THE PRODUCTION EFFECT IN CHILDREN LEARNING MODERN STANDARD ARABIC VOCABULARY AS AN ADDITIONAL LANGUAGE

BY

NOUF ALI ALHARBI

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF BIRMINGHAM

Department of English Language and Linguistics

School of English, Drama and Creative Studies

College of Arts and Law

University of Birmingham

August 2024

Abstract

This thesis investigates the production effect (PE), a phenomenon where actively producing words (reading aloud, writing) leads to better learning compared to passive exposure (reading silently). While extensively studied in adults, the PE's impact on children's foreign language acquisition remains unclear. This research explores the influence of the PE on noun acquisition in child learners of Modern Standard Arabic (MSA).

A pilot study examined the Production Effect (PE) in 72 bilingual children aged 4-11 with basic Modern Standard Arabic (MSA) knowledge. Participants engaged in an online MSA word-learning task under two conditions: listening only and listening then repeating. Testing occurred immediately, 24 hours, and nine weeks after learning. This pilot study highlighted the need to control for age and language background in subsequent experiments. It also suggested exploring a broader range of learning conditions to capture potential differences in retention better. These insights informed the design adaptations made in Experiments 1 and 2.

Experiment 1 explored the PE using a larger set of 64 MSA words while controlling for age and language background. Participants (6-year-old bilingual children) learned new words in either listen-only or repeat conditions. Testing occurred immediately, one week later, and two weeks later. Delayed tests revealed an overall improvement in performance across both groups. However, the experiment did not show a significant effect of production.

Experiment 2 expanded the investigation by including a writing condition alongside listening and repeating. This experiment focused on 8-year-old bilingual children. Testing was conducted immediately after learning, one week later, and two weeks later. The writing condition significantly enhanced retention and recall, particularly in delayed tests. These findings suggest that the PE's learning effects on children's MSA word acquisition are moderated by various factors, including age of the learners, the specific learning materials and the length of time between learning and testing.

Overall, this thesis contributes to understanding the PE in children's additional language learning. The results highlight the potential benefits of writing as a production mechanism in enhancing vocabulary learning in MSA children. The research also identifies factors influencing the effectiveness of the PE, paving the way for further investigation in the field of learning mechanisms and the development of more effective pedagogical approaches.

Dedication

To my beloved children, Mays, Qusai, and Mishary, and to the cherished memory of Setti Khadijah and Aunt Alia'a (May Allah grant them peace)

Acknowledgments

In the name of Allah, the Most Gracious, the Most Merciful, I express my deepest gratitude. Thanks to the Ministry of Education (Saudi Arabia) and the Saudi Arabian Cultural Bureau (London) for the scholarship and financial support that empowered this PhD journey.

I would like to thank my esteemed supervisors; Dr. Gareth Carrol, for his unwavering support and mentorship, Prof. Jeannette Littlemore for her invaluable expertise and feedback, and Dr. Bene Bassetti for her prior guidance, and teaching opportunities she offered me in her lectures.

My deepest gratitude goes to Sara Orfali for her captivating illustrations for the children in this thesis. I am also immensely thankful to the study participants: the parents for their trust, the children for their time, and the Amanah Muath Trust for facilitating data collection. Special thanks to Mr. Faisal, Ms. Muna, and the volunteer teachers for their invaluable support.

A heartfelt thanks to my amazing colleagues and friends for their endless support, stimulating discussions, and PhD journey camaraderie.

My deepest thanks to my parents, siblings, and entire family for their unwavering support throughout my PhD journey. A special thank you to my amazing children for your strength, belief, and endless love that made this possible.

