# Show your true colours:
# Can the use of colour aid memory for Chinese speakers' English L2 vocabulary acquisition?


**Leung Ka Po**
**Wolfson College**
**University of Oxford**

Dissertation submitted to the Department of Education, University of Oxford,
in partial fulfilment of the requirements for the degree of
*Master of Science in Applied Linguistics and Second Language Acquisition*

Trinity 2022

# Acknowledgements

I would like to take this chance to express my heartfelt thanks to my supervisor Dr Elizabeth Wonnacott. I cannot stress enough how lucky I am to have Dr Elizabeth Wonnacott as my supervisor as she has been so supportive to me throughout my whole Oxford journey. She is professional, conscientious, and organised in every aspect. I truly enjoy the intellectual exchange with her in every meeting. She is willing to teach and share academic knowledge in the field with me which is one of the most valuable experiences I have ever had in my life. She is a role model to demonstrate how to be scholarly and approachable at the same time since we can communicate well without barriers and she tries to calm my emotions when I feel vulnerable. Even though she had a hectic schedule, She still has given me many precious and constructive comments on my dissertation to guide me how to write better and expand the ideas further. I do appreciate her working attitude that not only do we need to make it good, but we have to make it perfect.

I cannot ask for more from Jo Deakin who has eagle eyes in picking up my careless mistakes. She is like a walking dictionary who has a large range of vocabulary and abundant English knowledge. She always gives me a helping hand when I want to discuss English language matters. It is also a simple pleasure to find someone who can be a grammar nerd like me and talk about grammar, select the best word choice together.

Of course, without the consent of the participating school, teachers and participants, it would not be possible to deliver this research study. I am glad that they demonstrated the efficiency of Hong Kong. All parties responded quickly and handled work efficiently, this shortened the time in all procedures.

Thanks to classmates in MSc Applied Linguistics and Second Language Acquisition, friends in Wolfson College and Oxford for being kind to me. They make me realise that this master's programme is not just about academic success, but also a time for personal growth and an opportunity to feel the moment.

Last but not least, my gratitude is beyond words for the unconditional love from my family.

# Abstract

The purpose of this study is to examine the use of colours in aiding memory for English L2 vocabulary acquisition. '4-colour memory' is an innovative vocabulary strategy that encourages language learners to have self-directed learning. By reviewing their word knowledge, learners use a colour-coding system to highlight words in a colour, indicating how well they know the word. For example, blue for full knowledge of the word, green for fundamental knowledge, yellow for only impression but no knowledge, and red for no knowledge at all. This study took the first step to empirically test this method and see if it helped learners to memorise words better. 99 secondary school students in Hong Kong participated in this study and they were assigned into 3 groups. During the reading comprehension task, the 'multiple-colour' group used the '4-colour memory' approach to highlight words according to their familiarity with the word, the 'single-colour' group only used the colour yellow, and the 'no-colour' group did not use any colours. After analysing the results from pre-tests and post-tests, participants improved from the pre-test to the post-test in general. However, there were no statistical differences between the learning session and conditions so it could not be discerned whether colours helped the participants to learn. Based on the observations of participants' performance in the 3 different groups, this paper discusses some practical pedagogical implications for educators to consider. Future research avenues are also discussed with the hope that learners can grow more independent in L2 vocabulary learning.

*Keywords: short-term memory, long-term memory, working memory, colour, vocabulary strategy, vocabulary acquisition*

# Content

# List of Abbreviations

| | |
|---|---|
| BCI | Brain-computer Interface |
| CEFR | Common European Framework of Reference for Languages |
| DREC | Departmental Research Ethics Committee |
| EEG | Electroencephalogram |
| fMRI | Functional magnetic resonance imaging |
| L1 | First language |
| L2 | Second language |
| LTM | Long-term memory |
| MTL | Medial Temporal Lobe |
| NIRS | Near-infrared spectroscopy |
| PET | Positron Emission Tomography |
| SLA | Second Language Acquisition |
| STM | Short-term memory |
| VST | Vocabulary Size Test |
| WM | Working memory |

# List of Figures



# List of Tables

# 1 Introduction

## 1.1 Statement of problem

In the second language acquisition journey, a learner has to accumulate knowledge of vocabulary, grammar, pronunciation and other aspects of language in order to be proficient in a certain language. Amongst all of these, vocabulary is arguably one of the foremost components in language development (Coombe, 2011) because it is still possible to communicate ideas with little or no grammar but it is impossible to do so without vocabulary (Wilkins, 1974). Past research has shown that compared with grammar knowledge, vocabulary knowledge has a stronger impact on reading comprehension (Zhang, 2012). In the process of reading comprehension, vocabulary is needed to unlock higher-level skills, such as understanding grammatical structures, creating schemata and text models (August et al., 2005). Thus, vocabulary learning is prominent in language learning.

Given the importance of vocabulary, Meara (1995) suggested that teaching a large amount of vocabulary quickly, say 2,000 words, in early language learning stages is reasonable so that learners can enjoy linguistic and psychological advantages. However, for beginners, vocabulary in a new language may look unfamiliar as they cannot recognise the word or associate the unknown word with the meaning. Laufer (1990) suggested situations that may lead to difficulties in learning a word, for example when new words in L2 (second language) are similar to each other in terms of form or morphology. To illustrate this with a word pair ‘loose and lose’, there are phonetic similarities between the words which are hard for the learner to distinguish given their L1 (first language). Also, words in L2 are different from those in learners’ L1 in syntactic patterning or pragmatic meaning and so on. In other words, trying to learn a few hundred fundamental vocabulary words is already a challenging task for a beginner, let alone 2,000 words.

In order to attain mastery of vocabulary, memorising newly learnt vocabulary is beneficial to language learners. Ding (2007) accentuated the importance of memorisation of vocabulary as top English learners see it as a common practice on the road to success,

especially if they have limited exposure to an English speaking environment. The problem is that most language learners can memorise a vocabulary word when they first encounter it but after a while, they may forget the meaning or the spelling of the word. Deficiency in memorising English vocabulary words is found to be an overarching issue among Chinese students and they often struggle with seeking vocabulary learning strategies (Yang, 2011).

Against this background, applied linguists have been interested in investigating a wide array of vocabulary memorisation strategies to aid vocabulary learning, for example, using word cards, associating the word with a picture, linking the word with a topic or a story (Pérez & Alvira, 2017), employing real objects, teachers illustrating the word with facial expressions and gestures (Alqahtani, 2015), performing repetition, classification and word analysis (Yao et al., 2000).

### 1.2 Research aim

Apparently, most of the previous research studies in vocabulary memorisation strategies were teacher-guided, conducted in a classroom setting or required extra learning tools (Yao et al., 2000; Alqahtani, 2015; Pérez & Alvira, 2017). Therefore, it is worth looking into some vocabulary memorisation methods that are student-centred, applicable for self-learning and simple to carry out. Investigating such a method is a goal of this dissertation. In particular, the goal is to investigate a method which can be easily applied in learners' everyday life. In this regard, colour is a feasible theme since it surrounds us in our everyday life and highlighter pens seem to be commonly used when students read.

A famous quote from the late Queen Elizabeth II is that "If I wore beige, nobody would know who I am." (Hanson, 2022, para.18) This has implied the visual function of colour in commanding attention. In a similar vein, the first part of this dissertation title is 'show your true colours' since this research aims to use colours to demand language learners' attention and show their true talent in a second language.

Takashi Ishii, a Japanese writer, proposed a method called '4-colour memory' in his book, entitled 'One-Minute Tips for Effective Memorising' (Takashi, 2013). It suggests a method for improving memory by asking the learner to highlight content in 4 different colours

according to the learners' knowledge of the vocabulary, such as i) no knowledge, ii) have an impression but not familiar with the vocabulary, iii) have basic knowledge of the vocabulary, iv) understand the vocabulary thoroughly.

The method of '4-colour memory' has set the scene for this dissertation. By putting this vocabulary memorisation strategy into practice, and comparing the performance of participants using this method against the performance of a control group, this research study aims to find out the impact of colours on vocabulary acquisition. The findings presented in this study can potentially provide educators and language learners with directions on how to achieve independent learning in vocabulary acquisition with the use of memorisation strategy, thus motivating learners to study English beyond the school context.

### 1.3 Dissertation structure

Inclusive of the current chapter, this dissertation includes six chapters in total. Chapter 2 provides a theoretical framework by reviewing the literature on short-term and long-term memory, types of memory, and how colour memory is possible to aid vocabulary learning. This chapter also critically reviews relevant research studies in the past and therefore explains the significance of this study. Chapter 3 describes the methodology undertaken in this research study, including the research design, recruitment of participants, data collection instruments, procedures and ethical approval. Chapter 4 reports the results of the study, presents data analysis and considers the issue of data reliability. Chapter 5 discusses the key findings in light of previous research studies, addresses limitations of the study and educational implications, and suggests possible fields for future research. Finally, chapter 6 serves as a concluding remark to summarise the major findings of the study.

# 2 Literature Review

This chapter reviews previous memory-related literature in the second language acquisition (SLA) field. With reference to the literature, the chapter is divided into seven sections in order to: i) define various types of memory, ii) look into memory from neuroscience's perspective, iii) address the importance of attention in memory, iv) describe the relationship between colour and memory, v) focus on the contribution of colour to vocabulary learning. At the end of the chapter, there are sections to vi) explain how this current study fills the research gap and vii) state the research questions.

## 2.1 Definitions

Regarding human memory, Baddeley (1997) defined it as a system for storing and retrieving information humans acquire through senses. In Sternberg's (1999) definition, memory is the means to bring the past experiences to the present. After reviewing data on memory, learning and other cognitive aspects, Baluška and Levin (2016) broadened the scope of definition of memory to include experience-dependent modification of internal structure that changes how the system reacts in response to stimuli in the future. In short, memory is a structural process to retrieve a certain piece of stored information.

Memory can be divided into three types: short-term memory (STM), long-term memory (LTM) and working memory (WM). STM refers to human mind's capacity for storing a small amount of information and maintaining it in a temporary manner. LTM is a large database that has contained knowledge and previous events for a long period of time. WM involves short-term memory and a bundle of working mechanisms that make short-term memory work in planning and carrying out actions (Cowan, 2008).

Three types of memory are intrinsically intertwined with each other. The Atkinson & Shiffrin memory model (1968), also known as the multi-store model or three-stage model, is a widely accepted theoretical framework that presents the fundamental concept of how three types of memory interact with each other. See Figure 1 below for the illustration of this human memory cycle. Walters (2020) explained the linear model in three steps. First, when a human perceives visual (**iconic memory**) or auditory (**echoic memory**) sensory

input, both types of memory fade quickly; they only have a duration of between one millisecond and four seconds. Second, only the information that we pay attention to enters the second stage, STM but mostly it is retained for not more than a minute. WM serves as a series of processes, such as mental replaying of auditory information (**phonological loop**), mental mapping of visual image (**visuospatial sketchpad**), and mental narrative (**episodic buffer**) to retain, make sense of, store, interpret and modify the information in STM. The central executive is a hub to direct these three components and determine the importance of each component. Through the maintenance rehearsal of the three processes in WM, it can prevent STM from decay. Third, in the final stage, by using selected information repetitively for an extensive period, STM can be switched to LTM and LTM can be recalled back to STM when needed. Moreover, only important and meaningful memories are encoded from STM to LTM by the means of iterative stimulation (Chang et al., 2011).

To illustrate WM with a common scene in daily life, university students mingle with many new people at social events during freshers' week of a new academic year and usually people start self-introduction with their names. In order to remember new names, WM is in use to replay the name in the mind, create an inner voice to repeat the name and visualise a linkage between the person and the name. A number of components work together to keep the names in STM.

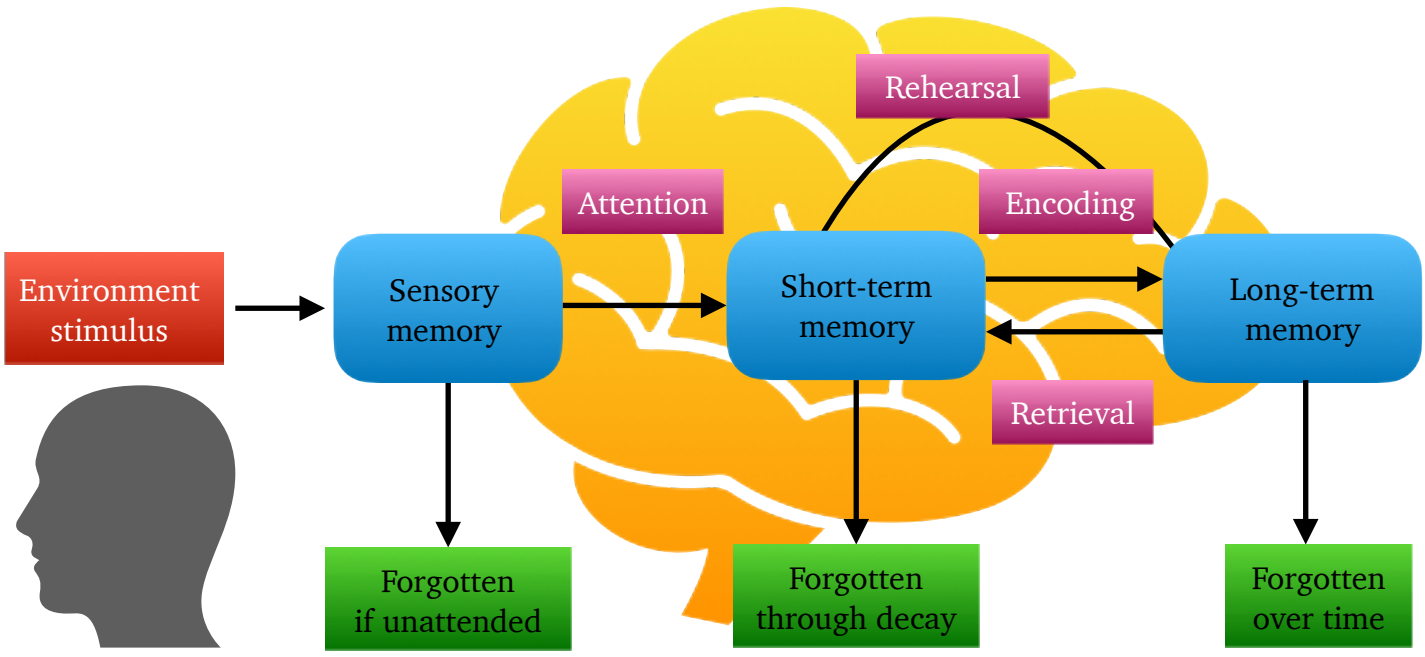



Figure 1 Illustration of human memory

## 2.2 Memory

### *2.2.1 Brain and memory*

The concept of memory has been explored in a wide array of aspects across numerous academic disciplines, including memory impairment of ageing population in neuroscience (Sachdev et al., 2010), collective memory in sociocultural psychology (Brown & Reavey, 2015), the relationship between implicit memory to advertising and consumer behaviour (Machouche, Gharbi, & Elfidha, 2017) and so on. This paper selectively focuses on the relevant literature in SLA and looks into the significance of memory in language learning.

First of all, when learning a language, what process does the brain go through to process a new word? What is the way to store information in LTM so that the word can be retrieved but not forgetten quickly? Before looking into language learning in particular, cognitive neuroscience provides an explanation of how memory is processed in brain areas in theory.

There is a dichotomy in LTM between declarative memory and non-declarative memory. Declarative memory (**explicit memory**) describes the recollection of facts and events in a conscious state of mind whereas non-declarative memory (**implicit memory**) is a collection of memory abilities in multiple brain regions that enables changes in behaviour unconsciously (Squire, 1992). Different types of information are stored at different brain areas. With regard to the corresponding brain areas, the hippocampus and surrounding entorhinal cortex and perirhinal cortex in Medial temporal lobe (MTL) are responsible for declarative memory so that the brain can recall acquired and learnt knowledge or spatial information. For non-declarative memory, the concerned brain domains are basal ganglia and cerebellum which help to form habitual behaviour, motor and perceptual skills, procedural memory for physical activities, resulting in reflexive and automatic responses (Gavin & Theibert, 2020).

The importance of the hippocampus area to memory can be illustrated by an epilepsy patient's case. According to Scoville & Milner (1957), a patient called H.M. had suffered from epileptic seizures for around 20 years since a bicycle accident at 9 years old. As a last resort to treat his epilepsy, this patient underwent a radical operation to have an extensive extent of MTL resected bilaterally. After the surgery, although the incidence and severity of

the seizures were ameliorated, the patient faced persistent memory deficit, known as amnesia. For instance, he could neither recall day-to-day events of his stay in the hospital nor the fact that his favourite uncle passed away three years ago. This clinical operation example suggested that the brain area hippocampal complex was vital to normal memory function.

*2.2.2 Brain, memory and language learning*

It is worth noting that memory, in essence, relies on the operation of a heterogeneous group of neural circuits which depend on the MTL critically. These neural circuits have a crucial position for the acquisition of new memories, recognition and recall of recent information, such as the ordinary facts and events (Reuber, 2008).

To relate brain and memory with learning, Clark (2017) addressed this in three parts. First, the neural pathways in memory have an impact on humans' behaviour and the changes in behaviour are usually observable. Second, learning can come into effect when environment stimuli are able to access the nervous system's pathway. Third, for learning to take place, there must be some changes or modifications in the nervous system as it acts as the neural substrate of learning as well as the resulting memory.

Following this idea, to learn a new word, the brain has to make some changes in the neural pathway for the semantic stimulus. Schütze (2016) elaborated on word processing in the brain with WM. Given that memory may decay shortly in STM, WM serves to lengthen the duration for the word to stay in the brain. For example, WM helps learners to rehearse the word subvocally and identify phonemes and lexemes in the meantime. In this case, the brain will be likely to create a record of the word and store it in the vicinity of hippocampus area. With the word record, the word can be recalled later.

Human brain can be conceptualised as a system of interconnected nodes linking different areal neural networks (Power et al., 2013). After creating a word record, the word will follow its neural pathway under the word's related association. When the brain receives and recognises external stimuli, the piece of information will be placed in an appropriate node. The nodes in memory are connected through links to indicate relations. For example,

the word ‘cat’ node is under ‘animal’ node as ‘subordinate’ and ‘cat’ node is seen as ‘coordinate’ relation with ‘tiger’ node and ‘leopard’ node. The characteristics of a node, the ‘cat’ in this case are described as ‘schemata’ and can be filled in ‘slots’ of the ‘cat’ node. Making associations between words allows stored items in LTM to be searched along the linking routes (Simon, 1981).

## 2.3 Attention

### *2.3.1 Attention and memory*

Prior to the word process in WM and retrieval in LTM, as illustrated in Figure 1, for acquisition of a new word, the word has to pass through an earlier stage of sensory memory on the condition that the human pays enough attention to the word so that the word can enter the LTM in the neural mechanism.

In cognitive neuroscience, Awh et al. (2006) summarised two main constructs concerning the limitations in information processing. First, maintenance of information in WM is temporary since the information is stored in a limited-capacity system that aims for efficient access and quick updating. It is believed that memory and attention are intimately intertwined. Likewise, based on Kahneman’s capacity model (1973), a human’s attention is assumed to be limited at any time, thereby limiting their capacity to perform mental work. Only the information that people pay attention is retained in STM. If people pay greater attention, the information will be transferred to LTM. Therefore, attention is immensely crucial for both STM and LTM (Russell & D’hollosy, 1992).

Second, given that both attention and WM are limited, the brain makes the best use of the resources and becomes selective in nature. To be exact, the selection of relevant information or stimuli is called ‘selective attention’ that delineates the preferential biases (Myers et al., 2017). Instead of dealing with an enormous amount of sensory information, the brain selectively pays attention to concerned targets by contrast, and emphasises the efficient encoding (Awh et al., 2006). In other words, the human’s brain prioritises relevant information over irrelevant information by giving selective attentional allocation.

