feedback from the physical exploration of objects

Notes: Adapted from *Psychology of Language* (p. 68) by Paivio and Begg (1981)

different types of information (Chun & Plass as cited in Wu, 2014). Such linkages give rise to increased retrieval avenues for accessing of vocabulary items. Any increase in dependency on recall is also largely determined/influenced by learner variables such as one's preferences and abilities. In sum, vocabulary learning is deemed profoundly enhanced when learning involves the simultaneous application of both verbal and nonverbal system.

The Picture Word Inductive Model (PWIM). The Picture Word Inductive Model (PWIM) is "...an inquiry-oriented language arts strategy that uses pictures containing familiar objects and actions to elicit words from children's listening and speaking vocabularies" (Calhoun, 1999, p. 21). Leaning on the inherent attributes of learners in learning, the approach strives to expand knowledge using language in print and simultaneously, to develop and hone skills on extricating and compiling information.

PWIM primarily uses pictures composed of commonly known objects and actions to exploit one's facility for thinking inductively to draw words from prevailing listening and speaking vocabulary. Learners are first steered into probing for words to acquire a gradually increasing number of words for sight-reading and writing. Learners' observation and analysis skills are further cultivated in detecting and learning of fundamental phonic elements and language structures before stretching to envelope skills such as reading, writing, comprehending and composition. Basically, the instructional sequence of the model cycles and recycles through the steps of the PWIM. The length of a complete sequence of a PWIM cycle is extremely versatile, largely determined by aspects such as richness of the image, the level of maturity and language proficiency of the learners as well as the intentions and language goals of the teacher (Calhoun, 1999).

PWIM further integrates concept attainment into its model of teaching, making it also a "safety net" programme that is exceedingly effective and popular as well with the more mature beginning readers and writers in secondary and high school, and adult learners too (Joyce et al., 2004). Generally, the instructional sequence of PWIM generates a learning environment that stresses on brain-compatible teaching

and learning, consequently serving as a base for subsequent learning. As an integral part of the language arts curriculum, PWIM has successfully aided learners in expanding their sight vocabularies whether individually or in groups (Calhoun, 1999; Joyce & Calhoun, 1998), irrespective of age, gender or ethnicity.

Inductive Thinking. Within PWIM, inductive thinking inculcates concept formation and application in learners (Joyce et al., 2009). The research efforts of Taba (as cited in Joyce et al., 2004; Wragg, 2012) on approaches to pursue and organise information, to construct hypothesis and assess it, and to illustrate correlations among sets of collected data have been incorporated into PWIM and subsequently refined by Joyce and Calhoun (1998, 1996).

Learners are perceived as natural conceptualises that observe both similarities and dissimilarities between items, incidents or experiences and sentiments or reactions (Joyce et al., 2009). Hence, typical of the PWIM approach in increasing and augmenting one's innate potentiality, learners are coaxed and motivated to build and expand the concept. Consistently employed, this strategy also made it possible to widen learners' scope of perspectives to better review information collected.

The lesson structure of PWIM is devised in accordance to four stages ascertained in the inductive model of teaching (Joyce et al., 2009), stage one is identifying and itemising of data while the second stage is classifying them according to general elements. In the third and final phase, data is first analysed and interpreted before being converted to either skills or hypotheses. It is crucial that learners are closely guided throughout the lesson to think inductively; they are not to be spoon-fed. Hence, PWIM aspires for learners to focus and attain conceptual control, and to convert conceptual understanding to skills; the classification, reclassification, and development of a hypothesis.

Concept Attainment. Concept attainment refers to "the search for and listing of attributes that can be used to distinguish exemplars from nonexemplars of various categories" (Bruner et al. as cited in Crawford & Nicklaus, 2013). The design of PWIM provides for instructions on distinct concepts that are precise in distinctiveness with plentiful opportunities for learners to further refine the various learning strategies. It is an effectual avenue that categorises information comprising vast scope of topics, ready to be dispensed to learners at the various juncture of learning. Within an environment that already enjoys formed groups or categories, learners just need to pinpoint the attributes of the groups. However, to successfully do so learners must first differentiate the dissimilarities from the similarities of *exemplars*, effectively segregating those with traits of the concept from those without. It is through comparing and contrasting of *exemplars* that learners gradually learn to grasp the fundamentals of forming hypotheses (Joyce et al., 2009). All in, the concept attainment model contains three phases that generally guide PWIM's

lesson structure. The initial two stages concern providing data to learners and assessing them on the attainment of the concept respectively while the later phase is to correctly label secondary examples not identified and then to go about creating their own examples (Joyce et al., 2009).

Developmental Priming Mechanisms. Ramey et al. (1995) first applied its developmental priming mechanisms to early intervention programmes comprising two generations. As a component of a theoretical framework derived from the General Systems Theory, its mechanisms chiefly concern the outcomes of children in terms of cognitive, social, and emotional facets though the progress of adults development is not overlooked either. The priming mechanism needs to be present on a consistent basis in children's daily lives for optimum positive outcomes (Ramey & Ramey, 1998). Otherwise, as Ramey and Ramey (2002) claimed, the children's development would be negatively affected cognitively, linguistically and socially.