Thank you all!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

PE – Production Effect

MSA - Modern Standard Arabic

L2 - Second Language

L1 - First Language

lmer - Linear Mixed-Effects Model

RT - Reaction Time

GLMM - Generalized Linear Mixed-Effects Model

ANOVA - Analysis of Variance

SE - Standard Error

AoA - Age of Acquisition

IPA - International Phonetic Alphabet

WM - Working Memory

STM - short term memory

LTM - Long term memory

CLT - Cognitive Load Theory

CTML - Cognitive Theory of Multimedia Learning

Index of Terms

- In order to convey the various stages of language acquisition and the factors that influence the learning process, it is essential to distinguish between the terms L1, L2, and additional language. L1, or First Language, refers to any language children typically acquire naturally from their environment during childhood. This process is effortless and occurs without formal instruction, allowing children to assimilate their mother tongue quickly. However, in many parts of the world, children grow up bilingual or multilingual, simultaneously acquiring multiple first languages (L1s). The second language, or L2, is any language learned after the first. Acquiring a second language usually requires dedication and effort, often necessitating immersion in an environment rich in the target language, either through formal education or informally in a community where the language is spoken by its native speakers. This process is typically less intuitive than the acquisition of L1. The term additional language can be used interchangeably with a second language and includes learning a third or subsequent language under conditions similar to those for learning a second language. This expands the learner's linguistic repertoire beyond their native and second languages.
- The connection between Classical Arabic, Fus'ha, Modern Standard Arabic (MSA), and colloquial dialects is integral to understanding the linguistic landscape of the Arab world. Classical Arabic, the language of the Quran and early Islamic literature, serves as the historical and linguistic foundation of Fus'ha and MSA. Fus'ha, often used interchangeably with MSA, encompasses the formal written and spoken Arabic used in official documents, education, and media. While Classical Arabic remains essential for religious and historical texts, MSA has evolved to accommodate contemporary needs, incorporating modern terminology and simplifying some grammatical structures. In contrast, colloquial dialects, which vary significantly across regions, are the everyday spoken languages of Arabic speakers. These dialects are not standardized and differ widely in vocabulary, pronunciation, and syntax. Despite these differences, MSA acts as a unifying standard, facilitating communication across diverse Arabic-speaking communities. The interaction between these forms of Arabic showcases the rich linguistic heritage and the practical adaptations necessary for modern communication.

Chapter 1: The Introduction

This thesis examines the role that production plays in additional language learning. Therefore, this first chapter introduces the core concepts of this study and provides an overview of the content. Initially, a general overview of the theory under investigation (i.e., The Production Effect) is presented. Then, the research gap and the purpose of the study are explained. The chapter concludes with an overview of the chapters included in this thesis.

1.1. Study Rationale: Investigating the Role of Active Production in Enhancing Second Language Learning in Children

1.1.1. A Brief Introduction to the Production Effect and its Mechanisms

Vocabulary learning is a multifaceted process that involves the utilization of various memory strategies to enhance the acquisition, retention, and retrieval of new words (Nation, 2001). These strategies often draw upon cognitive processes, such as encoding and retrieval, that support long-term retention of vocabulary. Within this framework, active production has gained attention for its role in improving memory performance. This section will introduce a general overview of the investigated theory (namely, the Production Effect), further discussing its modalities, mechanisms, applications, and implications in the following chapter. The production Effect (hereafter PE) is an umbrella theory that refers to the memory benefit from “producing” or vocalizing words aloud (MacLeod et al., 2010; Ozubko et al., 2012). The theory of the PE can be traced back to early studies by cognitive psychologists exploring how memory encoding and retrieval work. This theory is based on the idea that vocalizing information creates a unique memory trace in the phonological loop, the part of working memory that deals with spoken language. This rehearsal process enhances the encoding process and makes it easier to recall the information later. As early as 1928, Barlow found that studying nonwords aloud improved recall compared to silently reading words. This finding has been replicated in more recent elsewhere literature, where studies have shown a memory advantage for words and nonwords spoken aloud compared to those learned silently, suggesting that vocalization enhances memory encoding compared to silent learning for adult participants. (Kurtz & Hovland, 1953; Hopkins & Edwards, 1972; Conway & Gathercole, 1987; Gathercole & Conway, 1988; Dodson & Schacter, 2001). After ten years, interest in the production effect resurfaced in 2010 at work by MacLeod et al. In recent years, the advantage of PE has shown to be extended to various forms of production, including mouthing, whispering, singing, typing, and writing (Forrin et al., 2012; Jamieson & Spear, 2014; MacLeod et al., 2010).