Naghavi & Nyberg (2005) reviewed 12 studies that reported subjects' neural activation patterns in association with attention, WM, episodic retrieval and conscious perception through Functional magnetic resonance imaging (fMRI) and Positron emission tomography (PET). Anatomically, they showed a consistent and overlapping pattern of neuronal activities at the fronto-pariental area during the process of the aforementioned 4 specific cognitive demands. The neuroimaging evidence indicated that the brain is a 'global workspace' that enables different systems to interact through a single structure when they are activated. While selective attention gives access to conscious contents, this allows WM to process visual imagery, inner speech and conscious perception as well as the successful retrieval of information afterwards.

*2.3.2 Attention, memory and language learning*

If the amount of attention and memory are positively related, does it mean foreign language learners can memorise new words in an L2 easily as long as they pay enough attention? In fact, the literature suggests that this may not always be the case.

The study of Cop et al. (2015) aimed to find out if bilingual readers' eye movements when reading L2 sentences are different from L1 reading. 33 university students were asked to attend 4 reading sessions to finish an entire English novel. To make sure that they paid attention to the text, they had to complete content-related multiple-choice questions after each session. Texts were presented on a screen while participants' eye movements data were recorded. The findings revealed a large effect in sentence reading time. For within-subject measures, Dutch native speakers were 17% slower in reading L2 English than L1 Dutch ($F = 36.43$, $df = 24.70$, $p < 0.001$). The effect was larger on reading longer sentences reflected in the correlation analysis between language and number of words ($F = 9.92$, $df = 207.54$, $p < 0.005$), showing participants spending longer time on longer sentences due to more complex syntax. Apart from that, participants also had both a larger number and a longer duration of eyes fixations in reading L2 text than L1 supported by statistically significant data.

According to Randall (2007), compared with native speakers, the phenomenon of non-native speakers having a longer duration and larger number of eyes fixations at the written

text implies a paucity of storage capacity in WM. The reasons are threefold. First, for non-native speakers, their attention has to span across both content and function words but native speakers mostly skip functions words, provided their L1 knowledge of language structure is good. Second, non-native speakers need to process the text by using syntactic rules consciously. Third, non-native speakers do not have enough strategies to recognise words automatically. Therefore, these difficulties render L2 learners less able to memorise new words. Randall pointed out that efficient reading of textual information depends on the condition that the word has to stay long enough in WM for it to be processed and integrated with new information.

To counteract the disadvantages of non-native speakers in word recognition, this study aims to find a vocabulary learning strategy to prolong second language learners' attention or prioritise their selective attention so that new words can stay long enough to enter WM, attaining automatic word recognition in return. Colour is a selected main theme for this strategy.

## 2.4 Colour and memory

### *2.4.1 Colour and memory in non-language contexts*

Extensive research has been found in relation to colours and words. Stroop colour-word test (Stroop, 1992) has been a popular topic across decades (Weekes & Zaidel, 1996; Brown et al., 2002; Van der Elst et al., 2006; Hutchison, 2011; Shao et al., 2015; Kinoshita et al., 2018). Researchers tried to examine how subjects react when the name of a colour is incongruent with the ink-colour of its print (i.e. the word 'red' is printed in green) and subjects were asked to name the colour and determine the semantic meaning in the task. Besides, a great deal of research looks into the ability of individuals with clinical conditions such as semantic dementia, grapheme-colour synaesthesia, and down syndrome to memorise vocabulary or recognise objects with the use of colours (Laws, 2002; Radvansky et al., 2011; Ikea et al., 2016; Bibi et al., 2019; Smees et al., 2019).

Colours have been largely employed in neuroscience and psychology studies. The role of colour can be traced back to the importance of visual stimuli. Based on the aforementioned Atkinson & Shiffrin memory model (1968), the environmental stimuli are converted to

sensory input in accordance with the senses such as vision and hearing. From all senses, Medina (2011) argued that the sense of vision surpasses the rest. Seeing is a complicated process that consumes half of the brain resources. This is much more than all other senses. Vision also takes a dominant role to bring other senses into operation, such as hearing.

The superiority of vision over all other senses was discovered a long time ago. Colavita (1974) was the first to observe how visual stimuli dominate human behaviour. In 4 different experiments, when both visual and auditory stimuli were jointly presented, participants consistently responded to visual stimuli and were not even aware of the presence of auditory stimuli. The result was striking at that time. Since then, the Colavita visual dominance effect has been well studied by researchers and received similar robust results (Koppen & Spence, 2007; Sinnett et al., 2007; Li et al., 2017).

Among visual stimuli, colour, intensity and orientation are considered 'early' features that draw pre-attentive selection since they create a salient contrast with surrounding items. Importantly, these 'early' features have the ability to facilitate search in the brain more efficiently towards visual objects, and to disregard irrelevant information quickly (Divjak, 2019). To put it in an alternative way, colour has the power to win the competition for better attentional control and be the focus of attention.

The use of memory can be reflected in human's cognitive system, including pupil area and neural network. Eyes and the brain are closely linked to each other as a correlation has been found between pupillometry and neuromodulator systems which are responsible for regulating cortical arousal levels for perception, attention and memory (Kuipers & Phillips, 2022). Technically speaking, when memory is not in use, pupils are initially constricted. When the brain is in the course of encoding information, experiencing distraction and recalling memory, the pupils appear to dilate (Kucewicz et al., 2018). More specifically, Pajkossy et al. (2020) explained that presentation of environmental stimuli leads to pupil constriction in the first stage. When the stimuli enter WM processes and recalling phrases, the pupils get larger subsequently which shows mental effort is in use.

In respect of attention priority, Olmos-Solis et al. (2018) were curious about how humans' pupils reacted to relevant and irrelevant colour visual inputs from a neurophysiological perspective. Subjects sat in a room with dim indoor lighting and looked at a computer screen. First, a target colour was displayed and after a delay period, a number of different colours were shown at the end of each trial. Subjects were instructed to locate the specific target colour and press the corresponding arrow key on it. During the visual search, the target colour was seen as a template stored temporarily in the visual working memory so that the brain could differentiate stimuli and look for a matching target colour. Eye tracking apparatus was deployed to monitor subjects' reaction time, blinking times and pupil modulation. The data clearly showed that subjects' pupils became smaller when they first attended to the relevant colour as opposed to the irrelevant colour and then they gradually dilated. In this case, it is a sign that pupils are sensitive to task-relevant stimuli that match the colour representation in memory.

In comparison with other visual features, Kursawe & Zimmer (2015) found that colour can yield a better memory performance than shapes. They asked participants to detect changes in either colour, shape or both colour and shape in visual displays. The more difficult the mental task is, the more attentional effort it needs. The use of attentional effort results in an increase in pupil diameter as it is at work. The data analysis of pupil responses in three conditions suggested that memorising simple colours tends to use less attentional effort than complex shapes. That means colour costs less storage in visual WM. In contrast, memorising shapes is comparatively more complex and more prone to error. In view of limited capacity in memory, colour is an easy feature that can be stored in WM automatically without causing extra memory loads. Herlitz & Bäckman's study (1990) also delivered a consistent result that both patients with Alzheimer's disease and normal older adults remembered colours better than shapes.

Apart from pupillometry, neuronal activity also provides further evidence to support the contribution of colour to memory compared with other visual attributes, such as form, depth, and motion. The passage of visual information travels from retinas to visual cortex where it is dedicated to receive, integrate and process visual information (Huff et al., 2021). The visual cortex (V1 to V5) are responsible for different perceptual domains. For

example, V4 stores surface properties (Roe et al., 2012) and V5 is a pivotal motion area (Zeki, 1990). In relation to these cortical areas, Zeki & Marini (1998) revealed how colour is processed in the brain in three general stages. Colours first arrive in area V1 and V2 which are accountable for the presence and intensity of colour wavelengths as well as differentiating them. The second stage of colour processing goes to area V4 which helps to maintain constant colour perception even if the conditions of illumination differ. The third stage is inferior temporal and frontal cortex which are in control of colours of objects.

Regarding intricacy between colour and memory, neurons in visual cortical area V4 are concerned with perceptual, WM, and decision processes that are necessary for making memory-guided decisions (Hayden & Gallant, 2013). Importantly, area V4 is specialised for colour vision (Zeki et al., 1991; Heywoord et al., 1992; Gegenfurtner, 2003). That means it has more neuronal population dedicated to the processing of colour so colour is regarded as a higher-level visual characteristic. However, another visual characteristic, orientation of an object is more related to eye-centered coordinates. Besides the primary visual cortex, it seems that a particular visual area designated for processing orientation information is notably absent. It explained why attention for colour is more pronounced than orientation information in visual WM (Niklaus et al., 2017).

Colour is considered as an easy feature to be stored in humans' brains and colour has a specialised cortical area V4 for its functions so that colour enjoys a better memory performance than other visual attributes. In addition to these two advantages, the human brain is also adapted to coloured visuals instinctively. Tony Buzan is a leading and influential figure in memory. He expressed in an interview that taking notes in monochrome by using a black pen does not work well with memory. Instead, using colour and arrows to jot down notes helps to declutter the brain (D'Orio, 2010). Richmann (2006) also pointed out that the human brain is neither trained nor used to viewing things in black and white. Specifically, coloured images are 5-10% more beneficial to recognition memory than achromatic images (Wichmann et al., 2002).

Muhammad et al. (2016) measured participants' brain activities with the technology Near-infrared spectroscopy (NIRS) and observed the differences between monochrome and

chromatic visuals on attention and WM. 20 university students and 10 elderly people counted the number of circles displayed on a screen. The tasks were tested in 5 different background colours, including white, red, yellow, blue and green. They had to select a correct total number of circles on the tablet screen as a way to count their marks. Subjects wore a NIRS machine on their forehead so that their haemoglobin concentration levels were recorded during the experiment. Examination results showed that both young and old people performed better with a coloured background than a white background. To verify the findings, NIRS data of haemoglobin changes also indicated a significant difference between the white background colour and the coloured background for both groups.

The study (Muhammad et al., 2016) did demonstrate careful data handling as it used t-tests to check the interaction between colours. Also, NIRS data was able to detect brain activities around frontal regions during the tasks. However, it could have improved in a number of areas. The experiment aimed to look into two groups' attentional levels and WM when performing circle counting tasks but the design of the study could not objectively tell whether the subjects were paying full attention or whether WM was in operation. This study did not mention changes of attentional level in terms of neuronal activity in 5 different coloured backgrounds over the experiment duration. Yet, from a similar study, Shalev et al., (2017) presented relevant coloured stimuli and irrelevant distractor stimuli to test participants' selective attention. Researchers used 5 independent parameters to record visual attentional activities in the brain, including visual STM capacity, perceptual threshold of visual detection, speed of information processing, estimated attentional weightings over the course of time and space from neural substrates, and the top-down selectivity index. With a wide range of data, it is more precise to describe subjects' attentional levels. Besides, since WM has a limited capacity, usually studies would set a delay phase for participants to hold a small amount of visuals in memory. Participants were asked to memorise a cue that had to be searched for in a change detection task later (Moriguchi & Hiraki, 2013). Souza et al., (2016) proposed that focused attention was drawn to the cue so that this can avoid interference and enhance retrieval performance. This is called retro-cue benefit in WM test as it can see how a visual was temporarily held and recalled in WM. Yet, in Muhammad et al.'s study (2016), the circle counting task did not denote the temporary storage function of WM since the number of circles counted was

not for later use. A lack of a delayed cue and a recall task made the study ambiguous in interpreting subjects' performance in memory retention. On top of that, the participants were required to fill a self-reported questionnaire to indicate their levels of tiredness, readability, concentration and difficulty when viewing contents in 5 different screen background colours. The study simply used 2 tables to state that these were questionnaire results given by both young and elderly subjects. According to the self-reported data, young subjects had the best concentration on the white background. But it did not describe what kind of statistical data it was, for example, mean, mode, median, or other types of data. Therefore, it was not able to interpret the data in a meaningful way without this piece of information.

A few years later, Muhammad et al. (2018) conducted another similar study to examine the impact of coloured backgrounds on WM with a reading span task and compare the results with the prior study (Muhammad et al., 2016). In a similar setting, 30 Japanese university students and 7 elderly subjects were recruited to read 20 Japanese sentences in white, green and blue background colour on a tablet respectively. In order to test WM, subjects were asked to memorise and recall an underlined word in each sentence, amounting to 20 words in each set. NIRS probes were adopted to measure subjects' haemoglobin concentrations. Average percentage scores revealed that both groups achieved the best performance in green background colour, followed by blue colour and the worst in white colour. The recorded NIRS value indicated the haemoglobin level was the lowest in white background colour condition among young subjects.

There were a few areas that this study did not address clearly or could have designed better. Regarding data interpretation, first, although the NIRS data of the haemoglobin levels of elderly subjects were collected, the study did not have any disclosure of this at all. In this case, the study could not give a full picture to show the interaction between subjects' brain activity and the reading task in each background colour. Second, data showed that both young and elderly subjects performed the best when the green background was presented. However, no statistical differences were found between background colours and correct answers in ANOVA, meaning that a conclusion cannot be drawn from the data.

Therefore, with reference to statistically non-significant data, it is untenable to reach their claimed conclusion that white background is not advantageous to the reading task.

For the research design, first, all 7 elderly subjects were 65 years old or above. According to Centers for Medicare and Medicaid Services (2012), individuals who are 65 or above are categorised as older adults and they are inclined to experience visual changes that can affect reading adversely. These changes emerge in the way they perceive colours, contrast and sharpness of details which makes the older people hard to read written material. When conducting research studies with older adults with age-related vision loss, Trujillo Tanner et al. (2018) recommended some practices to mitigate the barriers, including magnification of font size, optimisation of lighting and seating arrangement. The study of Muhammad et al. (2018) did not describe whether the elderly participants had normal vision and whether reading aids were made for them to ensure they read the provided materials smoothly. It is equivocal to see if subjects' reading performance was affected by background colour or visual impairment. Next, the study did not address how it measured the correct rate of the reading span task. It just mentioned that subjects had to memorise 20 underlined words but the questions are: what were the recalling procedures? Were there any cues or delay phases? How to check if the subjects memorised the words correctly? It would be more clear if the study addressed the above questions in the research design section.

Moreover, as the study saw that a tablet computer's screen colour might possibly affect users' attention, the researchers could have added black background colour as one of the testing conditions since dark mode is also available in many smart devices' display settings. In this sense, the findings could have been more integral to see the impact of monochrome, but not just white background colour. Last but not least, their previous study (Muhammad et al., 2016) demonstrated the highest NIRS value and brain activity with a white background during the circle counting task. The researchers claimed it was a result of stress from looking at the white background. However, in this study (Muhammad et al., 2018), it showed a remarkably contradictory result that young subjects had the lowest NIRS value on average when they performed the reading span task with a white background. The study should have given an account of the NIRS value differences.

*2.4.2 Colour and memory in language contexts*

This section will focus on colour and language studies that are not within SLA vocabulary learning scope. The studies about colour in SLA vocabulary learning will be covered in section 2.5.3. In the light of WM's limitations, colour is a feature that can facilitate memory provided that it can lessen the attentional load with its simple characteristics, take advantage of its designated V4 cortical area to gather more neurons for attention, and generate better memory performance with its natural compatibility to the brain's operation. Considering these innate advantages, it is intriguing to see if these advantages can be carried on in language learning and become a memory aid in SLA.

The following section will dissect colour's different functions in SLA with three up-to-date research papers: i) effectiveness of yellow paper on English collocations learning in a group of South Asian university students in China, ii) effectiveness of using colours to develop Ukrainian grammar learning materials on foreign university students, and iii) effectiveness of colourful semantics on learning wh- questions among preschool children in Sri Lanka.

In the first study, Khan & Liu (2020) found that the performance of subjects using colour as an environmental stimulus in studying English collocations far exceeds that of the control group. 30 male university students in China were selected in the study. They were originally from Afghanistan, Pakistan and India and their L1s were Pashto, Urdu and Dari. Half of them were assigned to the control group and half to the experimental group. Both groups completed a pre-test in order to see who were qualified to proceed to the post-test. The pre-test contained 30 collocation multiple-choice questions on white papers. After the pre-test, they had 10 to 15-min lessons on these collocation meanings for three days. Two weeks later, the experimental group did the same collocation test as a post-test on yellow paper whereas the controlled group did it on white paper. Out of 30 questions, the experimental group and the control group scored 25.6 and 23.73 on average respectively. The difference was significant ($p<0.04$) which provided support to the claim that yellow coloured paper can yield a better WM performance than plain white paper in learning English collocations.

This research used randomisation in participant allocation to reduce bias and avoid imbalanced proficiency levels in both groups. Randomised controlled trial also serves as a

way to generate high-quality evidence for assessing the effectiveness and safety of an intervention (Bhide et al., 2018). Besides, the data also further consolidated that both groups' English proficiency levels were quite similar as there was no significant difference between groups in the pre-test. Moreover, it is appropriate for Khan & Liu to choose t-test as a statistical formula given two sample groups.

Nevertheless, some areas in the research design seemed to be problematic and some points did need taking into consideration in the data report. First of all, in the pre-test, the mean scores of experimental group and controlled group were 24.06 and 23.73 respectively. In other words, with a perfect score of 30 marks, both groups attained a rate of nearly 80% correct answers. It is reasonable to consider it reaching a ceiling effect and imply that the pre-test was not challenging enough. Concerning the problem of ceiling effects, Austin & Brunner (2003) described that it can pose a dangerous and unexpected statistical problem when one of the independent variables in a multiple regression displays a ceiling effect. This may even cause a higher rate of Type I Error (false significance) which will jeopardise the interpretability. Second, Khan & Liu could have disclosed more key information regarding participants' English proficiency in order to judge whether the English collocation tests were too easy. For example, participants had to choose the correct meaning of the collocation 'heavy rain' (question 3), 'keep quiet' (question 18) and so on. All participants were Masters degree and PhD university students with some taking English Literature as major, and all received English as a medium of instruction schooling previously. Hence, the level of the selected English collocations were too basic and had the potential to lead to a ceiling effect. Third, one of the major problems in this research is that both groups used white paper in the pre-test and also in the intervention part. Experimental group only used yellow paper for the post-test after 2 weeks. How could the researchers tell if the colour had an impact on WM as yellow colour was totally withdrawn during the lesson intervention? It also could not test the effectiveness of colour on memory retention from the experimental group as claimed since the subjects used plain white paper in the pre-test.

Moreover, it is advisable for Khan & Liu (2020) to take both gender effect and cultural effect into consideration for participant recruitment. All 30 participants in Khan & Liu's study were males and this may result in gender bias. For gender effect on perception of

colour, Venkatesh et al., (2002) discovered that females' visual reaction time over colour was longer than that of males. The authors ascribed this to the female sex hormones, affecting the nerve impulse velocity. A recent study conducted by Jakovljević et al., (2021) also suggested that colours may have different effects on the reading process between males and females. They provided physiological evidence that females experienced significantly higher Heart Rate Variability than males when reading texts particularly with yellow, red and orange backgrounds, contributing to a different reading experience between genders. Therefore, without female participants, Khan & Liu's findings (2020) on yellow paper towards English collocation learning might be potentially affected by the gender effect. Furthermore, the participants in Khan & Liu's study were from 3 different cultural backgrounds, including Afghanistan, Pakistan and India and they all spoke different L1s. People from different cultural backgrounds have different perceptions, emotions and preferences towards colours (Ozgen, 2004; Ou et al., 2012). Professor Anna Franklin illustrated this with an example. Russian has different words to describe the colour blue so that Russians are more sensitive to colours in this spectrum (Finnegan, 2016). In view of this, the participants in Khan & Liu's study (2020) had various cultural backgrounds and colour repertoire in L1 which may be a concern to influence their sensitivity to yellow paper in the post-test.