PWIM's instructional environment is closely oriented to the conditions expressed by Ramey and Ramey (Calhoun, 1999), as summarised in Figure 1 below:

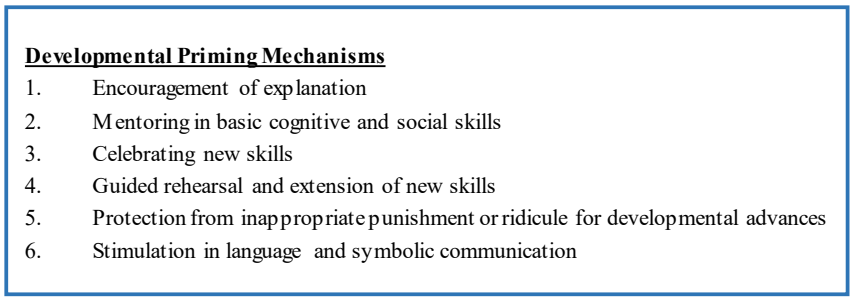

Figure 1. Developmental priming mechanisms. Adapted from *Early Intervention and Early Experience* by Ramey and Ramey (1998)

As asserted by Calhoun (1999), the instructional environment within the PWIM framework addresses "five of the six priming mechanisms for continued development"; Ramey and Ramey's fifth priming mechanisms was omitted. Thus, students in the intervention group similarly experienced the five priming mechanisms that exist within PWIM's instructional environment.

Research Framework

Figure 2 shows the research framework of this study.

METHODS

Research Design

The present quasi-experimental research utilised the pre-test, post-test and delayed posttest control group design, with purposive

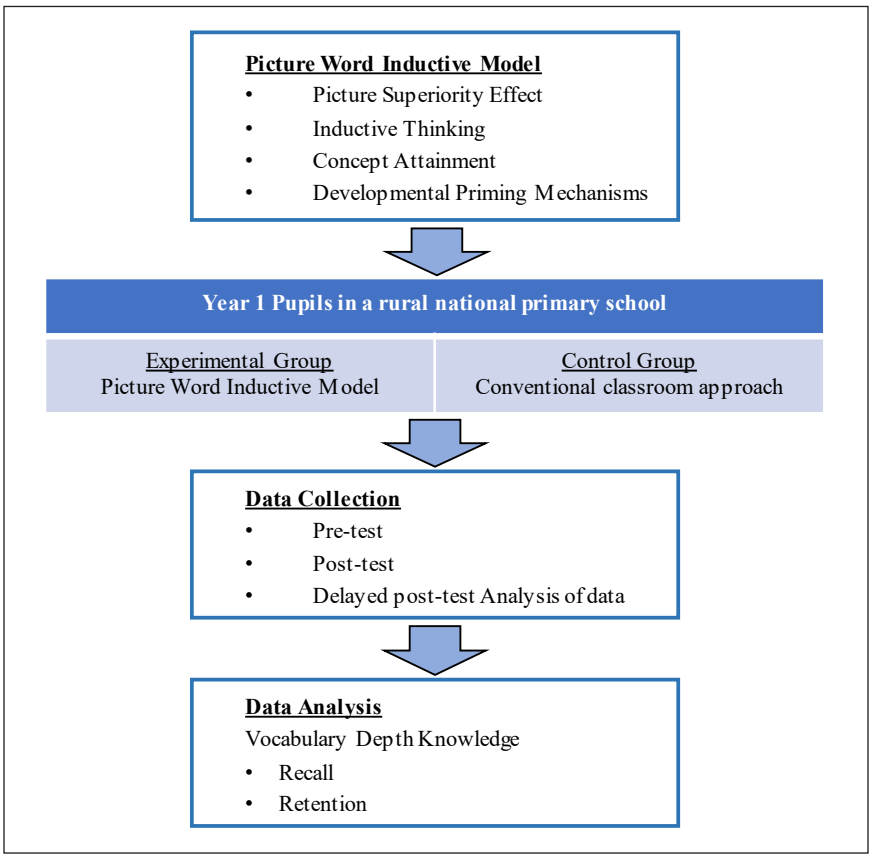

Figure 2. The framework of the present study

sampling. Such design is often employed for classroom experiments that come with naturally assembled experimental and control groups such as intact classes that are likely to be homogeneous (Best & Khan, 2006), particularly when the school refrains from the practice of streaming its pupils. The choice is further strengthened with Moore and McCabe's (2006) plausible recommendation for the experimental design as one of the most effective methods to ascertain causation within a design that extends control over possible variables. The presence or absence technique allows for manipulation of an independent variable in an effort to ascertain causal correlations, accomplished by conferring an intervention or condition on one group while denying the control or comparison group of it.

Participants

Subjects targeted in the study are young learners since past research findings deem it crucial to begin vocabulary knowledge building early (Hadley & Dickinson, 2018; Hoffman et al., 2014; McKeown & Beck, 2014) as those with smaller vocabularies have been found to lag behind their peers who possess larger vocabularies in literacy development (Biemiller & Slonim, 2001).

Hence, Year 1 pupils were purposively selected. As stated by Teddlie and Yu (2017), a purposive sample such as individuals, groups of individuals or institutions is selected with the precise intention related to seeking an answer to a research study's questions and hypothesis (Teddlie & Yu, 2007) and can be transferred absolutely to precise cohort (Barbie, 2008) because it is justifiable for researchers to form generalisations (Black, 1999). Utilising purposive sampling, a coin was tossed to resolve the random assignments of subjects to the experimental and control group. In the case of the present study participants comprised of two intact groups (n = 60) of Year 1 pupils enrolled in a rural national primary school.