Despite a focus on adults in production effect (PE) research, a limited number of studies (e.g., Icht & Mama, 2015; Zamuner et al., 2018) have investigated children using similar conditions (looking, listening, production). These studies reveal intriguing age-related variability in the PE. While Icht and Mama (2015) observed a positive PE across tasks, Zamuner et al. (2018) found that younger children (4.5 years) exhibited a reversed PE. In contrast, older children (6 years) positively affected by hearing and producing words. This trend is further supported by studies reporting positive PE in children aged 7-10 (Pritchard et al., 2019) and an age-dependent shift from reversed to positive PE (López Assef et al., 2021). These findings suggest that the mechanisms underlying vocabulary learning through production may evolve with age, potentially reflecting cognitive processing, memory capacity, or language development changes. Understanding these developmental shifts is crucial for elucidating the cognitive mechanisms that drive vocabulary acquisition in children and how they differ from adults. Research on the production effect generally suggests that pronouncing words aloud during encoding enhances their distinctiveness in memory, potentially leading to better retention in memory tests, even after longer intervals, and facilitating second-language learning (Ozubko et al., 2012; MacLeod & Bodner, 2017; Icht & Mama, 2019). However, the findings are inconsistent across studies involving adults and children (e.g., Icht & Mama, 2019). This variability points to the complexity of the mechanisms involved in vocabulary learning through production. Factors such as the depth of processing, the integration of phonological and semantic information, and the role of motor processes in speech production may all contribute to the PE. Further research is needed to disentangle these mechanisms and understand how they interact with developmental stages, individual differences, and specific learning contexts. Such insights could inform more effective vocabulary learning strategies tailored to different age groups and learning needs.

According to the distinctiveness account of the production effect, vocalized words receive deeper encoding in memory because they differ in processing compared to silently read words. As Hunt (2013, p.10) defines it, distinctiveness is the "processing of difference in the context of similarity." In this context, vocalizing a word creates a distinct processing experience compared to silent reading, even if multiple words are vocalized. Learners who study a word list with half the words spoken aloud and the other half read silently should exhibit better recall for the vocalized words due to this enhanced distinctiveness at encoding.

While the production effect often benefits memory, it can be disrupted under certain conditions. Research suggests that encountering unfamiliar or non-native sounds during vocalization (e.g., novel words with foreign phonemes) can weaken or even reverse the production effect (Kaushanskaya & Yoo, 2011; Cho & Feldman, 2013, 2016; Dahlen & Caldwell-Harris, 2013; Baese-Berk & Samuel, 2016; Zamuner et al., 2016). To illustrate, the study of Kaushanskaya and Yoo's (2011) with pseudowords containing foreign phonemes demonstrates this. Participants struggled to recall non-native phonemes when spoken aloud compared to silent reading. This aligns with the working memory model by Baddeley and Hitch (1974), where the phonological loop facilitates the processing of familiar sounds. Non-native sounds, lacking strong memory links, may overload the phonological loop, weakening the benefit of vocalization. This suggests that the production effect relies on some level of familiarity or ease of processing during vocalization. A detailed exploration of these specific mechanisms will be addressed later in the literature review (Chapter 2).

Drawing upon the observations made earlier, there are four possible effects for production on vocabulary learning as seen in the adult's and children's literature:

1. Positive effect: This occurs when learners actively produce words (e.g., saying them aloud) that contain familiar sounds (native phonemes) during learning. This reinforces memory through the production effect and strengthens memory encoding due to distinctiveness.
2. Reversed effect: This arises when learners attempt to vocalize words with unfamiliar sounds (non-native phonemes). The unfamiliar sounds can overload the phonological loop in working memory, hindering the production effect and potentially leading to poorer recall than silent reading.
3. Attenuated effect: This describes a situation where production still has some benefit on memory, but the advantage is reduced. This might occur with moderately difficult-to-pronounce words or when learners encounter partially familiar sound patterns in a new language. For instance, in language learning, when learners encounter words that contain a mix of familiar and unfamiliar sounds, the cognitive load of processing these combinations may diminish the advantages of production. For intermediate learners, words that are neither completely foreign nor entirely familiar in their phonetic

structure can lead to an attenuated production effect, as their cognitive resources are divided between recognizing familiar elements and grappling with unfamiliar ones.