In the second study, Zavaruieva et al., (2022) employed innovative colour coding methods to help foreigners to learn Ukrainian grammar in a systematic and effective way. The research aimed to harness colours' influence in humans' mental process and create associative connections between colours and English forms in the hope of evoking relevant information in memory. They surveyed 100 foreign students in a Ukrainian university about their emotions towards colours and their perceptions were collected for visual learning materials design. First, researchers selected a variety of colour palettes and colour combinations so that the colours had a visual balance. Second, the students were assigned with a card set with the selected colours and they had to rely on their first instinct to pick an attractive colour which would be regarded as a strong stimulus. Third, the students received a form with 7 grammatical categories, including nominative, genitive, dative and so on. They needed to match each category with a colour code immediately. Interestingly, more than 70% informants linked nominative with red, pink and yellow; genitive with blue,

pink and green; dative with green, blue and purple. After collecting this information, researchers generated a Ukrainian grammar table that used certain colours to represent assimilated parts of speech for easy memory.

Zavaruieva et al.'s study (2022) possessed multifaceted merits to make grammar learning more accessible. First, they capitalised on students' instinctive perception of colours to link them with certain grammar forms. They considered colours as a salient medium to transmit information and cause emotions. Lee (2006) mentioned that emotionally charged stimuli are able to capture attention and lead to the information processing stage in memory. In this case, the colour stimuli can play a part in the cognitive process. Liu et al., (2022) emphasised that emotional and cognitive engagement in conjunction can predict learning achievement. Therefore, by making linkages, colour serves as a powerful stimulus to assimilate symbolic meaning of English forms, creating cognitive automation. Second, the grammar table was carefully compiled containing various grammatical information. Using tables has been suggested to be an effective way to teach a foreign language since students can refer to a particular grammar rule from the table (Ilhomovna et al., 2020). According to Novawan (2019), a teacher can make use of visual aids (e.g. pictures, texts, tables, diagrams, charts) to raise students' consciousness towards isolated and specific grammar items. Referring to this purpose, the Ukrainian grammar table in Zavaruieva et al.'s study (2022) is an appropriate tool to learn phonetic rules, masculine and feminine nouns, categories of numbers and so on. It is arranged in an organised way that is comprehensible by different language users and suitable for early beginners. In this way, Ukrainian learners can independently refer to the table for revising particular grammatical categories in a systematic and logical manner.

Despite its considerations of colours' psychological effect and the various grammar forms, Zavaruieva et al.'s study (2022) has some demerits that render the use of the colour grammar table preliminary. First, the methodology part did not include enough details for replication. It did not provide any demographic information for the participants, such as age, sex, and length of time studying Ukranian language. The method of participant selection was also not clear since the inclusive and exclusive criteria remained unknown. Although it briefly stated the steps of how researchers ran the colour matching test with the

participants, it did not describe the details of the matching process. For example, how many colours could the participant select to match with each grammar category, one or more than one? After collecting participants' preferences, the research neither described the data analysis process nor reported the percentage of each colour in each grammar case in full. This led to doubts forming as to why 70% was set as a threshold to be considered as a general and common colour selected by the participants. Second, the table has not actually been tested whether it helps learning. The study did not collect the necessary data. (i.e. they did not test the effectiveness of the grammar table with a new group of learners). No quantitative statistics were available at all. Without descriptive data, it is not feasible to give an objective view of whether and how these 100 foreign students or a new group of learners improved their Ukrainian grammar after using the grammar table.

In the third study, Hettiarachchi et al. (2018) investigated the effectiveness of colourful semantics in an English programme in Sri Lanka. They divided 21 typically developing preschool children into 2 groups. For the experimental group, researchers used colour-coding cue cards to conduct English learning activities throughout 12 English lessons, including introductory games, story reading, and sentence building. Each card had a specific colour and sign for a wh- question, for instance, orange colour for who, yellow colour for what he/ she did, green for what object. On the other hand, the control group received regular English classes from their class teacher as normal which contained class time story telling activities. This research study employed Peter and the Cat-Narrative Assessment, a speech and language therapy assessment tool, for both pre- and post- test to measure the participants' expressive language skills in terms of sentence length and syntax. Students' qualitative responses were also collected to compare their answers to wh-questions before and after the intervention. Statistically significant results suggested that compared with the control group, a colourful semantic approach had an overwhelmingly positive influence on the experimental group at both group and individual levels as students could produce accurate and meaningful responses to wh- questions. Their answers were impressively longer with correct uses of conjunction and past tense markers.

Colourful semantics was firstly introduced by Bryan (1997), a Speech and Language Therapist. A 5-year-old client of Bryan named Gordon was diagnosed with language

disordered conditions as he could hardly produce a word unless he was provided with enough visual prompts and familiar social phases. In view of this, Bryan attempted to use colour-coding to develop a thematic role therapy so that Gordan could establish a mental connection between wh- question types and visual colour sequences. During an 8-week therapy programme, Gordon witnessed continuous lexical improvement at functional level. Although colourful semantics was originally designed for children with special education needs, Hettiarachchi et al.'s study (2018) was one of a few that looked into the effect on typically developing children. The robustness of the study is reflected in the study design. First, the research question was clearly identified by Hettiarachchi et al. and the selected measurement and data handling were suitable to address the question. The research question was 'does colourful semantics help children in the preschool stage to respond to wh- questions?'. The researchers listed out evaluation criteria to be qualified as an accurate response: semantic relevance, correctness, and length of utterance. The paired t-tests displayed the encouraging outcome that the experimental group outperformed the control group at both group and individual levels. Statistically significant results showed that participants in the experimental group improved in post-tests regarding accuracy of response and length of response. Besides, among 3 participants who demonstrated complexity of response, 2 of them belonged to the experimental group. Findings from previous colourful semantics studies were in line with their result: children with language impairments showing development in the use of verbs and arguments (Bolderson et al., 2011), intellectual disabled children exhibiting improvement in narrative skills and language structures (Hettiarachchi, 2016), deaf or hard-of-hearing children displaying understanding of wh- questions and giving appropriate responses (Hettiarachchi & Ranaweera, 2019). In aggregate, colourful semantics are a favourable learning tool in SLA among children.

There are some considerations that Hettiarachchi et al., (2018) could have addressed more. First, in order to specifically tell whether colour was important, the control group could have done the same as the experimental group in terms of everything except that the wh-questions cue cards were not coloured. The study mentioned that students in the control group had story telling activities as usual but it did not describe whether non-coloured wh-questions cue cards were used so it was hard to make a direct comparison with the

experimental group. The second issue is about researcher bias. The colourful semantics intervention in the experimental group was led by the first two authors. A research bias may exist when the researcher is involved in research design, sampling, field work, data analysis, or report writing (Tavakali, 2012). Since Hettiarachchi et al. were clear about the whole research design, it would be better if they could have given an account of how they avoided research bias.

In addition, from a macro view, with reference to the authors' research studies (Hettiarachchi, 2016; Hettiarachchi et al., 2018; Hettiarachchi & Ranaweera, 2019), all the concerned participants were either typically developing children or atypically developing children aged around 5 years old in Sri Lanka. This may potentially fall to another generalisation problem. In Applied Linguistics, if observations are drawn from a narrow scope of population distribution with particular individual backgrounds, the results are possibly compromised as a result of restricted range (Andringa & Godfroid, 2020). On a side note, although the effect of the colourful semantics on children of a particular age were largely studied, a big research gap seems to exist because colourful semantics have not been looked at much in adults or children of different ages. Moreover, the study of Hettiarachchi et al., (2018) was piloted with a preschool child on wh- questions in advance. However, they did not reveal details of the pilot test, such as the purpose of the pilot test, the performance in the pilot test, or what changes were made to the main study after the pilot test to make the experiment feasible. Lastly, a minor glitch was spotted as they claimed to have one class of 10 children in the experimental group but the entire result part presented data from 11 participants which caused confusion.

The above three scientifically-backed research studies explored the possibilities of colours in different scopes in SLA, ranging from English collocations, grammar, and wh- questions. Evidence suggests that colour is an effective tool that brings L2 learners remarkable learning achievements.

## 2.5 SLA Vocabulary learning

### *2.5.1 Importance of teaching and learning of vocabulary in SLA*

Susanto (2017) underscored the importance of vocabulary in language teaching since it is not possible for students to learn a language without words. To prevent students from forgetting vocabulary, language teachers have to teach students which vocabulary needs to be learnt, practised and revised. Similarly, vocabulary is considered important to language learners as well. Learners can use vocabulary as a tool to obtain communicative competence, as they need to express ideas and understand what other people say in words.

In the field of SLA, vocabulary is a determining factor for 4 core language skills: reading, writing, listening and speaking. For reading, vocabulary learning can generate a synergistic effect as a large vocabulary can prevent the likelihood of guessing word meanings as well as misinterpreting words in the wrong context (Coady & Huckin, 1997). For writing, writers communicate ideas in words and they choose the right words to convey the intended meaning (Johnson, 2000). For listening, aural vocabulary knowledge helps learners to comprehend the spoken language of others, so an extensive vocabulary is a success predictor for listening comprehension (Matthews, 2018). For speaking, a satisfactory command of vocabulary is inseparable from speech production, enabling speakers to create appropriate meanings as well as accurate syntactic, morphological, and phonological structures (Koizumi, 2012).

Given the important role of vocabulary learning in SLA, an important question is: how do learners assess how well they know a word? A learner's knowledge of vocabulary can be evaluated by two components: vocabulary breadth and vocabulary depth (Li & Kirby, 2015). Vocabulary breadth refers to the number of words that a learner knows. Vocabulary depth refers to a learner's understanding of various aspects of an individual word (Qian, 1999) Moghadam & Ghaderpour (2012) stressed that vocabulary depth and vocabulary breadth are equally important for a good understanding of words. While both dimensions are important, the current study does not aim to achieve thorough vocabulary knowledge through the intervention.

*2.5.2 Techniques for aiding memory of SLA vocabulary*

When it comes to memorising a word, Heyselaar et al. (2017) referred to an understanding of a word's sound, syntax and its meaning. In general, one has enough linguistic knowledge to distinguish these three important components. Over the course of language acquisition, sound, syntax and meaning are deemed as fundamental building blocks in language to be encoded and consolidated in memory. This is how memory works in the human language system. In psycholinguistics, there is a specific term called 'mental lexicon' for the words in memory.

A meta-analysis of the relation between reading and WM reviewing 197 studies concluded that when performing a vocabulary task, WM is activated to retrieve the existing background knowledge and concepts and combine them with the sources of visual information, such as the vocabulary words. Advanced learners can achieve automation in background knowledge retrieval from LTM and apply it to vocabulary tasks. In contrast, beginning learners may take more time and WM to integrate their background knowledge with the vocabulary since they are not familiar with phonological and orthographic representations (Peng et al., 2018).

Concerning the weaknesses of beginning learners in SLA vocabulary learning, Ibarra Santacruza & Martínez Ortegaba (2018) did a study into memory training strategies for the purpose of expanding learners' WM capacity and thus attaining success in vocabulary acquisition. 50 university students in Columbia attended an English course as a part of their graduation requirement. During 10 weeks, the control group did not use any strategies. On the other hand, each week the experimental group implemented one strategy from 10 WM boosting strategies which were listed in a WM classroom guide written by Gathercole & Alloway (2007). To name a few: creating images or drawings to illustrate a word and write sentences; using concept maps to make clusters between words and link to other expressions; connecting students' background knowledge with new vocabulary knowledge with the use of a graphic organiser; reading out the words aloud and so on. Compared with the control group, the experimental group presented better learning outcomes. The significant difference between pre- and post-tests revealed that the WM strategies were effective in vocabulary retention. With reference to the encouraging results, the researchers

also added that students in the experimental condition were not only capable of retrieving background information, but they were also able to use, recognise and transfer the acquired vocabulary to other listening, reading, writing and speaking tasks.

At the other end of the scale, Rambe's study (2018) targeted 17 gifted students in Indonesia who had high proficiency in L2 English. By using a questionnaire and an interview, this study surveyed the students' frequency of use of the vocabulary memory strategies recommended by Schmitt (1997). The top 4 strategies used by the participants were to memorise the spelling of a word, to read aloud the words in revision, to relate the word with a mnemonic, such as a visual image, and to make a word cluster. Interestingly, the last 3 strategies were also used by the participants in Santacruza & Martínez Ortegaba's study (2018). Regardless of level, language learners seem to find visual stimuli and making associations between words useful for vocabulary learning.

Vocabulary strategies are also crucial in training students to achieve independent learning. Rather than using rote learning, Liu (2011) highlighted that learners with higher English proficiency adopt more vocabulary strategies to focus on pragmatic use than low achievers and they use more self-learning and show higher awareness. Bernot & Metzler (2014) held a view that the prime goal of taking away teacher-led lessons is to train students to become independent learners. To achieve this, learners need to be equipped with various vocabulary learning strategies. With respect to limited mental resources for process ing information, English learners' vocabulary storage, processing and use are through repeated exposure to its verbal, visual and spatial forms. Teachers are advised to be familiar with different strategies for addressing learners' needs to store, retrieve and use vocabulary. When learners have plenty of opportunities to use different vocabulary strategies, they are empowered to take control of these in their self-learning (Geronimo, 2021).

*2.5.3 Colour and SLA vocabulary learning*

In recent years, researchers' attention has shifted to using colours for educational purposes. This section will look into 3 recent colour studies to see how they utilised colours in 3 different ways, assisting learners with vocabulary acquisition, including seeking an intuitive linkage between a lexical item and a colour for better memory, adjusting the font size and

font colour to memorise Chinese vocabulary, and setting different coloured words and coloured backgrounds for vocabulary retention.

In the first study, Randolph & Dubjelová (2019) tried to turn SLA vocabulary learning into a sensory experience. They believed that learners have a natural tendency to develop a network of connection with the vocabulary. This study took place during an English course at a university. The researchers guided L2 learners to create associations between vocabulary and colours as well as other senses. The more vivid the imagination, the better the memory. For example, a student shared how to memorise the phrasal verb 'come up with'. The student's instinct related this lexical item to different senses and emotions, such as blue (sight), ideas in a new book (smell), imagined conversation with the student's mother to tell her what she came up with during the daytime (hearing), excitement (emotion). Another example is the word 'bicycle' where the learner created a visual image in his mind. The scene is full of sensory imagination, for instance, red (sight), the smell of falling autumn leaves (smell), the voice of the learner's father who taught him how to ride a bicycle (hearing). The students reflected that this method was useful in memorising vocabulary because all of the associations were elicited from their unique personal experiences and they also incorporated their feelings towards colours and their own emotions into each word.

In the second study, Fu & Peng (2020) examined the effect of font size and font colour on memorising Chinese vocabulary. 80 of the most frequently used Chinese nouns were selected. There were 5 rounds of experiments with 16 words in each round. The 16 Chinese words were aligned in a 4 X 4 grid format. The combination of font colour and font size in each set was fixed at 4 red words at 16 point, 4 black words at 20 point, 4 black words at 28 point and 4 black words at 40 point. 10 adult participants had to sit in front of a computer screen and memorise as many words as they could in 1 minute. After each round of experiment, they were asked to solve a set of arithmetic questions so that researchers could observe the participants' remaining word memory after the time interval. Following this mathematical task, the participants had to recall as much vocabulary as they could. Researchers discovered that the font size had a more notable effect than the font colour in

memorising Chinese vocabulary. The participants expressed that their memory was improved after relaxation or engaging in conversation between the experiments.

However, a number of critical concerns made interpretation of the results problematic. Firstly, there was no control group in the study. As such, it was impossible to compare the results with a group of Chinese learners who participated in the study without the intervention of font size and font colour. Secondly, the number of red words and black words was not equal. There were only 4 red words whereas there were 12 black words so black words might have appeared more conspicuously on the screen. Also, there was no counterpart for comparison in each combination. That means in each experiment, only 4 red words at 16 point were available but no black words in equivalent size could be used for comparison. Thus, it was not possible to gauge whether the effect came from either the font size or from the font colour. Thirdly, the study did not specify the data analysis process. It did not mention what statistical tests were run or whether the results carried statistical significance or not. As a result, a conclusion was reached without enough supporting evidence. In addition, the researchers did not control the duration of the interval task, and the participants were given an unlimited period of time in which to complete the task. It is suspected that memory retention of the Chinese vocabulary might vary between participants, as the time spent by the participants on the arithmetic task may differ. Some participants might retain memory of more vocabulary regardless of the size and colour if they spent less time on the interval task and vice versa. The selection of words can also be seen as a problem. The researchers selected 80 frequently used words, but this does not mean these words have the same level of difficulty. Therefore, the participants might be inclined to memorise words that were easier. There were some minor factors which required more clarification as well. The study did not mention the participants' Chinese proficiency level, which would influence their knowledge of the words shown on the screen. Also, the study did not state the distance between the participants and the computer screen - if the distance was great, the red words at 16 point would be barely visible. If that was the case, the study could not determine whether the effect of the font size actually overrode that of the font colour.

In the third study, Folgieri et al. (2013) used a Brain-computer Interface (BCI) device to detect participants' brain activation in response to words displayed in different colours and backgrounds. The researchers also investigated the impacts of colour on word memory. 38 undergraduate students (19 male and 19 female) wore the BCI device at the frontal region of the scalp to measure brain activity. They partook in 4 experimental sessions which had a different combination of word colours and backgrounds: i) black words on a coloured background; ii) coloured words on a white background; iii) words in red or green or blue colour on a red or green or blue coloured background; and iv) words in cyan or magenta or yellow colour on a cyan or magenta or yellow coloured background. Each session had 6 different obsolete or non-existent Italian words to ensure that the participants' attention was related to colours. Participants were not required to complete extra tasks except for learning these 6 words on the screen which was placed 50 cm away from them. The study recorded participants' Electroencephalogram (EEG) signal, the words they remembered in each experiment and participants' answers to a cognitive questionnaire about the impacts of colours and the participants' attention levels. The behavioural data analysis suggested that participants who paid high attention could best memorise words in the following situations: black words on a cyan background, blue words on a white background, red words on a blue background, blue words on a red background, and cyan words on a magenta background. Language educators may take these colour combination findings as a reference for devising vocabulary learning materials.

*2.5.4 The 4-colour memory system*

Introduced by Takashi Ishii, '4-colour memory' (Takashi, 2013) is a vocabulary strategy that highlights vocabulary with respective colour to indicate learners' knowledge of the vocabulary. For instance, red for no knowledge, yellow for shallow impression but not familiar with the vocabulary, green for basic knowledge of the vocabulary, and blue for thorough understanding of the vocabulary. It also decides the study priority. That means to revise red vocabulary first, and then yellow, followed by green, and blue comes last. After a period of time, when learners revise the words they have learnt, they can switch to a new colour if their word knowledge level varies. The main objective of this vocabulary strategy is to turn all vocabulary into blue colour which means that learners have a full understanding of the words.

‘4-colour memory’ can be a student-led self-learning method and easy to carry out since learners can use highlighters to highlight vocabulary on a physical textbook or choose the highlighter function in a digital environment. Learners can study vocabulary whenever and wherever convenient to them. Also, as discussed above, WM has a limited capacity and attentional resources are allocated according to prioritisation. In view of these, ‘4-colour memory’ can accommodate both characteristics as colour is an easy feature to memorise and learners can follow the sequence of colours to allocate corresponding attentional resources. Hence, ‘4-colour memory’ seems to be a feasible strategy to acquire vocabulary effectively.

### 2.6 Research gap and current study

The literature review has shown that colour can play a role in attention which is important for WM and thus learning. Some direct evidence suggests that colour can broadly be helpful in language learning. An important question is whether it might be useful for SLA vocabulary learning specifically, since vocabulary is such a key area. Although the studies that have been reviewed have looked at this, they have had methodological problems. For example, there is no study where they have a pre- and post test design, or where they have a control group. In some studies, the colour has not been tested in the intervention. Hence, there is still a dearth of research on vocabulary acquisition strategies with the colour element, addressing these research design problems.