Instruments

Enlarged pictures are taken from the Year 1 textbook and vocabulary depth tests made up the instruments in the study. The list of words utilised for the intervention was randomly selected from the words listed in the word list in the *Dokumen Standard Kurikulum Pentaksiran* (Kementerian Pendidikan Malaysia [KPM], 2017). The teacher who participated in the study was given discretionary power to select the pictures used in the study, in line with Calhoun's (1999) suggestion that the pictures are selected by the teacher. Meanwhile, the vocabulary depth test employed in the study is based on Read's (1998) Word Associates Test, guided by (Schoonen & Verhallen, 2008). WAT has been validated with strong scores, for instance 0.93 (Read, 1998), 0.84 - 0.89 (Greidanus et al., 2004), 0.75 (Schoonen & Verhallen, 2008). It has also been subjected to a reliability test where at p = 0.891, the Sig. (p) value obtained is more than 0.05 ($p > 0.05$), denoting statistically no significant difference between group means. Thus, signifying the test's reliability in terms of stability over time. A sample of the WAT test employed in the present study to measure vocabulary depth is in Figure 3.

Scoring

The All-or-Nothing method was selected over the One-Point and Correct-Wrong

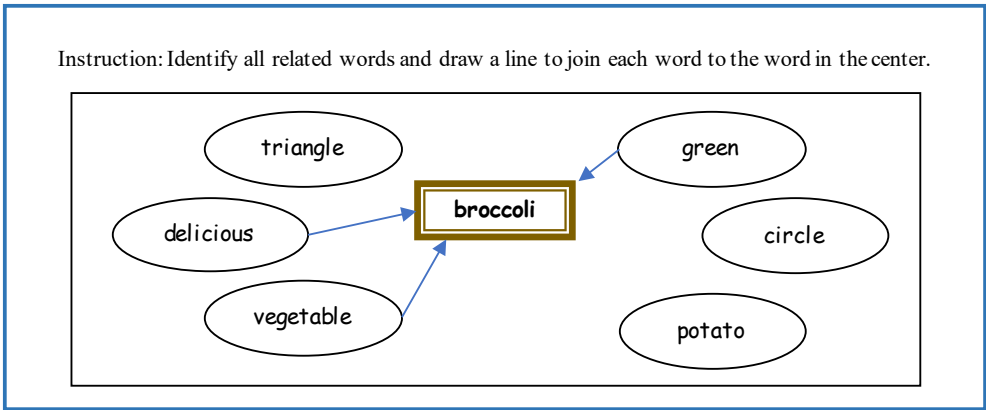

Figure 3. Vocabulary depth test adapted from Read's (1998) WAT

methods due to its simplicity in scoring and its strictness in awarding points. According to Schoonen and Verhallen (as cited in Zhang & Koda, 2017), the All-or-Nothing method awards points only for precise matching of correct answers which prevents testees from attempting to guess the correct answers. The One-Point and Correct-Wrong methods were rejected due to the weakness detected in them. For instance, the One-Point method only scores learners' answers according to their associate selection, completely disregarding the selection and non-selection of a distractor. On the other hand, the Correct-Wrong method was censured for its complexity and tediousness (Zhang & Koda, 2017). Both expose studies to threats of distorted data.

Scoring for the vocabulary depth test is mechanical and as such, tests to investigate inter-rater reliability are omitted. As stated by Cohen et al. (2018), inter-rater reliability only needs to be established to ensure the consistency of two or more raters, typical of studies pertaining to writing skills.

Procedure

Intervention. Guided by Calhoun's (1999) PWIM model of instruction, the structure of the PWIM was further adapted accordingly to accommodate the study based on its objectives. The steps of the intervention for the present study are in Figure 4.

Employing a single independent variable within a one-cycle design, the intervention occurred once a week at the participating school with the permission of its headmaster. Five words were taught during each session that lasted approximately 30 minutes; within one cycle ten words were taught to the experimental group while the control group received traditional classroom instructions. The present study stretched over a period of approximately six weeks.

Bearing in mind the Hawthorne effect, the researcher refrained from conducting the intervention. Instead, an English teacher in the participating school took charge of the experimental group upon receiving instructions in the steps of the intervention while the control group was held by the English teacher who was their class teacher at that point.

Steps of the intervention

1. Choose a picture, enlarge and hand on the board.
2. Point at an item and invite pupils to tell what they see
3. Tag the items identified. (Draw a line from the identified object or area, say the word, write the word; ask students to spell the word aloud and then pronounce it).
4. Read and review the picture word chart aloud
5. Ask students to read the words (using the lines on the chart if necessary) and to classify the words into a variety of groups. Identify common concepts (e.g., beginning consonants, rhyming words) to emphasize with the whole class.
6. Read and review the picture word chart (say the word, spell it, say it again)
7. Add words, if desired, to the picture word chart and to the word banks.

Figure 4. Lesson structure of the intervention in the present study

Attempts were made to standardise the conditions in which the study took place in order to minimise both external and internal threats. For instance, random selection and assignment of intact groups were practised to negate some threats. Also, being a single intervention effectively rules out the multiple-treatment interference, an external threat to a research study. Additionally, the pupils involved were not informed of their role in the study. To the knowledge of the authors and the participating teachers, subjects remained unaware of their status as research participants as the intervention took place during normal lessons and the picture used was taken from their Year 1 English textbook, *Supermind*. Teachers concerned were also cautioned against mentioning of the intervention to anyone. Hence, minimising another threat that could jeopardise research findings. As for internal threats, namely historical and maturation of subjects, they have been nullified to a greater extent with the existence of a control group and the comparatively short duration of the intervention as well. Regarding the mortality threat, it was non-existent since all subjects remained throughout the intervention and all were pre-, post- and delayed post-tested.