4. Null effect: In some cases, production may not show any significant difference in memory compared to silent reading. This could happen when the learning task is very simple, the learner is highly proficient in the language, or other factors unrelated to production come into play.

The literature reviewed in this section shows varied outcomes regarding the role of production in vocabulary learning and retention, particularly in second-language contexts. Some studies demonstrate that active production can enhance memory and recall, yet others report limited or inconsistent effects, especially over longer retention intervals. This divergence highlights an unresolved question about the effectiveness of production-based strategies for sustaining vocabulary retention in children learning a second language. To address this gap, this thesis investigates how production impacts vocabulary learning by assessing both immediate and delayed recall and recognition. It aims to clarify its role in supporting durable vocabulary retention among young language learners.

1.1.2. Child Language Acquisition

The process of learning vocabulary in a child's first language (L1) and second language (L2) is influenced by various factors. Factors such as exposure to language input, frequency of input, and vocabulary acquisition strategies play significant roles in vocabulary acquisition (Vermeer, 2001; Huckin & Coady, 1999; Aljadani, 2020; Radić-Bojanić, 2021). Access to native speaker input can affect the speed of language acquisition for young learners, especially concerning the oral and aural aspects (Andari, 2023). In addition, incidental and intentional vocabulary acquisition and vocabulary strategy use can negatively influence L2 vocabulary acquisition because they may lead to inconsistent or fragmented learning experiences, which can hinder the overall retention and application of new vocabulary (Alemi & Tayebi, 2011).

The vocabulary of a child's first language (L1) develops naturally through exposure and interaction with their environment. On the other hand, learning a second language (L2) can be influenced by factors like the child's age, L1 vocabulary, and the similarity between L1 and L2. For instance, a child with a strong L1 vocabulary and learning an L2 with many cognates (words with similar forms and meanings) will likely have an easier time expanding their L2 vocabulary (Tonzar et al., 2009). For instance, in Classical Arabic acquisition, research suggests a clear distinction between learning spoken dialects and Classical Arabic itself (Alzu'bi et al., 2023) and challenges the traditional view of modern dialects as a direct evolution of Classical Arabic. Their findings indicate that these dialects coexisted with Ancient Arabic rather than being a later development.

Research indicates that bilingual children often exhibit a combined vocabulary that surpasses that of monolingual children, highlighting the role of conceptual knowledge in their first language (L1) as a scaffold for acquiring vocabulary in their second language (L2) (Scheele et al., 2009). This foundational knowledge in L1 not only aids in L2 vocabulary learning but also underscores the interconnectedness between the two languages. Berghe et al. (2021) further emphasize that children with larger L1 vocabularies are likely to benefit more from L2 vocabulary interventions, demonstrating that the effectiveness of L2 vocabulary learning strategies is closely tied to L1 vocabulary size. Bilingual children's exposure to two languages allows them to draw from a broader linguistic repertoire, enhancing their vocabulary skills beyond what is typically observed in monolingual children. The cross-linguistic influences facilitated by bilingualism contribute to an enriched vocabulary repertoire and overall language proficiency, underscoring the advantages that bilingualism offers in vocabulary development.