Given this background, this study conducted new intervention study to explore the benefits of colour in SLA. ‘4-colour memory’ is currently used by teachers and students, but this method has not yet been tested experimentally. This study is going to test the effectiveness of this vocabulary learning strategy. The current study compared the performance of 3 different conditions. In the intervention, one group used 4 different colours to highlight vocabulary following the method of ‘4-colour memory’; another group only used one colour; the control group did not use any colour at all. This would allow researchers to see both if using colour is helpful at all and whether the ‘4-colour memory’ coding system is specifically useful.

### 2.7 Research questions

The research questions for this study are therefore:

1. Can the use of colour aid memory for Chinese speakers' English L2 vocabulary acquisition?
2. If so, which method is the most effective for L2 vocabulary acquisition by comparing using no colour, single colour and multiple colours?

It is hypothesised that the use of colour can aid memory for L2 vocabulary acquisition and using multiple colours is the most effective way among the three conditions.

# 3 Methodology

This chapter addresses the overall research design, participants' background and recruitment, data collection instruments, the research flow, and ethical considerations.

## 3.1 Design

This quantitative research centres around whether the use of colour can aid memory for Chinese speakers acquiring English vocabulary and takes the way students do their self-learning into consideration. The design of the research tries to imitate the students' self-study settings. Therefore, participants were asked to complete an online reading comprehension task and highlight the vocabulary in one colour or in multiple colours. Participants were randomly assigned to one of three groups: a control group (no colour) along with two experimental groups (single-colour, multiple-colour). This study used a pre-test and a post-test design whereby participants were tested on the vocabulary which was included in the intervention. The pre-test and post-test they did before and after the intervention were the same. This research is interested to see if participants improve on their scores in the post-test compared to the pre-test by measuring score differences. The extent of score differences for the three groups was also compared.

## 3.2 Participants

Hong Kong's secondary schools are divided into three bandings according to academic excellence with high-achieving students in band 1 schools and low-achieving students in band 3 schools (Tsang & Isaacs, 2022). Participants were recruited from a band 1 coed secondary school in Hong Kong which uses English as a medium of instruction. Teachers of the school expressed that the school has been interested in looking for innovative pedagogy to teach English vocabulary, thus setting up a working group. They found that this research study would be on a par with their teaching goals. In view of this, the school agreed to take part in and assist the administration of this research study.

According to the teachers, the average age of the participating students is 15 years old. Their first language is Cantonese, the principal language in Hong Kong. They received 5 English lessons per week and each lasted for 50 minutes.

*3.2.1 Inclusion criteria*

Participants were expected to have intermediate English proficiency as they were required to read a piece of an academic English article. According to the secondary school's English teachers, the school's Form 1 to Form 2 students are at pre-intermediate level and have relatively little exposure to academic English reading. On the other hand, although Form 5 and Form 6 students have achieved a more advanced level, they have had to prepare for public examinations. Therefore, Form 4 students (equivalent of Year 10 in the United Kingdom education system) were more appropriate for the intermediate English proficiency criteria so Form 4 was selected to participate in the study.

*3.2.2 Exclusion criteria*

Participants who have colour blindness have to be excluded as the research requires students to identify the colours within the selected vocabulary in the reading article. Also, participants who are not able to use a computer or perform online tasks were excluded as the research was conducted entirely online. Based on the above two exclusion criteria, no students were excluded.

**3.3 Participant recruitment**

Teachers helped to distribute an information sheet about the experiment to guardians and parents of all students in Form 4 which comprised 124 students in total, with approximately 30 students in each class. The school developed an online teacher-student portal so students' parents and guardians could indicate their preference electronically. Students who participated in the study received a certificate of participation at the end of the study.

After a week of recruitment, 7 students opted out of the experiment. The remaining participating students in each class were randomly divided into three groups, aiming for a similar number of participants from each class in each group. 99 students completed all tasks in the experiment, 33 in no-colour group, 34 in single-colour group, and 32 in multiple-colour group. Table 1 shows the flow of participant recruitment and the distribution of students in the three groups.

| Date | Flow of participant recruitment | Distribution of students | Number of students |
|---|---|---|---|
| From 10th May 2022 to 15th May 2022 | Students who received the information sheet | | 124 |
| From 16th May 2022 to 18th May 2022 | Students who did not opt-out | | 117 |
| From 19th May 2022 to 29th May 2022 | Students who completed a pre-test | | 109 |
| From 19th May 2022 to 29th May 2022 | Students who completed a post-test | | 99 |
| | | Students who were in no-colour group | 33 |
| | | Students who were in single-colour group | 34 |
| | | Students who were in multiple-colour group | 32 |

Table 1 Distribution of students at each stage of the study

The sample size was largely determined by practicalities and the number of participants who were available to do the tests at the time. In order to see the level of effect size for the key interaction, a power analysis was conducted, looking for the power for an interaction in an ANOVA between one within variable (test-session) and one between variable (group). As a result, in the one-way ANOVA study with a sample of 99 subjects divided into 3 groups, it attained a power of 80%. The assumption behind this power is a significance level of 0.05 in a non-central F test. The group subject counts are 32, 33, 34. The effect size is equal to 0.32 which is interpreted as a medium effect.

### 3.4 Data collection instruments

*3.4.1 Reading article*

When selecting a written article to be used as the reading in the intervention, the researcher aimed for an article of a level of difficulty suitable for Form 4 students. The article should be within participants' everyday context so that participants would find it familiar and interesting and be engaged with the material. After considering the aforementioned factors, a piece of reading material about Korean pop music was selected as this is within the knowledge and experience of the participants' age group.

*3.4.2 Pilot test*

Before the actual study, a pilot test was carried out to test the difficulty of the pre- and post-test to ensure it was not too easy which could lead to ceiling effects at pre-tests and to gauge the approximate time needed in the actual test. For the pilot test, 10 participants with similar English proficiency levels to the actual participants were recruited by creating a post on the researcher's personal social media in a private privacy setting or inviting personal contacts. Secondary school students in the actual study were excluded for the pilot test since the research team wished to maximise the number available to do the full experiment.

The pilot test was created with the use of Microsoft Form. The researcher selected 20 vocabulary words from the reading article, including both easy and advanced words and then used each word to form a statement. The 10 participants had to select an option with the corresponding meaning from 20 multiple-choice questions. On average, the participants answered 15.6 out of 20 questions correctly which was close to an 80% correct answer rate. As there were such a high percentage of correct answers in the pilot test, it was considered too easy.

With this in mind, 4 changes were made to avoid ceiling effects. First, 2 elementary words (band, young) were taken away as all participants got these right. Second, the number of questions was increased from 20 to 30 so more advanced words could be tested. Third, some options of the given answers in the multiple-choice questions were replaced by difficult words to prevent guessing behaviour. Fourth, some questions were decontextualised so that participants might not guess the word meaning. For example, for the word 'scandal', the original statement was 'some magazines contain nothing but scandal' and it was changed to 'it has become a public scandal'. Thus, participants might not relate the word 'scandal' to magazines.

With regard to the time spent on the pilot test, all participants finished 20 questions in around 5 minutes. This was shorter than the planned time for this part, 10 minutes. Adding 10 more questions to the test was believed to be the right move to utilise the experiment time more.

*3.4.3 Pre-test*

The pre-test followed the format of Vocabulary Size Test (VST) (Beglar, 2010; Chiu, 2010), testing participants' vocabulary knowledge in a multiple-choice question setting with answers shown in participants' L1, Traditional Chinese. 30 words were selected from the reading article in the intervention and were used to create a decontextualised statement.

By taking '4-colour memory' into consideration, the multiple-colour group had a chance to use 4 different colours to indicate the level of word familiarity in the intervention. In order to test the effectiveness of this method, Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001) has been taken as a reference in word selection. CEFR is a six-level scale reference for foreign language proficiency ascending from basic level to advanced level, with scales of A1, A2, B1, B2, C1 and C2. Likewise, a mix of words with a range of difficulties that are equivalent to these six levels were selected from the reading article so that participants could try to use all 4 colours in the intervention.

Among 30 specified words, some words were considered to be manageable by basic English users, such as culture, social media and expectation. At the other end of the spectrum, some advanced words were added to increase the test's difficulty, such as disillusioned, sordidness, and clandestine. All words were shuffled, regardless of difficulty level.

Examples of question 9 and 25 are shown below.

| | |
|---|---|
| 9. Thriving | 25. Taboo |
| It was a thriving place. | Do not break the taboo. |
| - 空無一人的<br>- 興旺的<br>- 神聖的<br>- 濕淋淋的 | - 規矩<br>- 沉默<br>- 法例<br>- 禁忌 |

Participants were required to look at the word and an example of the word in use. Then, they had to choose the meaning that most closely matches the word. The whole test can be referred to in the appendix A.

*3.4.4 Post-test*

The post-test was the same as the pre-test in terms of content and the order of questions. This helped to see if the participants got better at knowing a particular set of words, so the same set of words were used in both pre- and post- test. The scores from the post-test were collected to compare with those from the pre-test.

**3.5 Research procedures**

The research study was assigned as students' homework and students had flexibility to arrange their time to do the tasks as self-learning within the designated time frame. Participants were asked to complete three tasks: a pre-test of English vocabulary, a self-learning task with a reading article, and a post-test of English vocabulary.

The English vocabulary tests took approximately 10 minutes each. The self-learning task took approximately 45 minutes and participants were asked to read an article, learn and highlight new vocabulary and check the meanings in the dictionary. Some students used colours when highlighting the vocabulary and some did not, allowing the research team to see whether colours helped learning (i.e. did the group who used colour improve more from pre- to post- test than the control group?)

All groups had to read the same piece of an article. As they went through the article, the 30 specified words had a number next to them and were followed by a blank. Participants had to fill in the meaning of the word in Traditional Chinese by using a provided dictionary website. They could copy and paste the Chinese meaning they found suitable in the blank. After that, three groups were assigned a different task in the subsequent step.

For the multiple-colour group, they adopted the '4-colour memory' method in the intervention and highlighted both the word and its Chinese meaning in a colour to indicate how well they knew the word. For example, if they highlighted the word in blue, that meant they knew the word and the meaning before they looked it up in the dictionary. If they highlighted the word in green, that meant they had seen the word before but before they looked it up, they were not fully sure what it meant. If they highlighted the word in yellow, that meant they thought they had seen the word before but before they looked it

up, they did not know what it meant. If they highlighted the word in red, they had never seen or heard the word before.

For the one-colour group, they highlighted both the word and its Chinese meaning in yellow only. For the no-colour group, they needed to bold and underline both the word and its Chinese meaning but did not highlight any words in colour.

The participants were asked to play a game for approximately 15 minutes after the self-learning task in order that the test did not immediately follow training. We estimated that the whole study would take approximately 1.5 hours though the exact time varied. The tasks were completed as an individual and were student-friendly.

In terms of practicability, the school used Google Classroom as the students' online learning platform and participating students were divided into three learning groups in the Google classroom so that they were not able to view the instruction or testing content in the other two groups. Both pre-test and post-test were created with Microsoft Form and the links were embedded in the homework's instructions. After finishing the intervention, participants uploaded their self-learning task to Google Classroom individually. Figure 2 below illustrates the flow of the experiment in different stages.

Refer to Appendix B, C, and D for the examples of the three groups' reading comprehension.

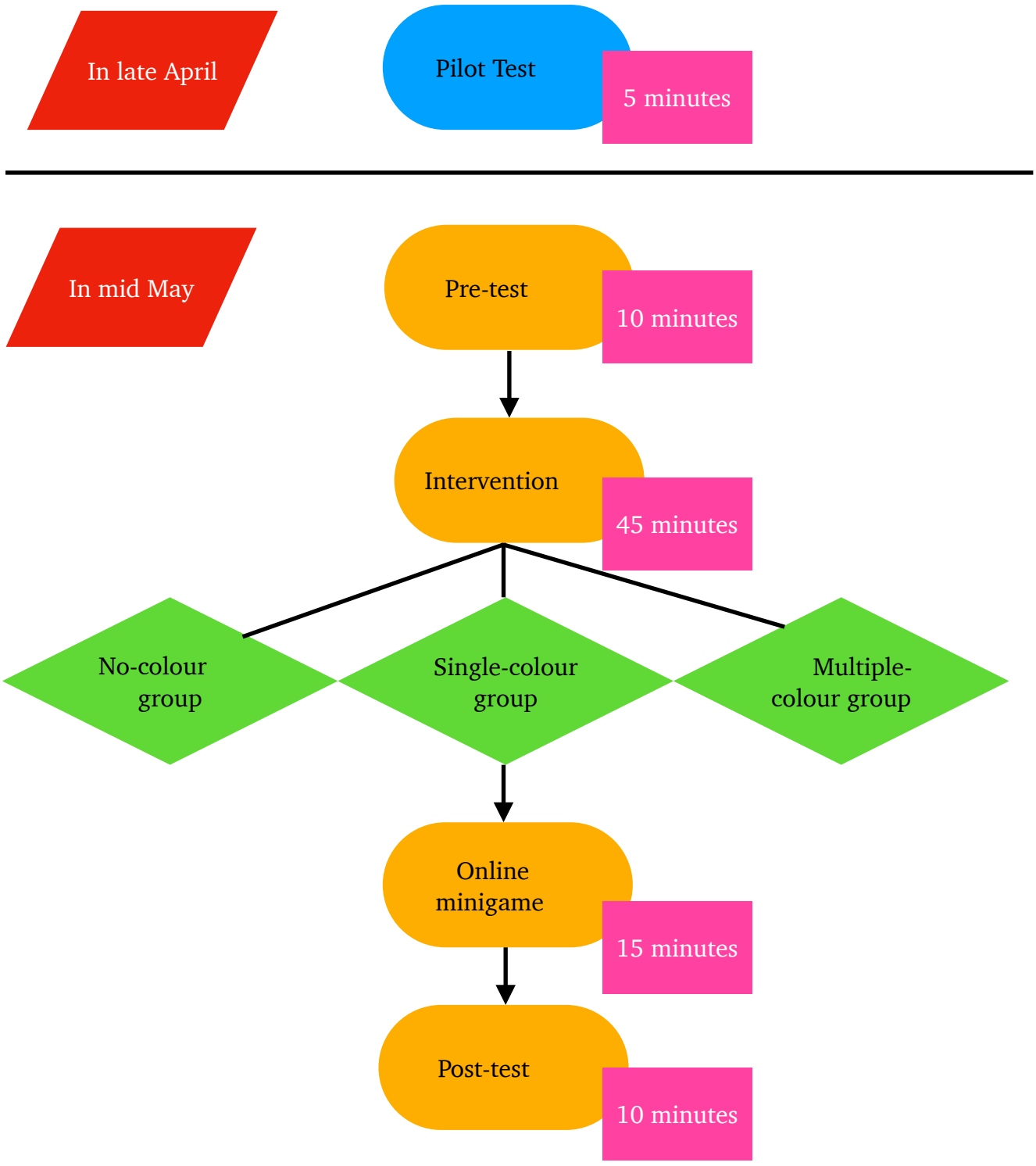



Figure 2 Flow chart showing the stages of the experiment

### 3.6 Ethical considerations

This research study received ethical approval from the Departmental Research Ethics Committee (DREC) of the Department of Education at University of Oxford. The reference number is CIA-22HT-051. (see Appendix E). Since the participants' age is 15 years old, this project followed Procedure 15 which is approved by the Ethics Committee for children aged under 16 years old in education settings. Prior to commencement of the project, the research team received consent from both the secondary school and participants' parents. Participants were well-informed of the voluntary nature of their participation. They had the right to withdraw from the study at any time without penalty.

# 4 Results

This chapter presents the analysis of the pre-test and post-test data of the research study. ANOVA is the major statistical analysis test. The data is analysed to answer research questions. Subsequently, numerous tests are run to check the data reliability.

## 4.1 Answering research questions

The research question is whether the use of colour aids memory for Chinese speakers' English L2 vocabulary acquisition. If so, which method is the most effective for L2 vocabulary acquisition by comparing groups using no colour, single colour and multiple colours?

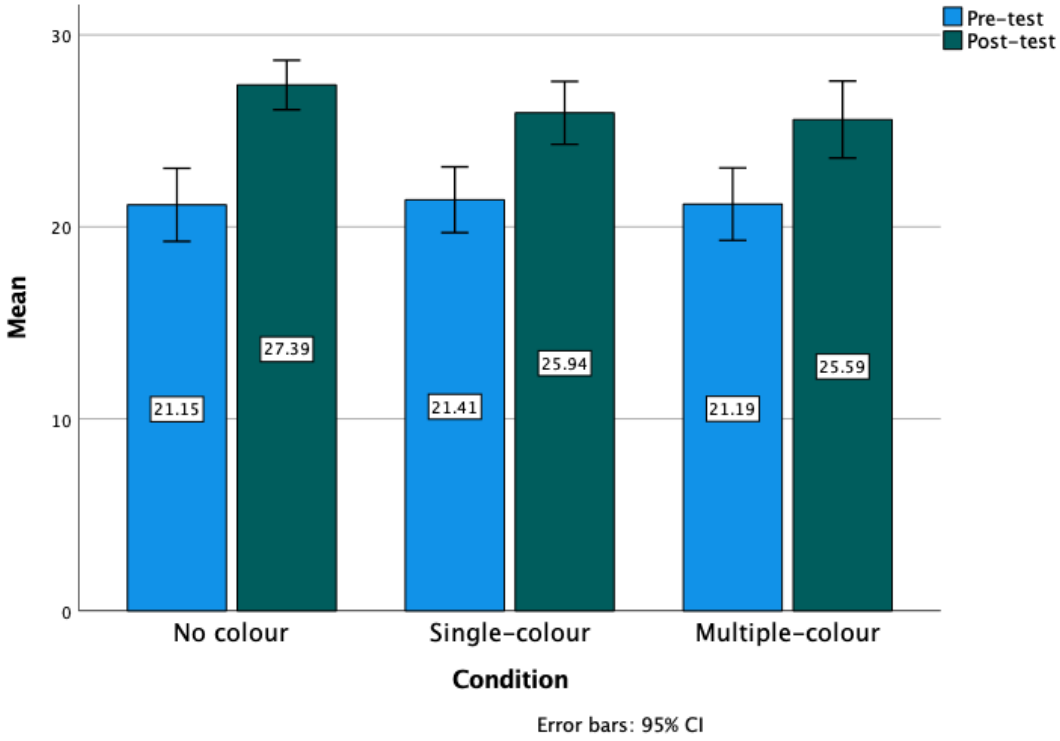



Figure 3 Bar chart showing the results of pre-tests and post-tests in the three conditions

Figure 3 shows results for pre-tests and post-tests for the three different experimental conditions. On inspection, it appears that the three groups scored similarly in the pre-test and all groups improved their performance in the post-test.

To see if there are differences in how the three groups improved from pre to post test, the planned analysis was a mixed ANOVA test with the 'within-participant' factor set with two levels (pre-test / post-test) and the 'between-participant' factor set with three levels (no colour, single-colour, multiple-colour). However, first, the data was tested to see if they met the assumptions necessary for ANOVA. First, the assumption of running ANOVA with the between-participants variables (condition) is that the variance within the three groups is similar (i.e. homogeneity of variance). To test if this assumption is established, Levene's test was conducted. Based on the Levene's test result, it did not show that the variances for the conditions either at pre-test ($F(2,96) = .630$, $p = .535$) or post-test ($F(2,96) = .630$, $p = .509$) are different. However, the assumption of normality was not met either at pre-test (Shapiro-Wilk $= .970$, $df = 99$, $p = .024$) or post-test (Shapiro-Wilk $= .970$, $df = 99$, $p = .024$). (For more details, histograms of both pre-tests and post-tests are in Appendix G and H).

Given that the assumption of normality was violated, for robustness of the study, the original planned ANOVA was run as well as two follow-up non-parametric statistical tests.