It is to be noted that this study is a piloting effort and therefore, is a scale down version of the actual study that numbered 180 subjects from a total of three primary schools located in rural Perak, Malaysia.

Data Collection. The participants in both experimental and control groups were administered vocabulary pre-tests, post-tests and delayed post-tests. Participants were subjected to pre-test before the intervention began while post-test was administered immediately upon completion of the intervention. Delayed post-tests took place approximately two weeks after the intervention was completed. The schedule is as summarised in Figure 5.

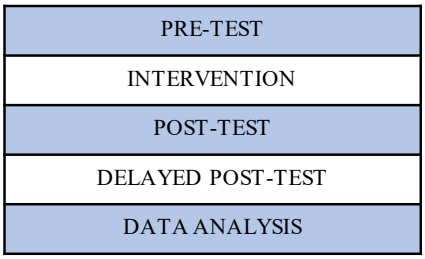

Figure 5. The schedule of the present study

Data Analysis. To investigate the effect of the PWIM on the acquisition of subjects' vocabulary depth knowledge, the scores obtained from pre-, post- and delayed post-tests were tabulated and duly analysed via the Statistical Package for the Social Sciences (SPSS). To assess the assumption of normality, Sharpiro-Wilk test was conducted as it is considered more appropriate than the Kolmogorov-Smirnov test due to the smaller sample size of the present study. Results suggested that samples are consistent with a Gaussian distribution as all the Sig. (p) values obtained are greater than 0.05 ($p > 0.05$). Thus, allowing for the acceptance of the null hypothesis that sample distribution is normal and non-violation of the assumption of normality. Parametric procedures were employed.

RESULTS

Levene's Test Results

Levene's Test for Equality of Variance was conducted to test the assumption of homogeneity of variances whereby the null-hypothesis (H_o: $\sigma^2_1 = \sigma^2_2$) assumes no difference between the groups' variances. The results as in Table 2.

Table 2
Levene's Test for the assumption of homogeneity of variances

Test	*F*	*Sig.*
Pre-test	0.489	0.487
Post-test	0.545	0.463
Delayed post-test	1.117	0.295

As shown, the Sig. (*p*) value obtained for the pre-, post- and delayed posttest was higher than 0.05 at 0.487, 0.463 and 0.0295 respectively; resulting in the null hypothesis of no difference between the groups' variances accepted given that value of $p > 0.05$. This indicates that the assumption of homogeneity of variances was met and no significant difference detected between the variances of the two groups.

Pre-test Results

To ascertain for baseline similarity in vocabulary depth knowledge between the two groups, pre-test scores were analysed via an independent sample t-test. Tabulated results at 2.60 (*SD* = 0.814) and 2.57 (*SD* = 0.728) for the experimental and control group respectively denote achievement of baseline similarity between these two groups. Table 3 shows that the pre-test mean scores for both experimental and control group differed only slightly, 0.03.

Table 3
Pre-test results: Vocabulary depth knowledge

Group	n	Pre-test	
		M	SD
Experimental	30	2.60	0.814
Control	30	2.57	0.728

Additionally, employing an alpha level of 0.05, an independent samples *t*-test was conducted to determine if statistically significant differences exist between the pre-test group means.

The null hypothesis of no significant difference in vocabulary depth at the pre-test level between groups was tested. Table 4 shows that at this level, with the Sig. (*p*) value obtained at more than 0.05 ($p > 0.05$) at $p = 0.868$, there is no statistically significant difference between group means. Thus, allowing for the acceptance of the null hypothesis of no significant difference in vocabulary depth between the Experimental group and the Control group; both groups

Table 4
Independent Samples T-test: Comparison between groups for pre-test

	t-test for Equality of Means					95% confidence interval of the difference	
Equal variances assumed	*t*	*df*	Sig.	M diff	SE diff	Lower bound	Upper bound
Pre-test	0.167	58	0.868	0.033	0.199	-0.366	0.432

possessed comparable vocabulary depth knowledge at the pre-test level.

Post-test Results

Vocabulary Knowledge Recall for Depth. To answer the first research question, the mean scores of the post-test for both experimental and control group were compared. As detected in Table 5, the mean score for the experimental group's post-test is 5.13 (SD = 0.900) as compared to 4.73 (SD = 0.691) for the control group. These numbers illustrated a higher score (M = 5.13) for the experimental group, thus indicating a higher vocabulary depth knowledge recall for those exposed to the intervention.

In addition, employing an alpha level of 0.05, an independent samples t-test was conducted to determine for statistically significant differences between the post-test group means; the first hypothesis tested. Referring to Table 6, the Sig. (p) value documented at p = 0.005 is less than 0.05 ($p < 0.05$), denoting statistically significant differences between group means at the post-test level. Therefore, overall results signify that the experimental group performed significantly better than the control group in vocabulary depth knowledge recall.