Understanding the intricate relationship between a child's first language (L1) and second language (L2) is crucial for optimizing vocabulary acquisition in bilingual toddlers. The phonological overlap between L1 and L2 languages can significantly impact vocabulary production, as demonstrated by Barachetti et al. (2021). By examining the impact of cross-language phonological overlap on bilingual and monolingual toddlers' word recognition, Holzen et al. (2018) demonstrated that words with greater phonological similarity between L1 and L2 were more easily recognized through eye-tracking experiments. These findings underscore the importance of understanding how phonological similarities between languages and the depth of vocabulary knowledge in L1 can shape bilingual children's vocabulary acquisition and learning experiences. Bilingual children with a more extensive vocabulary can better apply their L1 and L2 phonological knowledge, which would facilitate word learning and, therefore, impact word recognition (Holzen et al., 2018). Learners with substantial L1 vocabulary knowledge can rely on semantic or phonological similarities between novel words and words they have already learned. In contrast, novice learners cannot effectively utilize existing conceptual knowledge to facilitate new word learning and thus have to rely more on their phonological memory to establish form-meaning links for newly learned words (Berghe et al., 2021). This advantage stems from the rich conceptual and lexical networks established in their first language (L1), which serve as a foundation for learning new words in a second language (L2). Several key factors influence the development of the connection between L1 and L2 vocabulary. Strong social support at home can lead to a more extensive vocabulary in L1, thereby enhancing children's ability to learn new words in both L1 and L2 (Kan & Kohnert, 2011). The impact of the home environment extends beyond vocabulary acquisition; the relationship between L1 and L2 vocabulary skills can also influence reading comprehension development in bilingual children (Bosch & Segers, 2020). To leverage this L1-L2 connection, educators and researchers have identified effective strategies for promoting cross-language transfer and vocabulary development. These include using gestures and pictures and providing meaningful social and academic contexts in both languages (Andrä et al., 2020; Pham et al., 2017). Such approaches facilitate mapping new L2 words onto existing L1 concepts, enhancing overall vocabulary growth. The quantity and quality of language exposure play a crucial role in this process. Studies have shown that the amount of exposure to L1 and L2 at home can predict the vocabulary skills of sequential bilingual children (Cheung et al., 2018). This finding underscores the importance of rich language input in both languages for optimal bilingual vocabulary development.

As discussed earlier, phonological short-term memory plays a significant role in L1 and L2 vocabulary learning (Farnia & Geva, 2011). Moreover, learners with larger vocabularies tend to have better phonological acquisition in L2, emphasizing the role of lexical development in L2 learning (Bundgaard-Nielsen et al., 2011). In conclusion, the interplay between L1 and L2 vocabulary development is intricate, with factors like phonological overlap, social support, cognitive-linguistic skills, and exposure to language input playing crucial roles in children's vocabulary acquisition in both languages.

1.2. The Research Gap

Although a few studies have investigated the Production Effect (PE) in children's first language (L1) learning, there remains a critical gap in understanding its implications for second language (L2) acquisition, particularly in the context of Modern Standard Arabic (MSA). Existing research has primarily focused on short-term memory effects, demonstrating that active production can enhance recall for familiar and novel L1 words. However, little is known about the long-term retention of L2 vocabulary acquired through production, with no studies examining the durability of these effects beyond immediate testing. Moreover, the comparative impact of different production modalities (e.g., spoken versus written production) on memory retention remains largely unexplored in children. To address these gaps, this thesis incorporates delayed recall and recognition assessments, alongside an investigation of varied production methods, to evaluate how PE facilitates long-term vocabulary retention in child learners of MSA.

The current study examines two groups of children in the pilot study (Chapter 3), who primarily speak Modern Standard Arabic (MSA) as a second language (L2), with their first language (L1) being a colloquial Arabic dialect. This linguistic environment exemplifies the diglossic nature of Arabic, where MSA serves as the formal, high-prestige variety, typically acquired through formal education, while the local dialect is learned naturally in everyday communication. The participants in this pilot study are introduced to MSA words with familiar phonemes, focusing specifically on learners of Arabic, a population that has yet to be extensively explored in psycholinguistic research. The pilot study features a short list of MSA words to establish a baseline understanding of phoneme familiarity.

In contrast, Studies 1 and 2 shift focus on multilingual populations, where English is the dominant language. These children navigate a more complex linguistic landscape, with both MSA and their dialect existing alongside English. In these studies, participants are introduced to longer learning lists, and recognition tests are used to assess the influence of lexical competitors (i.e., similar-sounding words). Experiment 2 further innovates by incorporating writing as a production condition, alongside listening and repeating, to examine how different production methods affect vocabulary acquisition and recognition. The research investigates how these multilingual children acquire and process vocabulary in MSA, considering their prior exposure to multiple languages and the interaction between them. Each experiment's methodology is detailed in the corresponding chapters (Chapters 3, 4, and 5), highlighting these novel approaches to understanding MSA vocabulary learning within diverse linguistic contexts.