### 4.2 ANOVA

Overall, the main effect of Session on performance is significant ($F(1,96) = 105.512$, $p < .001$, $\eta_p^2 = .524$), denoting that participants improved from pre-test to post-test in general (mean pre-test=21.45, mean post-test=26.41). However, there was no significant main effect of Condition ($F(2, 96) = .36$, $p = .698$, $\eta_p^2 = .007$). Critically, there was no significant interaction between Session and Condition ($F(2, 96) = 1.445$, $p = .241$, $\eta_p^2 = .029$). In other words, no evidence was found to indicate that there were differences in the amount the different conditions improved by.

### 4.3 Non Parametric Tests

There is no direct non-parametric test equivalent of the mixed ANOVA available in SPSS. To see if there is an overall difference between sessions: a Wilcoxon signed rank test was executed to compare pre- and post scores for all participants. This found a significant difference ($z = -7.329$), $p < .001$). This supports the finding that there was a main effect of condition in the mixed ANOVA. To compare gains from training, a Kruskal-Wallis test serves as an alternative for a one-way ANOVA. This displayed no significant difference between

conditions (H=3.223, df=2, p=.2) - this presents no evidence that the conditions differ in how much the participants make gains from the intervention (i.e. it is not possible to tell how much the participants learn). This validates the absence of a significant interaction between condition and session in the previous ANOVA.

To answer the research question, given the conditions do not differ from each other overall in the amount they improve, it does not make sense to test individual contrasts between pairs of conditions.

# 5 Discussion

This chapter makes a comparison between the results of this study and previous studies and attempts to explain the differences, reviews the potential flaws of the study, proposes pedagogical practices for vocabulary teaching in accordance with subjects' performance in the experiment and suggests avenues for future research.

## 5.1 Key findings of the study

The current study attempted to delve into the effect of colour on memorising vocabulary in an L2. Instead of simply testing the effects between using one colour and no colour, this study goes further to examine the effect of multiple colours on vocabulary acquisition - with a multiple-colour coding method for learners to indicate their level of word knowledge. Therefore, 3 experimental conditions were set up to see if the use of colour can aid memory for Chinese speakers' English L2 vocabulary acquisition. If so, this study intends to find out the most effective method out of these 3 conditions.

This study used a pre- post design which allowed the researchers to see if the three groups improved after the activity and if the extent of improvement differed between groups. Based on the findings, there is a significant effect of sessions on subjects' performance. That means on average, participants had an improvement from the pre-test to the post-test. However, no significant differences were found in conditions or the interaction between the session and conditions. These results were further validated by the two non-parametric tests. It suggests that there is no evidence that the colour manipulations led to more improvement than the control condition.

## 5.2 Comparisons with the predictions

### *5.2.1 multiple-colour group vs other two groups*

The scores of pre-tests in the three groups are similar, ranging from 21.15 to 21.41 out of 30 which suggested that randomisation of participant allocation led to groups which were reasonably balanced in proficiency. In the post-tests, the scores of the three groups did not differ significantly, ranging from 25.59 to 27.39 out of 30. It was predicted that the

multiple-colour group would outperform the other two. The statistics shows that there is no evidence that any group has improved more than the other.

Why didn't using the 4-colour memory system help students learn the words? First of all, it is expected that all groups could at least gain some marks from the easy words in both pre- and post tests so the mark differences between the two tests might not be huge. In order to test the effect of 4-colour memory, words with different levels of difficulty were selected. Out of 30 selected words, some were set to be blue words, such as 'culture', 'social media' and 'expectation'. For the green words, some examples were 'harassment', 'scandal', 'compelling'. Some harder words such as 'debilitating', 'stalking', 'anonymous' were yellow words. The most advanced English vocabulary from the article were 'clandestine', 'sordidness' and 'juggernaut'. These were regarded as red words. Moreover, regardless of group conditions, it was expected participants could earn a certain number of marks in pre- and post- tests from the words in blue and green categories.

Another possibility is that the students might not be following the instructions in full. They were asked to use a colour to indicate how well they knew the word and encouraged to use all 4 colours. Looking at their coding of the sheets, although students managed to follow the instructions and thus highlighted the specified words in blue and red, quite a number of students did not make use of all 4 colours in the reading comprehension task as instructed. They only used red and blue colour to highlight the vocabulary but did not use yellow or green at all, representing that they either had no knowledge or full knowledge towards the vocabulary. The difference between green and yellow categories is subtle as green indicates that students have seen the word before but were not fully sure about the meaning while yellow indicates that students have seen the word before but did not know the meaning. Students may find it difficult to interpret the nuanced difference as both categories are not quite distinct from each other.

The third possibility is that some students were not cooperative. Their work reflected that they did not look words up in the dictionary but wrote down the meaning based on their own false knowledge. Some even left the task blank. To summarise, the students from 4-colour group condition were able to adopt the concept of using different colours to indicate

their familiarity towards the vocabulary. Despite this, it has been suspected that the ambiguity of definition between yellow and green as well as failure to complete the dictionary checking task may affect the result.

*5.2.2 single-colour group vs no-colour group*

It was hypothesised that the use of colour can yield better memory performance than using no colour. In fact there was no evidence that the groups differed in how much they improved, and in fact the no-colour group showed numerically higher performance than the single-colour group in the post-test. Results from an earlier study (Hosseini et al., 2022), however, found that colourful vocabulary, especially blue, orange and red colours are more beneficial to learning than black and white words. Why might the results of these studies differ?

In Hosseini et al.'s study, 20 participants at the age range between 18 and 29 had a chance to self-learn 144 vocabulary words in 6 sessions over 3 weeks. The vocabulary words that the participant had to learn were displayed on different coloured backgrounds, such as yellow, red, blue and so on. An immediate post-test was arranged in the second session of each week to measure short-term retention of the vocabulary. Delayed post-tests were set to test the participants' long-term vocabulary retention at least 2 weeks after the intervention. Researchers concluded that colours have a positive effect on recalling vocabulary in LTM. In this sense, participants had sufficient time to revise new words over study sessions and delayed post-tests were effective in testing vocabulary retention in the long term.

The importance of time intervals was emphasised in another vocabulary retention study (Goossens et al., 2012). Researchers believed that the spacing effect played a role in vocabulary learning which means words to be learned are distributed in different learning sessions. Children in the study were taught 30 new vocabulary words. Half of the words were delivered in one single study session. Another half were delivered over 3 learning sessions across several weeks. Results from vocabulary retention tests after 1 week and 5 weeks indicated that children had a better memory of the words that were acquired across multiple learning sessions instead of one. Thus, spacing effect helps vocabulary learning in general. It might be possible to see benefits of colour in a spaced design.

The above two research studies could possibly shed some light on why the single-colour group in the current study did not improve more than the no-colour group from pre- to post- test. One possible explanation might be the age factor. The age of the participants in Hosseini et al.'s study was between 18 and 29. They were all young adults. Comparatively, the participants in the current study were much younger and they were still adolescents. Different stages in cognitive development could possibly affect how individuals perceive colours. Besides, in the present study, students only had one study session for all 30 vocabulary words. In addition, there was no spacing. One study session might not be influential enough to test the effect of colour in vocabulary retention. To see effects of colour, it might be necessary to have multiple sessions. Besides, no delayed post-tests were arranged. The current study only had a 15-minute game break between the intervention and the immediate post-test. Regardless of colour conditions, this arrangement was more for testing the participants' STM but not related to colours as the delayed period was not long enough. Although it is not possible to tell which factors are critical on the current evidence, the above 3 points were still noteworthy.

### 5.3 Limitations of the study

First, '4-colour memory' vocabulary strategy is considered new in SLA research field as this strategy has not actually been tested experimentally. Likewise, it is also new to the participants. Based on the three groups' work, most of the students in the no-colour group and the single-colour group could follow the instructions strictly; however this was not the case for the multiple-colour group given that they mainly used red and blue colour without yellow and green colour at all. Therefore, to introduce this new vocabulary strategy, the definitions of 4 different colours could be addressed more carefully, especially dealing with the words falling between yellow and green colour categories. Instructions and examples in plain text might not be sufficient for students' understanding. In future work, it would be advisable to hold a separate learning session or film a step-by-step demonstration video to teach students how to use the strategy before the experiments.

Second, the time constraints of the experiment may not have led to a situation where the effect of colour could be seen. The whole experiment was conducted online and the

participants had to sit in front of a computer screen. In order to minimise the risk of eye strain, fatigue and boredom, the duration of all experimental tasks in total did not take longer than 1.5 hours. 45 minutes of the experiment time were assigned to the reading comprehension and self-learning session. This short period of time might not fully imitate students' self-learning time frame. More learning sessions could be arranged so that this could allow time to reinforce vocabulary knowledge in memory and students could experience the benefits of the spacing effect instead of learning in a massed condition. For the same reason, the time constraint made the students not possible to revise the words according to the colour over time. The colour coding method might specifically be useful to train students to be independent learners. It could be a self-paced strategy that learners could know how to pace themselves in learning vocabulary and doing revision. For example, the learners used '4-colour memory' to study a number of vocabulary last week and they would like to see which ones they should focus on for revision. They could quickly look at the ones in red. Then students could allocate proportionate time to revise words in each colour category. Moreover, since time intervals after intervention could be a useful measure to test long-term vocabulary retention, delayed post-tests could be arranged for students if the time constraints were not a concern.

Third, the study aimed to see how well the participants remembered the vocabulary, and see if colour helped them remember more in one case. However, since there was only a 15-minute break between the self-learning intervention and the post-test, students' memory of the vocabulary might still be fresh. The fact that the pre- and post-tests were repeated might mean that the students were learning the words in the test itself, and not learning from the intervention. Therefore, it has been suspected that the post-test served as a memory test of the vocabulary but was not relevant to the effect of colour. Future work could take an alternative approach that can test the isolated colour effect which omits the possibility of students remembering the answers straight from the intervention. A new intervention and a new pre- and post-test could be designed. The purpose is to test the effect of colour, maintaining the difficulty level whilst not using a repeated post-test. For the intervention part, the study could select 100 vocabulary words from the reading article with 25 words for each colour category, each category referring to a different difficulty level. Following this, the participants could colour code these 100 words in the intervention

to indicate their level of word knowledge. For the pre- and post-test, the study could set a certain number of questions for words in each difficulty level (i.e. 8 red words, 7 blue words). Then, the selected 25 words in each colour category could be randomly selected to meet the number of questions for each difficulty level. In addition, the study could create two tests and counter-balance the difficulty level so that half of the participants in each of the three groups could receive test A at the pre-test and test B at the post-test, and the other half could receive test B at the pre-test and test A at the post-test. In this case, each participant could receive different tests with the same level of difficulty.

### 5.4 Educational implications

The current work did not find evidence that colour was helpful for vocabulary learning, so researchers cannot draw direct implications from the finding. However, there are some observations that emerged from conducting the study that may be useful for teachers who are using the methods to consider. After reviewing students' work in the study, some pedagogical implications are recommended for L2 vocabulary acquisition. Participants in the study seemed to experience uncertainty in distinguishing the use of yellow and green colour when highlighting the specified vocabulary. It is recommended that the number of colours be lowered from 4 to 3. Words belonging to yellow or green colour categories can be grouped as one particular colour. Learners can justify their word knowledge based on 3 clear levels: full understanding, no full understanding, and no understanding. In this way, students could retain the idea of '4-colour memory' but attempt to use a simplified version.

Next, the ultimate goal of the '4-colour memory' is to change the colour of words to one level higher, for example, from red to yellow, yellow to green, then green to blue. The most ideal case is to turn all words into blue colour which means the learners have a full understanding of the words they learnt. Due to the time constraint in the study, it did not test this progressive phase. Language teachers could set a longer time frame, say 2 weeks or 2 months, preferably covering the entire period of formative assessments, such as weekly dictations, or term exams. Students are encouraged to review their progress of vocabulary learning by changing the colours from time to time. Teachers can then evaluate students' learning progress with reference to different assessments' results.

Moreover, in the self-learning task, students were provided with a dictionary website and they had to determine a Chinese meaning that matched the context and write it down in the blank. However, most participants across all groups made a mistake in one question - juggernaut. The word ‘juggernaut’ was set as a red word as it is not a common word in students’ everyday life. ‘Juggernaut' is polysemous in that it can mean i) a very large, heavy truck or ii) a large powerful force or organisation that cannot be stopped (Cambridge Dictionary, 2022). The sentence containing the word ‘juggernaut’ in the article was about the rising trend of Korean pop culture in the United States. To fit the context, the second meaning is a more appropriate choice. However, some students wrote down the first meaning instead, suggesting that they were lacking in dictionary skills.

Chen (2012) attempted to investigate the usefulness of dictionary use in vocabulary learning in reading context. In the reading task, participants using English-Chinese bilingual dictionaries were asked to underline the word they looked up while participants in non-dictionary conditions could guess the meaning based on the context. Results found that students in the experimental group achieved substantially higher scores than the control group in vocabulary retention tests. Chen emphasised the importance of dictionary use in vocabulary comprehension and incidental vocabulary acquisition by virtue of exposure to richer and more accurate lexical information. However, the study also noticed the same dictionary use problem among their participants as they also were confused about variant forms of the target words. Therefore, in respect of the students in the present study failing to select a more approximate meaning from polysemy, it is recommended that teachers can instil proper dictionary skills, for example, reading examples from each meaning, paying attention to both L1 translation and L2 definition, and detecting contextual meaning from sentence level but not just focusing on vocabulary level.

### 5.5 Suggestions for future research

To the best of the author’s knowledge, no study has ever tested ‘4-colour memory’ approach empirically so this study is pioneering in testing this vocabulary acquisition strategy. At the same time, there are still many uncharted areas of ‘4-colour memory’ that need exploring further. The present study looked into the effectiveness of ‘4-colour memory’ approach for vocabulary acquisition. Students only focused on one individual word each time. In the

future, research studies should redo the experiment with the changes that have already been described at word level. Then, the level can be extended to phrases, idioms or collocations. Given that WM has a limited capacity, it would be interesting to see how '4-colour memory' works if the number of words is higher (i.e. colour might actually be more beneficial in the situation where is more constrained)

Furthermore, this study used a quantitative method to assess L2 learners' vocabulary learning but did not elicit participants' personal feedback towards this self-learning strategy. Future research could collect additional qualitative data by inviting learners to share their opinions towards the strength and weakness of the '4-colour memory' approach. Learners can also use a journal to document their vocabulary learning journey with the use of '4-colour memory' on a regular basis and mark down the number of words they upgrade or downgrade the level of colour over time. The record format can also be flexible, for example, using a loose-leaf notebook to separate words from different colour categories and move the words to other colour pages when the learners observe a change in their word understanding. These can be useful materials for researchers to do qualitative analysis. Apart from the self-documentation, future research can further advocate autonomous learning by totally eliminating the role of teachers from the research. Students can have the freedom to choose the vocabulary and control their study schedule. Teachers can distribute self-evaluation vocabulary retention tests online for students to check their progress. In other words, teachers can take a coordinator role and let the students take up the dominant role and attain the purpose of self-learning. It might see a benefit if future research used a method where participants were more active in using the colour categories to aid their self-learning.

# 6 Conclusion

It is not uncommon for language learners to experience difficulties in memorising vocabulary in their L2 as memories fade quickly. This paper attempted to find a vocabulary strategy that can improve the memory performance of vocabulary learning. Past literature revealed that the use of colour could be an effective aid to facilitate memory for the reason that colours can draw more of humans' attention than other visual stimuli. Also, the neurons in visual cortex V4 are responsive to colours which are conducive to storing information in the working memory. In this sense, the method '4-colour memory' has potential to guide learners to memorise vocabulary words better.

'4-colour memory' is a new method that is student-oriented, suitable for self-learning and easy to carry out. '4-colour memory' requires learners to highlight words in 4 different colours based on their knowledge of the vocabulary, such as i) full understanding, ii) basic understanding, iii) have vague impression with almost no understanding, and v) no understanding. The present study used a pre- and post-test design to empirically test the effectiveness of '4-colour memory' on vocabulary acquisition. Participants were randomly assigned into 3 different experimental conditions to finish a self-learning reading comprehension task. When studying 30 specified vocabulary in the intervention, the 'multiple-colour' group adopted the '4-colour memory' approach to indicate their word knowledge with colours according to the description of colour categories, the 'single-colour' group only used the colour yellow to highlight the words, the 'no-colour' group did not use any colours. The results suggested that the participants generally improved from the pre-test to the post-test but no statistical differences were found between the learning session and conditions so it could not be ascertained whether colours helped the participants to learn. With reference to students' work, this paper shared some insights with both educators and learners into how to achieve mastery of vocabulary in self-learning.

# Reference


Adilin Mohd Anuardi, Muhammad Nur, Yamazaki, Atsuko K., & Eto, Kaoru. (2018). A Pre-Analysis of the Effect of White, Blue and Green Background Colours on Working Memory in a Reading Span Task. *Procedia Computer Science, 126*, 1847-1854.

Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International journal of teaching and education*, *3*(3), 21-34.

Andringa, Sible, & Godfroid, Aline. (2020). Sampling Bias and the Problem of Generalizability in Applied Linguistics. *Annual Review of Applied Linguistics, 40*, 134-142.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human Memory: A Proposed System and its Control Processes. *The Psychology of Learning and Motivation, 2*(C), 89-195.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice*, *20*(1), 50-57.

Austin, P. C., & Brunner, L. J. (2003). Type I error inflation in the presence of a ceiling effect. *The American Statistician, 57*(2), 97-104.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139*(1), 201-208.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. psychology press.

Baluška, František, & Levin, Michael. (2016). On having no head: Cognition throughout biological systems. *Frontiers in Psychology, 7*, 902.

Beglar, David. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101-118.

Bernot, M. J., & Metzler, J. (2014). A comparative study of instructor-and student-led learning in a large nonmajors biology course: Student performance and perceptions. *Journal of College Science Teaching*, *44*(1), 48-55.

Bhide, Amar, Shah, Prakesh S., & Acharya, Ganesh. (2018). A simplified guide to randomized controlled trials. *Acta Obstetricia Et Gynecologica Scandinavica, 97*(4), 380-387.

Bibi, Nosheen, Mumtaz, Nazia, & Saqulain, Ghulam. (2019). Effectiveness of colored visual prompts for three word phrase development in intellectually disabled (ID) children. *Journal of Medical and Allied Sciences, 9*(2), 42-47.

Bolderson, Sarah, Dosanjh, Christine, Milligan, Claudine, Pring, Tim, & Chiat, Shula. (2011). Colourful semantics: A clinical investigation. *Child Language Teaching and Therapy, 27*(3), 344-353.

Brown, Tracy L, Gore, Christopher L, & Carr, Thomas H. (2002). Visual Attention and Word Recognition in Stroop Color Naming. *Journal of Experimental Psychology. General, 131*(2), 220-240.

Brown, Steven D, & Reavey, Paula. (2015). Turning around on experience: The 'expanded view' of memory within psychology. *Memory Studies, 8*(2), 131-150.

Bryan, A. (1997) Colourful semantics. In S. Chiat, J. Law & J. Marshall (eds), *Language disorders in children and adults: psycholinguistic approaches to therapy (Whurr Publishers' books on communication disorders)*. London: Whurr, 143–61.

Cambridge Dictionary. (2022, September 28). *juggernaut - definition, audio pronunciation and more for juggernaut: 1. a very large, heavy truck 2. a large powerful force or*

*organization that cannot be stopped: See more in the Cambridge English-Chinese (Traditional) Dictionary - Cambridge Dictionary*. Retrieved October 3, 2022, from *https://dictionary.cambridge.org/dictionary/english-chinese-traditional/juggernaut*

Centers for Medicare and Medicaid Services. (2012). Part 9: Things to know if your written material is for older adults. *Toolkit for making written material clear and effective. Retrieved from https://www.cms.gov/outreach-and-education/outreach/writtenmaterialstoolkit/downloads/toolkitpart09.pdf*

Chang, Ting, Jo, Sung-Hyun, & Lu, Wei. (2011). Short-Term Memory to Long-Term Memory Transition in a Nanoscale Memristor. *ACS Nano, 5*(9), 7669-7676.