Table 5
Post-test results: Vocabulary knowledge recall for depth

Group	n	Post-test	
		M	SD
Experimental	30	5.13	0.900
Control	30	4.73	0.691

Post-test / Delayed Post-test Results

Vocabulary Knowledge Retention for Depth. The descriptive statistics in Tables 7 and 8 provide answers to the second research question and hypothesis. Table 7 shows the mean difference in the post- and delayed post-test scored by participants that measured the retention of vocabulary knowledge. The experimental group recorded 0.90 (SD = 0.548) while the control group obtained 1.33 (SD = 0.479), a larger difference in mean scores than the experimental group that indicates lower retention. Basically, retention is based on how well knowledge is maintained after a certain period of experiencing an intervention. Therefore, more knowledge lost (bigger difference) equates lower retention.

Following that, the existing differences were checked for statistic significance with the employment of an alpha level of 0.05. As seen in Table 8, the Sig. (p) value obtained is less than 0.05 ($p < 0.05$) at p

Table 6
Independent Samples T-test: Comparison between groups for post-test

	t-test for Equality of Means					95% confidence interval of the difference	
Equal variances assumed	t	df	Sig.	M diff	SE diff	Lower bound	Upper bound
Post-test	1.931	58	0.005	0.604	0.307	0.105	0.815

= 0.002, indicating statistically significant differences between group mean at the post- and delayed post-test level; the null hypothesis of no significant difference in the retention of vocabulary depth between groups was tested. Overall, results signify that the experimental group out-performed the control group significantly in vocabulary depth knowledge retention.

Table 7
Post- and Delayed post-test results: Vocabulary knowledge retention for depth

Group	n	Differences between Post-test and Delayed Post-test	
		M	SD
Experimental	30	0.90	0.548
Control	30	1.33	0.479

Table 8
Independent Samples T-test: Comparison between groups for differences in post- and delayed post-test

	t-test for Equality of Means					95% confidence interval of the difference	
Equal variances assumed	*t*	*df*	Sig.	M diff	SE diff	Lower bound	Upper bound
Post-test	-3.261	58	0.002	0.433	0.133	-0.699	-0.167

DISCUSSIONS

In employing PWIM on a group of young learners with the aim of exploring the effectiveness of the approach on vocabulary learning, specifically to measure vocabulary depth, opportunities for incidental vocabulary learning through speaking and listening as well as deliberate teaching and learning could be detected. The overall results have reflected significant differences in the mean scores of the post-test and post-test/delayed post-test between the groups. Descriptive statistics show that the experimental group which experienced the PWIM intervention recorded statistically significantly higher vocabulary recall and retention (see Table 6 & 8 respectively) than did the individuals in the control group. Thus, conclusions reached are that the recall and retention of vocabulary depth knowledge are effectively enhanced through the PWIM model of teaching.

Recall ability is likely enhanced through conceptual processing as observed within the framework of PWIM. This deduction is in accord with previous frameworks of vocabulary knowledge by Roediger (as cited in Lee et al., 2017) that advocated conceptual processing for word recall and recognition, and that improved conceptual coding forms the basis for PSE. Based on data collected from this research effort, it seems that pictures are superior in the recall due to being encoded with higher conceptual processing in addition to repetition for effective retrieval which lends support to Thornbury's (2004) research findings. As shown, PWIM's framework comprises multiple cycles that entail the repeating of new words induced. Obviously, meaningful

repetitions play a significant role in the storage and retrieval of information/ knowledge as found reiterated in the work of researchers renowned in this domain of study. For instance, Nagy et al's findings in 1987 were referred to by Krashen (2013) and Webb (2008).

Based on this study's positive findings on vocabulary depth recall, it appears that picture stimuli do surpass word stimuli every time for superior memory performance, as asserted in PSE. This was similarly highlighted in McBride and Dosher's (2002) work. In the Malaysian context, Tan and Parimala (2014) discovered in a more recent study that young learners demonstrated greater recall ability when pictures were elaborated with verbal accompaniment. Also, distractions for words tended to incur greater decline in memory performance compared to pictures.

The superior performance in vocabulary depth retention by those in the experimental group, as exhibited in this study's output (see Table 8), evinced the effectiveness of combining images with text in the context of retention of new information. A study carried out by Lee (2013) on using pictures with Malaysian pupils similarly documented improved vocabulary retention though it was the breadth and not depth dimension. Evidently, the combination of picture and text works towards fostering and enhancing of information retention. This supports the stance taken by researchers such as Nemati (2009) and Makoto and Jeffa (2002) on utilising pictures for enhanced second language learning. While Nemati (2009) emphasized that one should not undermine the efficacy of images for long term retention, Makoto and Jaffa (2002) expounded on the virtue of combining of picture and text with annotations for better recall and retention based on the post and delayed posttests results in their experimental study. Evidently, as discovered by Anderson (2009), the meaningfulness content of an item can be crucial to the retention of information.

Notwithstanding the markedly strong support for multiple repetitions for effective knowledge retention, Nemati (2009) had rebuffed the shallower processes of repetitions when it came to effective retention. Instead, a deeper level processing of the target words was advocated. Even as early as in the 1990s, Schmitt and Schmitt (1995) had advocated a more profound processing level as retention necessitates deeper and richer semantic processing; rote repetitions is simply not adequate if effective long term retention is the ultimate goal. Nevertheless, this study's significantly positive findings, as in Lee's (2013) research study on breadth, denote that multiple repetitions produce affirmative effects on vocabulary recall and retention. It is to be mentioned that PWIM upholds meaningful repetitions rather than rote repetitions in its lesson structure.