In conclusion, vocabulary acquisition in second language learners represents a multifaceted cognitive process fraught with challenges such as lexical attrition or word forgetting (Allen, 1999; Sharma, 2022). Various strategies have been developed to enhance retention, including oral and visual engagement during learning (Daniels, 2000; Hulstijn, 2001) and auditory input (Radojevic, 2006). Repetition and retrieval practice is critical to solidifying vocabulary knowledge (Yang, 2022). However, understanding the cognitive processes underlying word learning and memory remains essential. This thesis investigates the hypothesis that production, as a cognitive mechanism, facilitates vocabulary acquisition and enhances the durability of learning among young second-language learners. This research aims to develop a comprehensive framework for improving vocabulary learning and retention by examining critical cognitive and instructional strategies, such as repeating words aloud. The findings seek to enrich our understanding of how young learners acquire and retain vocabulary in a second language, providing valuable insights into the mechanisms of L2 learning among children.

1.3. The purpose of the study

In an attempt to contribute to studies of the Production Effect in children, this thesis aims to examine the role that production plays in the process of second language learning, primarily how production affects Modern Standard Arabic learning during childhood. Despite the existing literature indicating a potential connection between production and adult L2 learning, research into the effect of production on the durability of learning in child learners remains unexplored in the existing literature. Specifically, A critical examination of the current literature reveals several gaps and inconsistencies in the understanding of production's effect on learning durability, particularly in the context of young second language learners. Thus far, only four studies have examined the production effect in children, but none have tested the durability of learning beyond the immediate test (i.e., using a delayed test). This thesis addresses the existing gap by including delayed tests and combining two memory tests (i.e., recall and recognition) to evaluate the participants' long-term retention and depth of learning. Therefore, a combination of different learning strategies was implemented to provide a clear understanding of the effect of production on L2 child learners, including listening, repeating, and writing.

1.4. Outline of the thesis

This thesis is organised as follows: Chapter 2 provides a general literature review to summarise key elements of vocabulary learning and the studies that have investigated the Production Effect on vocabulary learning. The chapter discusses vocabulary acquisition theories and vocabulary learning methods. The focus is on the Production Effect and how studies have looked at it in adults and children. This helps to provide a general grounding for developing the following empirical work. The acquisition of Modern Standard Arabic is also discussed as it is used as the L2 under investigation in the thesis.

Each empirical chapter is preceded by a short abstract and an introduction to contextualize it within the broader discourse of the thesis and link it to other studies. Chapter 3, the first empirical chapter, presents a pilot study that explores the acquisition of nouns in Modern Standard Arabic (MSA) and investigates the potential impact of production on language learning. Child learners of Arabic (aged four to eleven) participated in a word learning online task designed to examine the effects of different learning conditions: listening only and listening then repeating. A recognition test was employed to assess accuracy and to gain insight into hesitancy and competition in the responses across three-time points: an immediate test, a 24-hour delayed test, and a nine-week delayed test. The study compares the effectiveness of listening only versus listening and repeating in word learning. Additionally, it examines the influence of age and language background on accuracy and response times to correct answers. This investigation contributes to our understanding of the role of production in vocabulary learning and retention among young learners of MSA.

Chapter 4 presents Experiment 1, which offers a modified version of the pilot study. The participants' age and language background were controlled, and an extended learning list of 64 words was divided into two lists based on the condition (listen and repeat). The number of items was refined to increase the likelihood of items being unknown. Lexical competitors (words that sound similar to or have similar meanings to the target word) were also included in the testing phase. This study aims to investigate the influence of word similarity on retrieval and determine whether these competitors would show a production bias, where participants choose phonological or semantic competitors over the target words. To test the effect of PE on the durability of learning, a free recall test and a recognition test were used across three test times (an immediate test, a one-week delayed test, and a two-week delayed test). Results are discussed in terms of the differential impact of testing conditions, delay intervals, and distractor types on the efficacy of the Production Effect in memory retention.