Chen, Yuzhen. (2012). Dictionary use and vocabulary learning in the context of reading. *International Journal of Lexicography, 25*(2), 216-247.

Chiu, J. (2010). *Test your word knowledge*. https://My.Vocabularysize.Com/. Retrieved August 8, 2022, from https://my.vocabularysize.com/

Clark, R.E. (2017). A History and Overview of the Behavioral Neuroscience of Learning and Memory. In R.E. Clark, S. Martin (eds), Behavioral Neuroscience of Learning and Memory. *Current Topics in Behavioral Neurosciences*, vol 37. Springer, Cham. https://ezproxy-prd.bodleian.ox.ac.uk:2102/10.1007/7854_2017_482

Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge University Press.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409-412.

Coombe, C. (2011). *Assessing vocabulary in the language classroom*. Kuala Lumpur, Malaysia: The Malaysian English Language Teaching Association.

Cop, Uschi, Drieghe, Denis, & Duyck, Wouter. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PloS One, 10*(8), E0134008.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research*, *169*, 323-338.

Ding, Yanren. (2007). Text memorization and imitation: The practices of successful Chinese learners of English. *System (Linköping), 35*(2), 271-280.

Divjak, D. (2019). Frequency in language: *Memory, attention and learning* (Cambridge core). Cambridge.

D'Orio, Wayne. (2010). Tony Buzan. Scholastic Administrator, 10(3), 22.

Finnegan, G. (2016, November 15). *How we perceive colour depends on our culture and language – Prof. Anna Franklin*. European Commission. Retrieved September 14, 2022, from https://ec.europa.eu/research-and-innovation/en/horizon-magazine/how-we-perceive-colour-depends-our-culture-and-language-prof-anna-franklin

Folgieri, Raffaella, Lucchiari, Claudio, & Marini, Daniele. (2013). Analysis of brain activity and response to colour stimuli during learning tasks: An EEG study. *Proceedings of SPIE - The International Society for Optical Engineering, 8652*, 86520I-86520I-12.

Fu, Tse-Wei, & Peng, Jui-Wen. (2020). The Effect and the Efficiency Balance of Font Size and Font Color Change on the Human Memory in Chinese Vocabulary. In *Engineering Psychology and Cognitive Ergonomics. Mental Workload, Human Physiology, and Human Energy* (Vol. 12186, Lecture Notes in Computer Science, pp. 52-61). Cham: Springer International Publishing.

Gathercole, S. E., & Alloway, T. P. (2007). *Understanding working memory. A classroom guide*. London, UK: Harcourt Assessment.

Gavin C.F., & Theibert A.B. (2020). Learning & memory. Amthor F.R., & Theibert A.B., & Standaert D.G., & Roberson E.D.(Eds.), *Essentials of Modern Neuroscience*. McGraw Hill. https://ezproxy-prd.bodleian.ox.ac.uk:9744/content.aspx?bookid=2938&sectionid=247989325

Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience, 4*(7), 563-572.

Geronimo, A. (2021). Meeting the Needs of All English Learners Through Vocabulary Instruction. *KOTESOL Proceedings 2021*, 237.

Goossens, Nicole A. M. C., Camp, Gino, Verkoeijen, Peter P. J. L., Tabbers, Huib K., & Zwaan, Rolf A. (2012). Spreading the words: A spacing effect in vocabulary learning. *Journal of Cognitive Psychology (Hove, England), 24*(8), 965-971.

Hanson, J. (2022, Sep 9). Communications Tips From The Accidental Queen: What we can learn from Elizabeth II. *Forbes*. https://www.forbes.com/sites/janehanson/2022/09/09/communications-tips-from-the-accidental-queen-what-we-can-learn-from-elizabeth-ii/?sh=1cc615e5177d

Hayden, Benjamin Y., & Gallant, Jack L. (2013). Working memory and decision processes in visual area V4. *Frontiers in Neuroscience, 7*(7), 18.

Herlitz, Agneta, & Bäckman, Lars. (1990). Recall of object names and colors of objects in normal aging and Alzheimer's disease. *Archives of Gerontology and Geriatrics, 11*(2), 147-154.

Hettiarachchi, Shyamani. (2016). The effectiveness of Colourful Semantics on narrative skills in children with intellectual disabilities in Sri Lanka. *Journal of Intellectual Disabilities, 20*(1), 18-33.

Hettiarachchi, Shyamani, Ranaweera, Mahishi & Walisundara, Dilini. (2018). Colour Me... Orange? Incorporating Aspects of the Colourful Semantics Approach into a TESOL Programme at Preschool. *Asian Journal of English Language Teaching, 27*, Asian journal of English language teaching, 2018, Vol.27.

Hettiarachchi, Shyamani, & Ranaweera, Mahishi. (2019). The effectiveness of using the Colourful Semantics approach to support language development with children who are deaf or hard-of-hearing. *Deafness & Education International, 21*(4), 157-173.

Heyselaar, Evelien, Segaert, Katrien, Walvoort, Serge J.W., Kessels, Roy P.C., & Hagoort, Peter. (2017). The role of nondeclarative memory in the skill for language: Evidence from syntactic priming in patients with amnesia. *Neuropsychologia, 101*, 97-105.

Heywood, CA, Gadotti, A, & Cowey, A. (1992). Cortical area V4 and its role in the perception of color. *The Journal of Neuroscience, 12*(10), 4056-4065.

Hosseini, S. Fatemeh, & Ghabanchi, Z. (2022). What's in a Color? A neuropsycholinguistic study on the effect of colors on EEG brainwaves, immediate emotional responses, and English language vocabulary retention among Iranian young adults. *Journal of Neurolinguistics, 63*, 101083.

Huff, T., Mahabadi, N., & Tadi, P. (2021). Neuroanatomy, Visual Cortex. In *StatPearls*. StatPearls Publishing.

Hutchison, Keith A. (2011). The Interactive Effects of Listwide Control, Item-Based Control, and Working Memory Capacity on Stroop Performance. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(4), 851-860.

Ibarra Santacruza, Diana, & Martínez Ortegaba, David. (2018). Can Working Memory Strategies Enhance English Vocabulary Learning?/¿Pueden las estrategias para entrenar la memoria de trabajo mejorar el aprendizaje de vocabulario en inglés? *HOW - A Colombian Journal for Teachers of English*, *25*(2), 29.

Ikeda, M., Patterson, K., Graham, K. S., Ralph, M. L., & Hodges, J. R. (2006). A horse of a different colour: do patients with semantic dementia recognise different versions of the same object as the same?. *Neuropsychologia*, *44*(4), 566-575.

Ilhomovna, D. Z., Ruzimurodovna, M. F., & Rashidovna, X. X. (2020). New methods in the study of English grammar. *Middle European Scientific Bulletin*, *4*, 34-35.

Jakovljević, T., Janković, M. M., Savić, A. M., Soldatović, I., Mačužić, I., Jakulin, T. J., ... & Ković, V. (2021). The effect of colour on reading performance in children, measured by a sensor hub: From the perspective of gender. *PloS one*, *16*(6), e0252622.

Johnson, D. D. (2000). Just the right word: Vocabulary and writing. *Perspectives on writing*, 162-186.

Kahneman, D. (1973). *Attention and effort* (Prentice-Hall series in experimental psychology). Englewood Cliffs: Prentice-Hall.

Khan, Jahangeer, & Liu, Chengyu. (2020). The impact of colors on human memory in learning English collocations: Evidence from south Asian tertiary ESL students. *Asian-Pacific Journal of Second and Foreign Language Education*, *5*(1), 1-10.

Kinoshita, Sachiko, Mills, Luke, & Norris, Dennis. (2018). The Semantic Stroop Effect is Controlled by Endogenous Attention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(11), 1730-1742.

Koizumi, R. (2012). Vocabulary and speaking. *The encyclopedia of applied linguistics*, 1-7.

Koppen, Camille, & Spence, Charles. (2007). Seeing the light: Exploring the Colavita visual dominance effect. *Experimental Brain Research, 180*(4), 737-754.

Kucewicz, Michal T., Dolezal, Jaromir, Kremen, Vaclav, Berry, Brent M., Miller, Laura R., Magee, Abigail L., . . . Worrell, Gregory A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports, 8*(1), 4949-7.

Kuipers, Jan Rouke, & Phillips, William A. (2022). Variations in Pupil Size Related to Memory for Recently Presented Words and Event Related Potentials. *Journal of Cognitive Neuroscience, 34*(7), 1119-1127.

Kursawe, Michael A., & Zimmer, Hubert D. (2015). Costs of storing colour and complex shape in visual working memory: Insights from pupil size and slow waves. *Acta Psychologica, 158*, 67-77.

Laufer, Batia. (1990). Ease and difficulty in vocabulary learning: Some teaching implications. *Foreign Language Annals, 23*(2), 147-155.

Laws, Glynis. (2002). Working memory in children and adolescents with Down syndrome: Evidence from a colour memory experiment. *Journal of Child Psychology and Psychiatry, 43*(3), 353-364.

Lee, J. D. (2006). Affect, Attention, and Automation. In A. F. Kramer, A. Kirlik & D. A. Wiegmann (Eds.), *Attention: From Theory to Practice* (pp. 73-89). Cary: Oxford University Press.

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics, 36*(5), 611-634.

Li, You, Liu, Mingxin, Zhang, Wei, Huang, Sai, Zhang, Bao, Liu, Xingzhou, & Chen, Qi. (2017). Neurophysiological correlates of visual dominance: A lateralized readiness potential investigation. *Frontiers in Psychology, 8*, 303.

Liu, Z. L. (2011). A study on English vocabulary learning strategies for non-English majors in independent college. *Cross-cultural communication*, *6*(4), 152-164.

Liu, S., Liu, S., Liu, Z., Peng, X., & Yang, Z. (2022). Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement. *Computers & Education*, *181*, 104461.

Machouche, H., Gharbi, A., & Elfidha, C. (2017). Implicit effects of online advertising on consumer cognitive processes. *Academy of Marketing Studies Journal*, *21*(2), 1-17.

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, *72*, 23-36.

Meara, P. (1995). The importance of an early emphasis on L2 vocabulary. *LANGUAGE TEACHER-KYOTO-JALT-*, *19*, 8-11.

Medina, J. (2011). *Brain rules: 12 principles for surviving and thriving at work, home, and school*. ReadHowYouWant.com.

Moghadam, S. H., Zainal, Z., & Ghaderpour, M. (2012). A review on the important role of vocabulary knowledge in reading comprehension performance. *Procedia-Social and Behavioral Sciences*, *66*, 555-563.

Moriguchi, Yusuke, & Hiraki, Kazuo. (2013). Prefrontal cortex and executive function in young children: A review of NIRS studies. *Frontiers in Human Neuroscience, 7*, 867.

Muhammad, Nur Adilin Mohd Anuardi, Shinohara, Hideyuki, & Yamazaki, Atsuko K. (2016). A Pre-study of Background Color Effects on the Working Memory Area of the Brain. *Procedia Computer Science, 96*, 1172-1178.

Muhammad, Nur Adilin Mohd Anuardi, Yamazaki, Atsuko K., & Eto, Kaoru. (2018). A Pre-Analysis of the Effect of White, Blue and Green Background Colours on Working Memory in a Reading Span Task. *Procedia Computer Science, 126*, 1847-1854.

Myers, Nicholas E, Stokes, Mark G, & Nobre, Anna C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences, 21*(6), 449-461.

Naghavi, Hamid Reza, & Nyberg, Lars. (2005). Common fronto-parietal activity in attention, memory, and consciousness: Shared demands on integration? *Consciousness and Cognition, 14*(2), 390-425.

Niklaus, Marcel, Nobre, Anna C., & Van Ede, Freek. (2017). Feature-based attentional weighting and spreading in visual working memory. *Scientific Reports, 7*(1), 42384.

Novawan, A. (2019). *Visually-Based Grammar Teaching. January 2011*, 161–172.

Olmos-Solis, Katya, Van Loon, Anouk M., & Olivers, Christian N. L. (2018). Pupil Dilation Reflects Task Relevance Prior to Search. *Journal Of Cognition, 1*(1), Journal of cognition. Volume 1:Issue 1 (2018).

Ou, L. C., Ronnier Luo, M., Sun, P. L., Hu, N. C., Chen, H. S., Guan, S. S., ... & Richter, K. (2012). A cross-cultural comparison of colour emotion for two-colour combinations. *Color Research & Application, 37*(1), 23-43.

Ozgen, E. (2004). Language, Learning, and Color Perception. *Current Directions in Psychological Science : A Journal of the American Psychological Society, 13*(3), 95-98.

Pajkossy, Péter, Szőllősi, Ágnes, & Racsmány, Mihály. (2020). Pupil size changes signal hippocampus-related memory functions. *Scientific Reports, 10*(1), 16393.

Peng, Peng, Barnes, Marcia, Wang, CuiCui, Wang, Wei, Li, Shan, Swanson, H. Lee, . . . Tao, Sha. (2018). A Meta-Analysis on the Relation Between Reading and Working Memory. *Psychological Bulletin, 144*(1), 48-76.

Pérez, Libia Maritza, & Alvira, Roberto. (2017). The Acquisition of Vocabulary Through Three Memory Strategies. *Colombian Applied Linguistics Journal, 19*(1), 103-116.

Power, Jonathan D., Schlaggar, Bradley L., Lessov-Schlaggar, Christina N., & Petersen, Steven E. (2013). Evidence for Hubs in Human Functional Brain Networks. *Neuron (Cambridge, Mass.), 79*(4), 798-813.

Qian, David D. (1999). Assessing the Roles of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension. *Canadian Modern Language Review, 56*(2), 276-307.

Radvansky, Gabriel A, Gibson, Bradley S, & McNerney, M. Windy. (2011). Synesthesia and Memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*(1), 219-229. doi: 10.1037/a0021329.

Rambe, Sojuangon. (2018). STRATEGIES USED BY GIFTED STUDENTS FOR MEMORIZING ENGLISH VOCABULARY AT ENGLISH DEPARTMENT OF IAIN PADANGSIDIMPUAN. *Tarbawi (Institut Agama Islam Negeri Kerinci. Fakultas Tarbiyah Dan Ilmu Keguruan), 14*(1), 65.

Randall, M. (2007). *Memory, psychology and second language learning* (Language learning and language teaching ; v. 19). Amsterdam ; Philadelphia: John Benjamins Pub.

Randolph, P. T., & Dubjelová, N. (2019). Creating neural connections: Personal color associations and vocabulary acquisition. *ORTESOL Journal, 36*, 25-27. Retrieved from https://www.proquest.com/scholarly-journals/creating-neural-connections-personal-color/docview/2544548801/se-2

Reber, Paul J. (2008). Cognitive Neuroscience of Declarative and Nondeclarative Memory. In *Advances in Psychology* (Vol. 139, pp. 113-123). Elsevier Science & Technology.

Richmann, K. J. (2006). Short term memory retention, how time and colour play a role. *St Martin's Biology journal*, *1*, 51-6.

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., ... & Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron*, *74*(1), 12-29.

Russell, Elbert W., & D'hollosy, Marilyn E. (1992). Memory and attention. *Journal of Clinical Psychology, 48*(4), 530-538.

Sachdev, Perminder S, Brodaty, Henry, Reppermund, Simone, Kochan, Nicole A, Trollor, Julian N, Draper, Brian, . . . Lux, Ora. (2010). The Sydney Memory and Ageing Study (MAS): Methodology and baseline medical and neuropsychiatric characteristics of an elderly epidemiological non-demented cohort of Australians aged 70–90 years. *International Psychogeriatrics, 22*(8), 1248-1264.

Schmitt, N. (1997). Vocabulary learning strategies. In D. N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 199–227). Cambridge University Press.

Schütze, U. (2016). Memory, Language, and the Brain. In *Language Learning and the Brain: Lexical Processing in Second Language Acquisition* (pp. 25-48). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316665619.004

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11.

Shalev, Nir, De Wandel, L, Dockree, P, Demeyere, Nele, & Chechlacz, Magdalena. (2017). Beyond time and space: The effect of a lateralized sustained attention task and brain stimulation on spatial and selective attention. *Cortex, 107*, 131-147.

Shao, Zeshu, Roelofs, Ardi, Martin, Randi C., & Meyer, Antje S. (2015). Selective Inhibition and Naming Performance in Semantic Blocking, Picture-Word Interference, and Color-Word Stroop Tasks. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 41*(6), 1806-1820.

Simon, Herbert A. (1981). Information-processing models of cognition. *Journal of the American Society for Information Science, 32*(5), 364-377.

Sinnett, S, Spence, C, & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception and Psychophysics, 69*, 673-686.

Smees, Rebecca, Hughes, James, Carmichael, Duncan A., & Simner, Julia. (2019). Learning in colour: Children with grapheme-colour synaesthesia show cognitive benefits in vocabulary and self-evaluated reading. *Philosophical Transactions. Biological Sciences, 374*(1787), Philosophical transactions. Biological sciences. Volume 374:Number 1787 (2019).

Souza, Alessandra S, Rerko, Laura, & Oberauer, Klaus. (2016). Getting More From Visual Working Memory: Retro-Cues Enhance Retrieval and Protect From Visual Interference. *Journal of Experimental Psychology. Human Perception and Performance, 42*(6), 890-910.

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research, 176*, 245-258.

Squire, Larry R. (1992). Declarative and Nondeclarative Memory: Multiple Brain Systems Supporting Learning and Memory. *Journal of Cognitive Neuroscience, 4*(3), 232-243.

Sternberg, R. J. (1999). *Cognitive psychology (2 nd ed.)*. Fort Worth, TX: Harcourt Brace College Publishers.

Stroop, J. Ridley. (1992). Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology. General, 121*(1), 15-23.

Susanto, A. (2017). The teaching of vocabulary: A perspective. *Jurnal Kata: Penelitian Tentang Ilmu Bahasa Dan Sastra*, *1*(2), 182-191.

Takashi, Ishii. (2013). *One-Minute Tips for Effective Memorizing*. SBクリエイティブ.

Tavakoli, H. (2012). *A dictionary of research methodology and statistics in applied linguistics*. Rahnama press.

Trujillo Tanner, C., Caserta, M. S., Kleinschmidt, J. J., Clayton, M. S., Bernstein, P. S., & Guo, J. W. (2018). Conducting research with older adults with vision impairment: Lessons learned and recommended best practices. *Gerontology and Geriatric Medicine*, *4*, 2333721418812624.

Truscott, J. (2014). Consciousness and Second Language Learning. Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781783092673

Tsang, Chi Lai, & Isaacs, Talia. (2022). Hong Kong secondary students' perspectives on selecting test difficulty level and learner washback: Effects of a graded approach to assessment. *Language Testing, 39*(2), 212-238.

Van der Elst, Wim, Van Boxtel, Martin P. J., Van Breukelen, Gerard J. P., & Jolles, Jelle. (2006). The Stroop Color-Word Test. *Assessment (Odessa, Fla.), 13*(1), 62-79.

Venkatesh, D., Ramachandra, D. L., Suresh Babu, N., & Rajan, B. K. (2002). Impact of psychological stress, gender and colour on visual response latency. *Indian journal of physiology and pharmacology, 46*(3), 333-337.

Walters, S. (2020). 10.1 Memory Models and Systems. *Psychology-1st Canadian Edition*.

Weekes, N.Y., & Zaidel, E. (1996). The Effects of Procedural Variations on Lateralized Stroop Effects. *Brain and Cognition, 31*(3), 308-330.

Wichmann, F. A. , Sharpe, L. T. & Gegenfurtner, K. R. (2002). The Contributions of Color to Recognition Memory for Natural Scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28* (3), 509-520. https://doi.org/10.1037/0278-7393.28.3.509

Wilkins, D. (1974). *Second-language learning and teaching*. London: Edward Arnold.