Though numerous studies on PWIM can be detected of late, especially by Indonesian researchers, many of these studies explored its impact on writing skills. Take, for instance, Yuniati (2015) who carried out a study on writing recount

text while Yuniyarsih and Saun (2014) focused on the narrative writing skills of high school students via PWIM. As for Nurani and Rosyada (2017), their area of research covered high school students' communicative competence via PWIM. Positive findings may be documented for the above-mentioned research but then such studies cannot be considered comparable since the focuses of the studies differ from that of the present study.

Previous PWIM studies on vocabulary acquisition specifically include efforts by Yurfalah (2014), Jiang (2014), Lee (2013) and Li (2011). Even so, differences between theirs and the present study lie in the domain measured, test instruments, age of subjects among others. Yurfalah (2014) measured her subjects for vocabulary breadth recall and at the same time examined their perceptions on how PWIM impacted their vocabulary learning in class. As for Jiang, a mainland Chinese who conducted his study on Chinese learners in 2014, he combined both vocabulary breadth and depth in one test instrument to investigate vocabulary recall. Likewise, Li (2011) who similarly integrated both depth and breadth into a single test to examine fourth-grade Swedish pupils' ($n = 16$) second language vocabulary acquisition that did not extend beyond a post-test. Hence, there is no differentiation in the vocabulary dimension when findings were reported. On the other hand, Lee (2013) assessed her subjects for vocabulary retention but her study did not extend to cover depth. Therefore, among recent comparable studies mentioned, either it was limited to exploring vocabulary recall only or recall and retention for breadth, or else it examined breadth and depth as one. Although all studies mentioned have posted positive findings, at best it can only be generalized that this study's research findings are in tandem with the findings of previous PWIM studies on vocabulary learning that similarly documented positive results. To the researchers' best knowledge there are no published studies that similarly explored the influence of PWIM on the recall and retention of young learners' vocabulary depth as what this paper has achieved.

Modestly, it can be claimed that this study's findings are more comprehensive and extensive as compared to other studies similar in nature in that subjects were measured for both recall and retention, targeting vocabulary depth knowledge precisely though admittedly the results are far from being all-conclusive.

CONCLUSION

Recall and retention of vocabulary knowledge, akin to learning, could be explored against a milieu of hypotheses that strive to elucidate the why and how of these two processes. Substantiated findings of this study indicate effective recall and retention of vocabulary depth among young learners when pictures are creatively and diligently applied for language learning. Both hypotheses are also supported in that pictures facilitate recall and retention of vocabulary knowledge when the concept attainment and inductive thinking model

are duly incorporated into a framework that is aligned to an instructional setting that adheres to Ramey and Ramey's (2002) development priming mechanisms.

Nonetheless, a blanket acknowledgement of the results is deemed unacceptable to some as the effectiveness of visual imagery may not appertain to lexical across the board. For instance, pictures may not work as successfully with abstract words. Concreteness versus vagueness in terms of word substance could be vital if not deciding factor in the dynamics of language learning via a picture-text combo. Providing food for thought, this notion stands to prompt future research on the effectiveness of pictures against lexical of varying abstractness for enhanced vocabulary teaching and learning, consequently bridging a gap in vocabulary acquisition research via pictures.

ACKNOWLEDGEMENT

The authors wish to acknowledge and thank all 8th ILLC 2019 committee members.

REFERENCES

Adibah Halilah, A. M., Rashidah, A. K., Rashidah, R., & Faizah, A. M. (2014). Vocabulary learning strategies among Malaysian TEVT students in German Malaysian Institute

(GMI). *Procedia - Social and Behavioral Sciences, 123*(2014)*,* 361-368.

doi:10.1016/j.sbspro.2014.01.1434

Ally, B. A., & Budson, A. E. (2007). The worth of pictures: Using high-density event-related potentials to understand the memorial power of pictures and the dynamics of recognition memory. *NeuroImage, 35*(1), 378-395. doi:10.1016/j.neuroimage. 2006.11.023

Anderson, J. R. (2009). *Cognitive psychology and its implications* (7th ed.). New York: Worth Publishers.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble features outside the focus of attention. *Psychological Science, 19*(4), 392-398.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, New Jersey: International Reading Association.

Babbie, E. R. (2008). *The basics of social research* (4th ed.). Belmont, California: Thomson Wadsworth.

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, *27*(1), 101-118. doi:10.1177/0265532209340194

Best, J. W., & Khan, B. (2006). *Research in education.* London: Prentice-Hall International.

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology, 2*(3 & 4), 323-335.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498-520. doi:10.1037/0022-0663.93.3.498

Black, T. R. (1999). *Doing quantitative in the social sciences*. London: Sage Publications.

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique, 114*(4), 647-662.

Calhoun, E. (1999). *Teaching beginning reading and writing with the Picture-Word Inductive Model*.

Alexandria: Association for Supervision and Curriculum Development.

Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology, 1*(1), 60-81.

Cohen, M (2012). Strengthening science vocabulary through the use of imagery interventions with college students. *Creative Education, 3*(7), 1251-1258. doi: 10.4236/ce.2012.37184

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). New York: Routledge.

Crawford, G., & Nicklaus, J. (2013). *Philosophical and cultural values: Ethics in schools*. New York: Routledge.

Dickinson, D. K., Golinkoff, R. M., & Pasek, K. H. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher, 39*(4), 305-310. doi:10.3102/0013189X10370204

Duppenthaler, P. (2007). Vocabulary acquisition: The research and its pedagogical implications. *Studies and Essays, XLII,* 1-14.

Greidanus, T., Bogaards, P., Van der Linden, E., Nienhuis, L., & de Wolf, T. (2004). The construction and validation of a deep word knowledge test for advanced learners of French. In P. Bogaards & B. Laufer-Dvorkin (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 191-208). Amsterdam: John Benjamins.

Hadley, E., & Dickinson, D. (2018). Measuring young children's word knowledge: A conceptual review. *Journal of Early Chlidhood Literacy, 0*(0), 1-29. doi:10.1177/1468798417753713

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., & Nesbitt, K. T. (*2016*). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly*, *51*(2), 181-198. doi:10.1002/rrq.130

Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Harlow, England: Pearson Longman.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27*(1), 4-9.

Hoffman, J. L., Teale, W. H., & Paciga, K. A. (2014). Assessing vocabulary learning in early childhood. *Journal of Early Childhood Literacy, 14*(4), 459-481. doi:10.1177/1468798413501184

Joyce, B., & Calhoun, E. (1996). *Creating learning experiences.* Alexandria, Egypt: Association for Supervision and Curriculum Development.

Joyce, B., & Calhoun, E. (1998). *Learning to teach inductively.* Boston: Allyn & Bacon.

Joyce, B., Hrycauk, M., & Calhoun, E. (2001). A second chance for struggling readers. *Educational Leadership, 58*(6), 42-46.

Joyce, B., Weil, M., & Calhoun, E. (2004). *Models of teaching* (7th ed.). Boston: Pearson Education Inc.

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

Joyce, B., Weil, M., & Calhoun, E. (2015). *Models of teaching* (9th ed.). Boston: Pearson Education Inc.

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

Kementerian Pendidikan Malaysia. (2017). *Dokumen Standard Kurikulum Pentaksiran Tahun 1 (Bahasa Inggeris)*. Retrieved January 12, 2019,

from https://sumberpendidikan.com/2016/07/26/dskp-bahasa-inggeris-tahun-1/

Krashen, S. (2013). Reading and vocabulary acquisition: Supporting evidence and some objections. *Iranian Journal of Language Teaching Research, 1*(1), 27-43.

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

Laufer, B., & Sim, D. D. (1985). Measuring and explaining the threshold needed for English for academic purposes texts. *Foreign Language Annals, 18*(5), 405-413. doi:10.1111/j.1944-9720.1985.tb00973.x

Lee, B. C. (2013). *A quasi-experimental study on the impact of the Picture Word Inductive Model on the vocabulary acquisition of year 1 learners in Malaysia* (Master's thesis), Universiti Sains Malaysia, Malaysia.

Lee, B. C., Tan, D. A. L., Manjet Kaur Mehar Singh, & Ganapathy, M. (2017). The significance of the academic word list among ESL tertiary students in a Malaysian public university. *3L: The Southeast Asian Journal of English Language Studies, 23*(4), 56-65. doi:10.17576/3L-2017-2304-05

Li, X. (2011). *The Picture Word Inductive Model and English vocabulary acquisition: A study in a Swedish primary school* (Master's thesis). Kristianstad University, Sweden.

Makoto, Y., & Jeffa, F. (2002). Second language incidental vocabulary retention: The effects of picture and annotation types. *CALICO Journal, 20*(1), 33-58.

Malaysia Education Blueprint (2013-2025). (n.d.). Retrieved January 10, 2019, from https://www.moe.gov.my/dasar/1207-malaysia-education-blueprint-2013-2025/file

Malaysia English Language Roadmap (2013-2025). (n.d.). Retrieved January 10, 2019, from https://www.moe.gov.my/dasarmenu/pelan-pembangunan-pendidikan-2013-2025

McBride, D. M., & Dosher, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: A process dissociation analysis. *Consciousness and Cognition, 11*(3), 423-460. doi:10.3758/BF03194397

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

Misbah, N. H., Mohamad, M., Yunus, M. M., & Ya'acob, A. (2017). Identifying the factors contributing to students' difficulties in English language learning. *Creative Education, 8*(13), 1999-2008. doi: 10.4236/ce.2017.813136

Moore, D. S., & McCabe, G. P. (2006). *Introduction to the practice of statistics* (5th ed.). New York: W. H. Freeman.

Nation, I. S. P. (1994). Review of working with words, teaching and learning vocabulary, vocabulary in action, and vocabulary. *System, 22*(2), 283-287.

Nation, I. S. P. (2006). Vocabulary: Second language. In K. Brown (Ed.), *Encyclopaedia of Language and Linguistic*s (2nd ed.) (pp. 448-454). Oxford: Elsevier.

Nagy, W. E., Anderson, R. C., & Herman, P. A., (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237-275. doi: 10.2307/1162893

National Early Literacy Panel. (2010). A summary of the process and the report. *Educational Researcher, 39*(4), 279-285. doi:10.3102/0013189X10369172

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Encoding strategy and sensory and semantic

interference. *Memory and Cognition, 5*(4), 462-67. doi: 10.3758/BF03197387

Nelson, D. L., Reed, V. S., & Walling, J. R. (1976). Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning & Memory, 2*(5), 523-528. doi:10.1037/0278-7393.2.5.523

Nemati, A. (2009). Memory vocabulary learning strategies and long-term retention. *International Journal of Vocational and Technical Education, 1*(2), 14-24. doi: 10.13189/lls.2013.010102

Nurani, S., & Rosyada, A. (2017). Implemented PWIM in developing students' communicative competence of SMK Islam Wijaya Kusuma Jakarta. *Lingua Cultura, 11*(1), 25-30.