Yang, W., & Dai, W. (2011). Rote Memorization of Vocabulary and Vocabulary Development. *English Language Teaching, 4*(4), 61-64.

Yao, M., Wu, J., & Pang, H. (2000). The English vocabulary memorizing strategies of junior middle school students. *Psychological Science (China), 23*(6), 682–685.

Zavaruieva, Inna, Bondarenko, Larysa, & Fedko, Olha. (2022). The role of colour coding of educational materials when studying grammatical categories of the Ukrainian language by foreign students. *Review of Education (Oxford), 10*(1), N/a.

Zeki, S. (1990). The motion pathways of the visual cortex. *Vision: Coding and efficiency*, 321-345.

Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., & Frackowiak, R.S. (1991). A direct demonstration of functional specialization in human visual cortex. *The Journal of Neuroscience, 11*(3), 641-649.

Zeki, S., & Marini, Ludovica. (1998). Three cortical stages of colour processing in the human brain. *Brain (London, England : 1878), 121*(9), 1669-1685.

Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. *The modern language journal*, *96*(4), 558-575.

# Appendices

### Appendix A - Pre-test / Post-test

## Vocabulary Test A (30 點)

Show your true colours: Can the use of colour aid memory for Chinese speakers' English L2 vocabulary acquisition

1. There are 31 questions in this test
2. Look at the word and an example of the word in use
3. Choose the meaning that most closely matches the word

1. Please enter your participant ID (The code started with VL)
   Please confirm that you use the same participant ID throughout the study)
   輸入您的答案

2. Allegation
   She has made an allegation. (1 點)

   宣言
   協議
   介紹信
   指控

3. Compelling
   How can we create compelling content? (1 點)

   新奇有趣的
   引人入勝的
   數碼的
   互動的

4. Fandom
   Sports fandom (1 點)

   狂熱喜愛
   潮流
   場地
   裝備

5. Scandal
   It has become a public scandal. (1 點)

   主張
   領域

醜聞
諮詢

6. Bashing
The bashing has to stop. (1 點)

批評
口角
恐懼
短暫的放縱

7. Juggernaut
Here comes the juggernaut. (1 點)

強大力量
智力測驗題
收購
兇惡的人

8. Debilitating
He has a debilitating disease. (1 點)

致命的
會傳染他人的
慢性的
令人感到虛弱的

9. Thriving
It was a thriving place. (1 點)

空無一人的
興旺的
神聖的
濕淋淋的

10. Bullying
He hated bullying. (1 點)

隱瞞
愛爾蘭式曲棍球
欺淩
優柔寡斷

11. Ambivalence
I experienced ambivalence when I was about to leave my country. (1 點)

迷惘
處之泰然
緊張
心情矛盾

12. Social media
Some teenagers are addicted to social media. (1 點)

社交媒體
電子遊戲
網上購物
酒精

13. Veneer
What is behind the veneer? (1 點)

甲板
掩飾物
蠟燭台
接縫

14. Stalking
He was arrested for stalking. (1 點)

非法跟蹤騷擾
綁架
縱火
偽造文書

15. Candid
Please be candid. (1 點)

小心的
表示同情的
坦率的
公正的

16. Unparalleled
They did unparalleled research studies. (1 點)

領先的
空前的
初步的
全面的

17. Hunt down
They want to hunt down their enemies. (1 點)

攻擊
找到
戰勝
抵抗

18. Culture
She's interested in different countries' cultures. (1 點)

藝術
烹飪
文化
時裝

19. Expectation
You should manage your expectations. (1 點)

期望
時間
計劃
財政

20. Jubilant
He was in a jubilant mood. (1 點)

幸福的
激動的
歡欣雀躍的
慈祥的

21. Harassment
The law can protect people from harassment. (1 點)

騷擾行爲
暴力行為
詐騙行為
網路上偽裝自己的行為

22. Disillusioned
She felt disillusioned with her job. (1 點)

煩躁不安
疲倦

氣餒
理想破滅

23. Sordidness
The book revealed the sordidness of the industry. (1 點)

趣聞
不道德
演變
燦爛

24. Panopticon
People live in a panopticon society. (1 點)

萬花筒
世俗的
封建的
監獄

25. Taboo
Do not break the taboo. (1 點)

規矩
沉默
法例
禁忌

26. Devotion
He shows devotion to his family. (1 點)

關懷
摯愛
善意
惡意

27. Anonymous
They accept anonymous donations. (1 點)

企業的
國際的
匿名的
巨額的

28. Coincidence
There were a series of coincidences. (1 點)

不幸事件
不確定因素
碰巧的事
謎團

29. Heyday
It happened in her heyday. (1 點)

過去的時期
違法的時期
反叛時期
全盛時期

30. Bolster
They are going to bolster the invasion. (1 點)

發動
擊退
阻撓
提高

31. Clandestine
They are going to hold a clandestine wedding. (1 點)

華麗的
盛大的
暗中的
簡單的

## Appendix B - Sample of no-colour group's intervention

**Instruction**
Your aim in this task is to study 30 specified words which each have a number next to them.

1. Please click "File" --> "Save as" --> Save as a new document in your desktop and named the document as **your participant ID** and submit it to Google classroom after you finished the task
2. Please read the article below. As you go through, some of the words have a number and are followed by a blank. Your first job is to fill in the meaning of the word in Chinese.

   Use the link below to check the Cambridge dictionary, copy and paste the **Chinese** meaning you find suitable next to the word https://dictionary.cambridge.org/dictionary/english-chinese-traditional/
3. Next, we want you to **bold** and underline both the word and its Chinese meaning.

For example,
As a teen **member 成員，一分子** of the band U-KISS, Kevin Woo felt the hot glare of the K-pop limelight.

# At KCON L.A., frank talk about mental health amid the ear-splitting meet-and-greets

As a teen member of the band U-KISS, Kevin Woo felt the hot glare of the K-pop limelight. His time in the band showed the California-born singer what life in the particular pressure cooker of South Korean fame meant, and how the [1]**expectations 期盼；盼望；希望** can be [2]**debilitating 使虛弱；削弱，使衰弱**.

"In training, [the record labels] take a lot of privacy from you," said Woo, now a solo artist who will appear at this weekend's KCON convention in downtown L.A. "They take your phone away when you debut. You can't date, which was very shocking for me. They want trainees to have a certain figure, so you're dieting and there's pressure to get plastic surgery."

Record labels mold and cast band members when they're young, and exert total control over their lives and images. That can drive you to dark places. "I've seen artists who have been affected by that, when you feel like you're being watched too closely," Woo said. "You get scared of people, of going out. You constantly fear someone taking pictures. You can't live a comfortable private life."

For young fans of the genre, the pressures of modern life can mirror those faced by K-pop stars. We live in a [3]**social media 社交媒體（網站或電腦程式）** [4]**panopticon （看守室在中央的）全景監獄，環形監獄**, where one false move can destroy your reputation. Perfection is expected. Gigantic corporations have colonized our lives.

So it's no [5]**coincidence 同時發生；（尤指令人吃驚的）巧合，碰巧的事** that some of the most interesting presentations and panels at this year's KCON L.A., a four-day concert and fan event that

draws well over 100,000 people, deal with mental health issues in the scene. The main concert sports sets from rising stars like Ateez, Stray Kids and Loona, alongside a bevy of idol meet-and-greets, Korean beauty tutorials and dance workshops.

There's immense value in the connection K-pop creates between fans, many of whom come from marginalized backgrounds. But it can also be lonely and exhausting onstage for artists in such a rigid system. Fans' all-consuming [6]**devotion 忠誠，忠實；摯愛** to their idols can turn threatening, both online and off-.

At KCON, fans and scholars are acknowledging that the scene is both an asset and a challenge when it comes to fans' and artists' mental health.

"When I was younger, I went through my own mental health difficulties, which just weren't talked about in my [7]**culture 文化（尤指某一群體在某一時期普遍的風俗習慣和信仰）**," said Janet Ly, a Chinese American family therapist and hallyu fan who will speak at a panel on mental health and K-pop. The scene helped her "say whatever I needed to say, and validated my experiences. There's a strong emphasis on community because K-pop is not as mainstream in the U.S. That helps you feel connected to other people."

But also, Ly added, "Being [8]**anonymous 匿名的；不知姓名的；名字不公開的** through a screen is scary. You only see words, not the lives being affected. They can have a great impact on somebody."

K-pop's rise in the U.S. elevated the scene (long established in Asia) from an internet-driven curiosity here to a [9]**thriving 茁壯成長；興旺，繁榮** subculture and, finally, into a multimedia [10]**juggernaut 重型貨車** with bands selling out stadiums and signing to major labels. With it came an ultra-passionate fan base where young audiences' devotion to acts is [11]**unparalleled 無雙的，無比的；空前的** since the boy-band/"TRL" [12]**heyday 全盛時期，鼎盛時期** of the late '90s.

BTS performs at the Rose Bowl, May 4, 2019.(Kent Nishimura / Los Angeles Times)

For most, that community built around outsider pop [13]**fandom 狂熱喜愛** is an asset. As the genre grew, events like KCON became hubs for meeting idols and one another. For a millennial generation where, in one recent study, 22% claimed they had zero close friends, that's nourishing.

"I met my roommate, my best friend and been to weddings because of people I met at KCON," said Shelby Moses, a KCON fan organizer and a speaker on a panel about hyper-devoted fans (or "stans"). "It created space for everyone to show up."

Many of these fans are from racial, gender or sexual identities under threat today, and the upbeat K-pop scene is a godsend. Many of them come from cultural backgrounds (often Asian American but others as well) where mental health can be a [14]**taboo （宗教或社會習俗方面的）禁忌，忌諱** topic, or where affordable resources for treatment are scarce.

"K-pop provides space for people who have felt outside of their own communities. You hear K-pop fans say that it's been their mental health savior, that they feel connected and supported and affirmed when they can't find it in their own world," said Patty Ahn, a professor at UC San Diego who studies South Korean pop culture (and who is speaking on cultural clashes at KCON).

Ahn is working on a documentary about black K-pop fans and has seen firsthand how the genre can [15]**bolster 支撐；加固；提高；改善** those who don't neatly fit into any one culture.

"The fandom tracks with a lot of outsiders, be it gender or black or Latinx outsiders within their own communities, and that's particularly [16]**compelling （理由、論證等）令人信服的，很有說服力的**," Ahn said. "I'm not sure I've seen that in other fandoms."

But the pressures of fame and the fever pitch of fandom can exact its own toll.

The 2017 suicide of SHINee member Jonghyun shook the fandom and brought to light the genre's long-standing issues around mental health. Even BTS, the biggest K-pop group in the world, admitted to similar feelings sometimes. "I really want to say that everyone in the world is lonely and everyone is sad," BTS member Suga told Billboard after Jonghyun's death. "I hope we can create an environment where we can ask for help."

"Every day is stressful for our generation. It's hard to get a job, it's harder to attend college now more than ever," said BTS' RM in the same interview. "Adults need to create policies that can facilitate that overall social change."

A lot of that pressure for idols' constant perfection comes from the all-powerful record labels. But sometimes, it can come from fans as well.

"I turn to fans when I need support, and that always boosts my confidence. But there are fans who take it to the next level, who are too obsessive with idol groups because they love them so much they become controlling," Woo said. "When a member messes up or has a [17]**scandal 醜行；醜聞；震驚；反感**, so much [18]**bashing 針對…的嚴厲批評** has a negative effect."

The kind of [19]**harassment 騷擾行爲** that can happen in K-pop has echoes of the same [20]**stalking （在一段時期內）非法跟蹤騷擾** dynamics, both physical and digital, that many young people, especially young women, face in their own lives.

"There are a lot of instances of people [21]**hunting down 搜捕，找到** idols and breaking into their hotels. BTS had a plane delayed because people bought tickets just to take pictures of them," Moses said. "We saw how deteriorating that can be — there have been suicides. It's really sad."

Moses has had to moderate her own fandom to avoid harassment over the years. "You see a lot of [22]**bullying 恃強淩弱者，以大欺小者** for very stupid reasons," she said. "My friends and I used to do K-pop cosplay with elaborate outfits, and got bullying and death threats after someone posted pictures of us. We were like, 'OK, this has changed,' and we backed out of that scene."

Many fans were hurt and [23]**disillusioned 理想破滅的；不抱幻想的** by the still-ongoing Burning Sun scandal, which implicated several high-profile K-pop stars and industry figures in a ring of sex crimes. Charges included [24]**allegations （未經證實的）指責，指控** of [25]**clandestine （尤指從事不被官方允許的活動）秘密的，暗中的，私下的** videotapes of assaults at a popular Seoul nightclub shared over text-message chains.

The scale and [26]**sordidness 骯髒和不愉快的品質** of that scandal pulled back the curtain on the image of squeaky-clean K-pop idols. But fans still wrestle with what that means for their own attachments and communities. At best, it opens up space for fans to be [27]**candid （尤指對棘手或令人痛苦之事）率直的，坦誠的，直言不諱的** about their own experiences with abuse. But

time will tell if the scene is ready to listen.

“We have so much emotionally wrapped up in music and a sense of belonging to it. This is a time where we need to be open and have space for dialogue,” Ahn said. “Idols and celebrities need to be held accountable,” Ahn added, mentioning Seungri, a central figure charged in the Burning Sun scandal. “Idols have a tremendous amount of power and it’s getting difficult to hold any [28]**ambivalence （心情）矛盾的；模棱兩可的，含糊不定的**. You can love an idol but have complicated feelings.”

Underneath the shiny [29]**veneer 飾面薄板；鑲板** and tight corporate control of K-pop, profound pain and abuse can boil over. Now that K-pop has matured in the U.S., most in the scene are glad that more difficult conversations are finally happening in public. It’s not as [30]**jubilant （尤指因成功而）歡欣的，喜氣洋洋的** as a new Blackpink single, but it might be more important.

“I like the awareness that it’s not just happiness and sunshine and rainbows,” Ly said. “They’re people too, they make mistakes and struggle. They’re humans but very famous humans.”

## Appendix C - Sample of single-colour group's intervention

**Instruction**

Your aim in this task is to study 30 specified words which each have a number next to them.

1. Please click "File" --> "Save as" --> Save as a new document in your desktop and named the document as **your participant ID** and submit it to Google classroom after you finished the task
2. Please read the article below. As you go through, some of the words have a number and are followed by a blank. Your first job is to fill in the meaning of the word in Chinese.

   Use the link below to check the Cambridge dictionary, copy and paste the **Chinese** meaning you find suitable next to the word https://dictionary.cambridge.org/dictionary/english-chinese-traditional/
3. Next, we want you to highlight both the word and its Chinese meaning in yellow

For example,

As a teen member 成員，一分子 of the band U-KISS, Kevin Woo felt the hot glare of the K-pop limelight.

# At KCON L.A., frank talk about mental health amid the ear-splitting meet-and-greets

As a teen member of the band U-KISS, Kevin Woo felt the hot glare of the K-pop limelight. His time in the band showed the California-born singer what life in the particular pressure cooker of South Korean fame meant, and how the [1]expectations ___期盼_____ can be [2]debilitating ___削弱_____.

"In training, [the record labels] take a lot of privacy from you," said Woo, now a solo artist who will appear at this weekend's KCON convention in downtown L.A. "They take your phone away when you debut. You can't date, which was very shocking for me. They want trainees to have a certain figure, so you're dieting and there's pressure to get plastic surgery."

Record labels mold and cast band members when they're young, and exert total control over their lives and images. That can drive you to dark places. "I've seen artists who have been affected by that, when you feel like you're being watched too closely," Woo said. "You get scared of people, of going out. You constantly fear someone taking pictures. You can't live a comfortable private life."

For young fans of the genre, the pressures of modern life can mirror those faced by K-pop stars. We live in a [3]social media ___社交媒體_____ [4]panopticon _全景監獄，環形監獄_____, where one false move can destroy your reputation. Perfection is expected. Gigantic corporations have colonized our lives.

So it's no [5]coincidence __巧合_____ that some of the most interesting presentations and panels at this year's KCON L.A., a four-day concert and fan event that draws well over 100,000 people, deal with mental health issues in the scene. The main concert sports sets from rising stars

like Ateez, Stray Kids and Loona, alongside a bevy of idol meet-and-greets, Korean beauty tutorials and dance workshops.

There's immense value in the connection K-pop creates between fans, many of whom come from marginalized backgrounds. But it can also be lonely and exhausting onstage for artists in such a rigid system. Fans' all-consuming [6]devotion ___摯愛_____ to their idols can turn threatening, both online and off-.

At KCON, fans and scholars are acknowledging that the scene is both an asset and a challenge when it comes to fans' and artists' mental health.

"When I was younger, I went through my own mental health difficulties, which just weren't talked about in my [7]culture ____文化_____," said Janet Ly, a Chinese American family therapist and hallyu fan who will speak at a panel on mental health and K-pop. The scene helped her "say whatever I needed to say, and validated my experiences. There's a strong emphasis on community because K-pop is not as mainstream in the U.S. That helps you feel connected to other people."

But also, Ly added, "Being [8]anonymous ____匿名的_____ through a screen is scary. You only see words, not the lives being affected. They can have a great impact on somebody."

K-pop's rise in the U.S. elevated the scene (long established in Asia) from an internet-driven curiosity here to a [9]thriving ____興旺，繁榮_____ subculture and, finally, into a multimedia [10]juggernaut __不可抗拒的強大力量_____ with bands selling out stadiums and signing to major labels. With it came an ultra-passionate fan base where young audiences' devotion to acts is [11]unparalleled ___空前的_____ since the boy-band/"TRL" [12]heyday ____全盛時期_____ of the late '90s.

BTS performs at the Rose Bowl, May 4, 2019.(Kent Nishimura / Los Angeles Times)

For most, that community built around outsider pop [13]fandom ___狂熱喜愛_____ is an asset. As the genre grew, events like KCON became hubs for meeting idols and one another. For a millennial generation where, in one recent study, 22% claimed they had zero close friends, that's nourishing.

"I met my roommate, my best friend and been to weddings because of people I met at KCON," said Shelby Moses, a KCON fan organizer and a speaker on a panel about hyper-devoted fans (or "stans"). "It created space for everyone to show up."

Many of these fans are from racial, gender or sexual identities under threat today, and the upbeat K-pop scene is a godsend. Many of them come from cultural backgrounds (often Asian American but others as well) where mental health can be a [14]taboo ___禁忌_____ topic, or where affordable resources for treatment are scarce.

"K-pop provides space for people who have felt outside of their own communities. You hear K-pop fans say that it's been their mental health savior, that they feel connected and supported and affirmed when they can't find it in their own world," said Patty Ahn, a professor at UC San Diego who studies South Korean pop culture (and who is speaking on cultural clashes at KCON).

Ahn is working on a documentary about black K-pop fans and has seen firsthand how the genre can [15]bolster ____支撐_____ those who don't neatly fit into any one culture.

"The fandom tracks with a lot of outsiders, be it gender or black or Latinx outsiders within their own communities, and that's particularly [16]compelling ___扣人心弦的_____," Ahn said. "I'm

not sure I've seen that in other fandoms."

But the pressures of fame and the fever pitch of fandom can exact its own toll.

The 2017 suicide of SHINee member Jonghyun shook the fandom and brought to light the genre's long-standing issues around mental health. Even BTS, the biggest K-pop group in the world, admitted to similar feelings sometimes. "I really want to say that everyone in the world is lonely and everyone is sad," BTS member Suga told Billboard after Jonghyun's death. "I hope we can create an environment where we can ask for help."

"Every day is stressful for our generation. It's hard to get a job, it's harder to attend college now more than ever," said BTS' RM in the same interview. "Adults need to create policies that can facilitate that overall social change."

A lot of that pressure for idols' constant perfection comes from the all-powerful record labels. But sometimes, it can come from fans as well.

"I turn to fans when I need support, and that always boosts my confidence. But there are fans who take it to the next level, who are too obsessive with idol groups because they love them so much they become controlling," Woo said. "When a member messes up or has a [17]scandal ___醜聞_____, so much [18]bashing ___針對...的嚴厲批評_____ has a negative effect."