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

Paivio, A. D., & Begg, I. (1981). *Psychology of language.* Englewood Cliffs, New Jersey: Prentice-Hall.

Poor command of English among major reasons for graduates unemployment in Malaysia. (2018). Retrieved February 15, 2019, from https://www.hrinasia.com/recruitment/poor-command-of-english-among-major-reasons-for-graduates-unemployment-in-malaysia/

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing, 25*(7), 1635-1664. doi:10.1007/s11145-011-9336-5

Qian, D.D. (1998). *Depth of vocabulary knowledge: Assessing its role in adults' reading comprehension in English as a second language* (Doctoral dissertation**),** University of Toronto, Ontario, Canada.

Qian, D.D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536. doi:10.1111/1467-9922.00193

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53*(2), 109-120. doi:10.1037/0003-066X.53.2.109

Ramey, C. T., & Ramey, S. L. (2002). *Early childhood education: From efficacy research to improved practice***.** Retrieved December 8, 2018, from https://www2.ed.gov/teachers /how/early/cognitivedevsummit02/page_pg2.html

Ramey, C. T., Ramey, S. L., Gaines, R., & Blair, C. (1995). Two-generation early interventions: A child development perspective. In S. Smith (Ed.), *Two-generation programs for families in poverty: A new intervention strategy* (pp. 199-228). Norwood, New Jersey: Ablex Publishing Corp.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal, 19*(2), 12-25. doi:10.1177/003368828801900202

Read, J. (1989). Towards a deeper assessment of vocabulary knowledge. Paper presented at *The 8th World Congress of Applied Linguistics* (pp. 16-21). Washington, D.C.: ERIC Clearinghouse on Languages and Linguistics.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). Mahwah, New Jersey: Lawrence Erlbaum.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*(9), 1043-1056. doi:10.1037//0003-066x.45.9.1043

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research, 95*(5), 259-272. doi.org/10.1080/00220670209596600

Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing, 25*(2), 211-236. doi:10.1177/0265532207086782

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: Theoretical underpinnings and practical suggestions. *ELT Journal, 49*(2), 133-143. doi:10.1093/elt/49.2.133

Shen, Z. (2008). The roles of breadth and depth of vocabulary knowledge in EFL reading performance. *Asian Social Science, 4*(12), 135-137. doi:10.5539/ass.v4n/2p135

Stanovich, K. E., Cunningham, A. E., & West, R.F. (1998). Literacy experiences and the shaping of cognition. In S. G. Paris & H. M. Wellman (Eds.), *Global prospects for education: Development, culture, and schooling* (pp. 253-288). Washington, DC: American Psychological Association. doi:10.1037/10294-009

Susanto, A. (2017). The teaching of vocabulary: A perspective. *Jurnal KATA, 1*(2), 182-191. doi:10.22216/jk.v1i2.2136

Taba, H. (1966). *Teaching strategies and cognitive functioning in elementary school children* (Cooperative Research Project No. 2404). San Francisco, CA: San Francisco State College.

Tan, D. A. L. (2016a). 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.

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

Tan, K. E., & Chandrasekaran, P. (2014). The use of contextualised storytelling to enhance Malaysian primary school pupils' reading comprehension. *The English Teacher, XLIII* (2), 79-92.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*(4), 381-398. doi:10.1207/s1532799xssr1004_3

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research*, *1*(1), 77-100. doi: 10.1177/1558689806292430

Thornbury, S. (2004). *How to teach vocabulary.* Essex, UK: Pearson Education Limited.

Wagner, R. K., Torgeson, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*(3), 468-479. doi:10.1037//0012-1649.33.3.468

Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. *Kiyo, The Occasional Papers of Notre Dame Seishin University, 46*(1), 35-41.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232-245.

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

Weldon, M. S., Roediger, H. L., & Challis, B. H. (1989). The properties of retrieval cues constrain the picture superiority effect. *Memory & Cognition, 17*(1), 95-105. doi:10.3758/BF03199561

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

Wragg, E. C. (2012). *An introduction to classroom observation* (Classic ed.). New York: Routledge.

Wu, Q. (2014). Learning ESL vocabulary with smartphones. *Procedia-Social and Behavioral Sciences, 143*(2014), 302-307. doi:10.1016/j.sbspro.2014.07.409

Yuniati, D. (2015). *The effectiveness of Picture Word Inductive Model (PWIM) for teaching recount text* (Undergraduate thesis), Universitas Muhammadiyah Purwokerto, Indonesia.

Yuniyarsih, P., & Saun, S. (2014). Using Picture Word Inductive Model (PWIM) to teach junior high school students in writing a descriptive text. *JELT, 2*(2), 193-199.

Yurfalah, D. (2014). *The use of Picture Word Inductive Model in teaching vocabulary: A study at the seventh-grade students of a state junior high school in Cidaun Cianjur* (Master's thesis), Universitas Pendidikan Indonesia, Indonesia.

Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education, 2*(1), 1-30. doi:10.1186/s40862-017-0024-0