The kind of [19]harassment __騷擾行爲_____ that can happen in K-pop has echoes of the same [20]stalking ____非法跟蹤騷擾_____ dynamics, both physical and digital, that many young people, especially young women, face in their own lives.

"There are a lot of instances of people [21]hunting down ___找到_____ idols and breaking into their hotels. BTS had a plane delayed because people bought tickets just to take pictures of them," Moses said. "We saw how deteriorating that can be — there have been suicides. It's really sad."

Moses has had to moderate her own fandom to avoid harassment over the years. "You see a lot of [22]bullying ____欺淩_____ for very stupid reasons," she said. "My friends and I used to do K-pop cosplay with elaborate outfits, and got bullying and death threats after someone posted pictures of us. We were like, 'OK, this has changed,' and we backed out of that scene."

Many fans were hurt and [23]disillusioned __理想破滅的_____ by the still-ongoing Burning Sun scandal, which implicated several high-profile K-pop stars and industry figures in a ring of sex crimes. Charges included [24]allegations ___指控_____ of [25]clandestine ____秘密的，暗中的_____ videotapes of assaults at a popular Seoul nightclub shared over text-message chains.

The scale and [26]sordidness ___不道德_____ of that scandal pulled back the curtain on the image of squeaky-clean K-pop idols. But fans still wrestle with what that means for their own attachments and communities. At best, it opens up space for fans to be [27]candid ___坦誠的_____ about their own experiences with abuse. But time will tell if the scene is ready to listen.

“We have so much emotionally wrapped up in music and a sense of belonging to it. This is a time where we need to be open and have space for dialogue,” Ahn said. “Idols and celebrities need to be held accountable,” Ahn added, mentioning Seungri, a central figure charged in the Burning Sun scandal. “Idols have a tremendous amount of power and it’s getting difficult to hold any [28]ambivalence ___矛盾的_____. You can love an idol but have complicated feelings.”

Underneath the shiny [29]veneer ___虛飾_____ and tight corporate control of K-pop, profound pain and abuse can boil over. Now that K-pop has matured in the U.S., most in the scene are glad that more difficult conversations are finally happening in public. It’s not as [30]jubilant ____歡欣的_____ as a new Blackpink single, but it might be more important.

“I like the awareness that it’s not just happiness and sunshine and rainbows,” Ly said. “They’re people too, they make mistakes and struggle. They’re humans but very famous humans.”

## Appendix D - Sample of multiple-colour group's intervention

**Instruction**

Your aim in this task is to study 30 specified words which each have a number next to them.

1. Please click "File" --> "Save as" --> Save as a new document in your desktop and named the document as **your participant ID** and submit it to Google classroom after you finished the task
2. Please read the article below. As you go through, some of the words have a number and are followed by a blank. Your first job is to fill in the meaning of the word in Chinese.

   Use the link below to check the Cambridge dictionary, copy and paste the **Chinese** meaning you find suitable next to the word https://dictionary.cambridge.org/dictionary/english-chinese-traditional/
3. Next, we want you to use a colour to indicate how well you know the word. Please use this colour index.

   Blue: I know the word well and I knew the meaning before I looked it up
   Green: I have seen this word before but before I looked it up, I was not fully sure what it meant
   Yellow: I think I have seen this word before but before I looked it up, I did not know what it meant
   Red: I have never seen or heard this word before

Highlight both the word and its Chinese meaning, for example,

As a teen member 成員，一分子 of the band U-KISS, Kevin Woo felt the hot glare of the K-pop limelight.

**Please try to make use of all four colours as you do the task**

# At KCON L.A., frank talk about mental health amid the ear-splitting meet-and-greets

As a teen member of the band U-KISS, Kevin Woo felt the hot glare of the K-pop limelight. His time in the band showed the California-born singer what life in the particular pressure cooker of South Korean fame meant, and how the [1]expectations 期盼 can be [2]debilitating 削弱.

"In training, [the record labels] take a lot of privacy from you," said Woo, now a solo artist who will appear at this weekend's KCON convention in downtown L.A. "They take your phone away when you debut. You can't date, which was very shocking for me. They want trainees to have a certain figure, so you're dieting and there's pressure to get plastic surgery."

Record labels mold and cast band members when they're young, and exert total control over their lives and images. That can drive you to dark places. "I've seen artists who have been affected by that, when you feel like you're being watched too closely," Woo said. "You get scared of people, of going out. You constantly fear someone taking pictures. You can't live a comfortable private life."

For young fans of the genre, the pressures of modern life can mirror those faced by K-pop stars. We live in a [3]social media 社交媒體 [4]panopticon 全景監獄, where one false move can destroy your reputation. Perfection is expected. Gigantic corporations have colonized our lives.

So it's no [5]coincidence 巧合 that some of the most interesting presentations and panels at this year's KCON L.A., a four-day concert and fan event that draws well over 100,000 people, deal with mental health issues in the scene. The main concert sports sets from rising stars like Ateez, Stray Kids and Loona, alongside a bevy of idol meet-and-greets, Korean beauty tutorials and dance workshops.

There's immense value in the connection K-pop creates between fans, many of whom come from marginalized backgrounds. But it can also be lonely and exhausting onstage for artists in such a rigid system. Fans' all-consuming [6]devotion 忠誠 to their idols can turn threatening, both online and off-.

At KCON, fans and scholars are acknowledging that the scene is both an asset and a challenge when it comes to fans' and artists' mental health.

"When I was younger, I went through my own mental health difficulties, which just weren't talked about in my [7]culture 文化," said Janet Ly, a Chinese American family therapist and hallyu fan who will speak at a panel on mental health and K-pop. The scene helped her "say whatever I needed to say, and validated my experiences. There's a strong emphasis on community because K-pop is not as mainstream in the U.S. That helps you feel connected to other people."

But also, Ly added, "Being [8]anonymous 匿名的 through a screen is scary. You only see words, not the lives being affected. They can have a great impact on somebody."

K-pop's rise in the U.S. elevated the scene (long established in Asia) from an internet-driven curiosity here to a [9]thriving 興旺 subculture and, finally, into a multimedia [10]juggernaut 重型貨車 with bands selling out stadiums and signing to major labels. With it came an ultra-passionate fan base where young audiences' devotion to acts is [11]unparalleled 無雙的 since the boy-band/"TRL" [12]heyday 全盛時期 of the late '90s.

BTS performs at the Rose Bowl, May 4, 2019.(Kent Nishimura / Los Angeles Times)

For most, that community built around outsider pop [13]fandom 狂熱喜愛 is an asset. As the genre grew, events like KCON became hubs for meeting idols and one another. For a millennial generation where, in one recent study, 22% claimed they had zero close friends, that's nourishing.

"I met my roommate, my best friend and been to weddings because of people I met at KCON," said Shelby Moses, a KCON fan organizer and a speaker on a panel about hyper-devoted fans (or "stans"). "It created space for everyone to show up."

Many of these fans are from racial, gender or sexual identities under threat today, and the upbeat K-pop scene is a godsend. Many of them come from cultural backgrounds (often Asian American but others as well) where mental health can be a [14]taboo 禁忌 topic, or where affordable resources for treatment are scarce.

"K-pop provides space for people who have felt outside of their own communities. You hear K-pop fans say that it's been their mental health savior, that they feel connected and supported and affirmed when they can't find it in their own world," said Patty Ahn, a professor at UC San Diego who studies South Korean pop culture (and who is speaking on cultural clashes at KCON).

Ahn is working on a documentary about black K-pop fans and has seen firsthand how the genre can [15]bolster 支撐 those who don't neatly fit into any one culture.

"The fandom tracks with a lot of outsiders, be it gender or black or Latinx outsiders within their own communities, and that's particularly [16]compelling 令人信服的," Ahn said. "I'm not sure I've seen that in other fandoms."

But the pressures of fame and the fever pitch of fandom can exact its own toll.

The 2017 suicide of SHINee member Jonghyun shook the fandom and brought to light the genre's long-standing issues around mental health. Even BTS, the biggest K-pop group in the world, admitted to similar feelings sometimes. "I really want to say that everyone in the world is lonely and everyone is sad," BTS member Suga told Billboard after Jonghyun's death. "I hope we can create an environment where we can ask for help."

"Every day is stressful for our generation. It's hard to get a job, it's harder to attend college now more than ever," said BTS' RM in the same interview. "Adults need to create policies that can facilitate that overall social change."

A lot of that pressure for idols' constant perfection comes from the all-powerful record labels. But sometimes, it can come from fans as well.

"I turn to fans when I need support, and that always boosts my confidence. But there are fans who take it to the next level, who are too obsessive with idol groups because they love them so much they become controlling," Woo said. "When a member messes up or has a [17]scandal 醜行, so much [18]bashing 針對 has a negative effect."

The kind of [19]harassment 騷擾行為 that can happen in K-pop has echoes of the same [20]stalking 非法跟縱騷擾 dynamics, both physical and digital, that many young people, especially young women, face in their own lives.

"There are a lot of instances of people [21]hunting down 搜捕 idols and breaking into their hotels. BTS had a plane delayed because people bought tickets just to take pictures of them," Moses said. "We saw how deteriorating that can be — there have been suicides. It's really sad."

Moses has had to moderate her own fandom to avoid harassment over the years. "You see a lot of [22]bullying 恃強欺弱者 for very stupid reasons," she said. "My friends and I used to do K-pop cosplay with elaborate outfits, and got bullying and death threats after someone posted pictures of us. We were like, 'OK, this has changed,' and we backed out of that scene."

Many fans were hurt and [23]disillusioned 理想破滅的 by the still-ongoing Burning Sun scandal, which implicated several high-profile K-pop stars and industry figures in a ring of sex crimes. Charges included [24]allegations 指責 of [25]clandestine 秘密的 videotapes of assaults at a popular Seoul nightclub shared over text-message chains.

The scale and [26]sordidness 骯髒 of that scandal pulled back the curtain on the image of squeaky-clean K-pop idols. But fans still wrestle with what that means for their own attachments and communities. At best, it opens up space for fans to be [27]candid 率直的 about their own experiences with abuse. But time will tell if the scene is ready to listen.

"We have so much emotionally wrapped up in music and a sense of belonging to it. This is a time where we need to be open and have space for dialogue," Ahn said. "Idols and celebrities need to

be held accountable," Ahn added, mentioning Seungri, a central figure charged in the Burning Sun scandal. "Idols have a tremendous amount of power and it's getting difficult to hold any [28]ambivalence 矛盾的. You can love an idol but have complicated feelings."

Underneath the shiny [29]veneer 鑲板 and tight corporate control of K-pop, profound pain and abuse can boil over. Now that K-pop has matured in the U.S., most in the scene are glad that more difficult conversations are finally happening in public. It's not as [30]jubilant 歡欣的 as a new Blackpink single, but it might be more important.

"I like the awareness that it's not just happiness and sunshine and rainbows," Ly said. "They're people too, they make mistakes and struggle. They're humans but very famous humans."

## Appendix E - Ethical Approval

**From:** **Hamish Chalmers** hamish.chalmers@education.ox.ac.uk
**Subject:** CUREC Ref: CIA-22HT-051
**Date:** 29 March 2022 at 2:32 PM
**To:** Polly Leung polly.leung@wolfson.ox.ac.uk
**Cc:** Elizabeth Wonnacott elizabeth.wonnacott@education.ox.ac.uk, Student CUREC student.curec@education.ox.ac.uk

HC

Dear Polly

Title: Show your true colours: Can the use of colour aid memory for Chinese speakers' English L2 vocabulary acquisition?

Ref: CIA-22HT-051

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethics approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse
http://www.nspcc.org.uk/globalassets/documents/information-service/factsheet-child-abuse-reporting-requirements-professionals.pdf

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Best wishes
Hamish Chalmers
Member of the DREC

Dr Hamish Chalmers
Departmental Lecturer in Applied Linguistics and SLA
15 Norham Gardens, Oxford, OX2 6PY
Tel: +44 (0)1865 284 091
Email: hamish.chalmers@education.ox.ac.uk

Director: The International Database of Education Systematic Reviews IDESR.org

**Recent publications**:
Jitpaisarnwattana, N. & **Chalmers**, H. (2022). Can I MOOC to Catch up? The Effects of Using an LMOOC as a Remedial Tool for EFL Students in Thailand. English Language Teaching, 15(1), 37-52. doi: 10.5539/elt.v15n1p37.
**Chalmers**, H. & Murphy, V. (2022). Multilingual Learners, linguistic pluralism and implications for education and research. In Macaro, E., & Woore, R. (Eds.) Debates in Second Language Education. Abingdon: Routledge.

## Appendix F - Information Sheet for Parents / Guardians

**UNIVERSITY OF OXFORD**
DEPARTMENT OF EDUCATION

15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865274024 Fax: +44(0)1865274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

# Show your true colours: Can the use of colour aid memory for Chinese speakers' English L2 vocabulary acquisition?

## INFORMATION SHEET FOR PARENTS / GUARDIANS

Ethics Approval Reference: CIA-22HT-051

We are inviting your child to join in a research study. I am Leung Ka Po, a MSc student in Applied Linguistics and Second Language Acquisition, Department of Education, University of Oxford, under the supervision of Dr Elizabeth Wonnacott.

In partnership with researchers at the University of Oxford, Wa Ying College has agreed to take part in a research study investigating how colour memory affects English vocabulary acquisition. We would like to invite your child to be part of this study. We very much hope you would like your child to be involved, but before you decide, it is important that you understand why the study is being done and what it will involve.

**What are we trying to find out?**
This study aims to examine how the use of colour memory affects vocabulary acquisition in English. We are recruiting English as a Foreign Language learners in secondary schools in Hong Kong. Research has shown that the use of colour may facilitate attention, retention and memory performance among adult learners of English as a second language, and we are interested in finding out whether the same results are observed in teenagers. By participating in this study, your child will contribute to our research project on this question. Given that vocabulary plays an important role in English learning, the results from this study may help your school to plan activities targeting vocabulary teaching.

More information about the study can be obtained by contacting the researcher.

**Why has my child been invited to take part?**
We are inviting your child to take part because he/she is a young person, aged 15, attending Wa Ying College.

**Does my child have to take part?**
No. You can ask questions about the study before deciding whether or not to allow your child to participate. If you do agree to participation, you may withdraw your child from the study at any time, without giving a reason and without penalty, by advising the school or researchers of this decision.

**What will happen if my child takes part?**
If you agree your child to take part, your child will be asked to complete three tasks: a pre-test of English vocabulary test, a self-learning task with a reading article, and a post-test of English vocabulary test. The

English vocabulary tests take approximately 10 minutes each. The self-learning task takes approximately 45 minutes. The participants will be asked to play a game for approximately 15 minutes after the self-learning task. The whole study may take approximately 1.5 hours. The exact time may vary, as the tests have not been taken by Chinese secondary students previously. The tasks will be completed as an individual and are student-friendly.

**What are the advantages / disadvantages of taking part?**

There is no intended benefit for the participants directly but the results of the study may benefit your child's school to plan activities targeting vocabulary teaching. There is no risk for your child in participating in the study.

**What happens to the results of the study?**

Results for each child are kept strictly confidential. Only numbers will be used to identify children, and all information and results are kept in a locked room or encrypted and password protected hard drive respectively. Third parties may be given access to research data for monitoring and/or audit of the study, or for data storage purposes. Regular summaries of our summaries of our findings will be given to the school and will be available to interested families.

Any information your child provides during the study is the research data. No personal data about your child will be stored beyond the duration of the study. Opt-out forms will be retained by the school for the duration of the study, and for as long as the school determines appropriate after research activities have concluded at the school.

**Will the research be published?**

The research will be written up as a thesis. On successful submission of the thesis, it may be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible. The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research. In addition, the research may be submitted for publication in a research journal.

All research data and records will be stored for a minimum retention period of 3 years after publication or public release of the work of the research. In addition, some processed data may also be shared with other researchers via websites such as the Open Science Framework. It will not be possible to identify your child's data in this dataset (no personal data will be stored).

**Who is conducting this research?**

The research is organised by Leung Ka Po, who is a MSc student in Applied Linguistics and Second Language Acquisition, Department of Education, University of Oxford, under the supervision of Dr Elizabeth Wonnacott.

**Ethics**

This study has been reviewed by, and received ethics clearance through, the University of Oxford's Central University Research Ethics Committee, CIA-22HT-051.

**What if there is a problem?**

If you have a concern about any aspect of this study, please contact Leung Ka Po at +44(0)1865274024 / polly.leung@wolfson.ox.ac.uk or Dr Elizabeth Wonnacott at +44(0)1865274024 / elizabeth.wonnacott@education.ox.ac.uk, and we will do our best to answer your query. I/we will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee**; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

**Data Protection**

The University of Oxford is the data controller with respect to any personal data collected, and as such will determine how this personal data is used in the study.
The University will only process personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest.
Further information about your rights with respect to personal data is available from http://www.admin.ox.ac.uk/councilsec/compliance/gdpr/individualrights/.

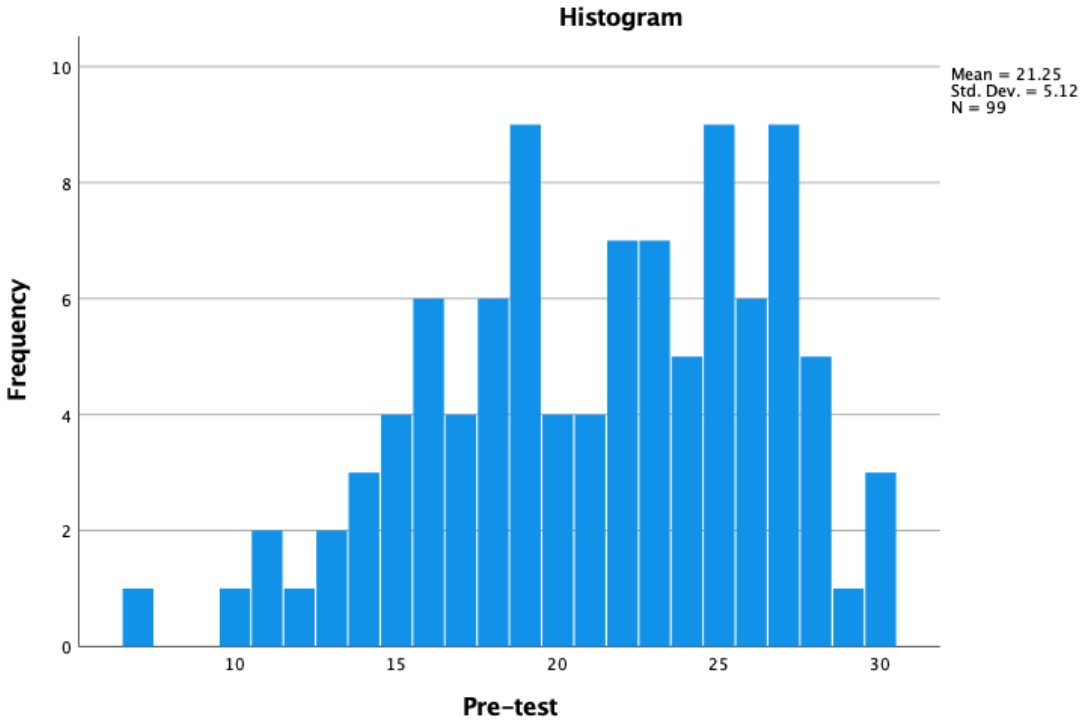
**What should I do next?**

Please fill in the enclosed form and return it to your child's class teacher if you would not like your child to take part in this study. Please remember that you may withdraw your child at any time, without penalty and without giving a reason, by notifying the researcher.

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Leung Ka Po
Department of Education
15 Norham Gardens, Oxford
UK, OX2 6PY
University tel: +44(0)1865274024
University email:

## Appendix G - Histogram of pre-test




## Appendix H - Histogram of post-test