Chapter 5 (i.e., Experiment 2) extends the previous investigations of the PE by examining writing as a production condition with a focus on eight-year-old children. The participants were early elementary school children who can write complex sentences in their L1 and are confident to apply critical writing skills in their L2 writing (i.e., analysing a simple word's phonemes when hearing and writing it). The aim was to re-examine the typical PE conditions (i.e., listen and repeat) with the current age group, with an additional type of production introduced (i.e., writing). Results are discussed regarding the durability of learning of L2 children across three testing periods (immediate, one-week, and two-week).

Chapter 6 synthesizes and reviews the findings from the three conducted experiments exploring the Production Effect (PE) in language learning. It analyses how PE manifests under different conditions (positive and reversed effects) for children acquiring Modern Standard Arabic (MSA) vocabulary. Further, it investigates writing as a form of production and its potential to enhance learning. The chapter then transitions to a critical review of existing PE literature, examining how the current thesis' findings align with or diverge from the current understanding of PE. Finally, the chapter identifies thesis limitations and proposes avenues for further research. It outlines possible further investigations into the intricate dynamics that underlie the Production Effect.

Chapter 7 presents the concluding remarks, and it gives a concise summary of the studies' discoveries.

Chapter 2. Literature Review

This chapter provides a broad foundation for the following empirical chapters. It reviews the literature on the critical aspect of the production effect, as it is the theory under investigation concerning vocabulary learning. A few remarks about memory and language learning are needed to provide a theoretical base to describe some terms used throughout the thesis. I then review the literature on the key concepts and factors affecting language learning, focusing on vocabulary learning, mainly existing work on child vocabulary learning. It should be noted that while bilingualism is not the primary focus of the present study, the participants are multilingual (i.e., speaking English and additional languages). This multilingual background, with English as a dominant language, plays a significant role in understanding how the participants acquire and process Modern Standard Arabic vocabulary. Hence, I include an overview of bilingualism and bilingual child learners as they are the investigated group. As this study examines the effect of production on acquiring MSA as an additional language, an overview of MSA and its characteristics is also discussed. Specific overviews related to individual studies will be addressed in the appropriate chapters. The final section includes the thesis objectives and the research questions.

2.1. Vocabulary Learning: Concepts, Factors, and Methods

2.1.1. Key Concepts of Learning and Acquisition

In language research, the terms "vocabulary learning" and "vocabulary acquisition" are frequently used interchangeably, with "vocabulary learning" being more prevalent in contemporary scholarship (Laufer, 2003; Kersten, 2010). Vocabulary learning encompasses two fundamental approaches: incidental and intentional learning. Incidental vocabulary learning refers to acquiring new lexical items through exposure to language in context, without the learner consciously aiming to study or memorize specific words (Pellicer-Sánchez, 2016; Spivey & Cardon, 2015). This process is typically seen in activities such as reading or listening, where learners encounter new words and internalize their meanings through repeated contextual exposure.

Conversely, intentional vocabulary learning is characterized by deliberate and structured efforts to commit words to memory. This often involves explicit instruction and pedagogical strategies, such as word lists, flashcards, or focused learning tasks to enhance vocabulary retention (Nation, 2001; Webb, 2007). While both incidental and intentional learning processes contribute to vocabulary development, they represent distinct pathways to word acquisition, each with its benefits and challenges.

Incidental learning significantly broadens vocabulary knowledge, mainly through exposure to language in meaningful contexts. For instance, research by Pellicer-Sánchez (2016) underscores the efficacy of reading for incidental vocabulary learning, wherein learners acquire new words without a targeted focus on memorization. On the other hand, intentional vocabulary learning provides a framework for deeper engagement with lexical items, enabling learners to actively process and retain new words more effectively (Guo, 2021). This is especially critical in second language (L2) learning contexts, where intentional study supports the development of more complex vocabulary, which learners might not acquire solely through incidental exposure.

The interaction between these two approaches is complementary and mutually reinforcing. Research suggests that explicit instruction can enhance incidental learning outcomes by providing learners with strategies to process and retain lexical items more efficiently (Gabay et al., 2015; Bisson et al., 2015). In this sense, the two learning processes are not mutually exclusive but operate to promote a more comprehensive and nuanced approach to vocabulary acquisition (Csizér & Kormos, 2009; Dunn & Iwaniec, 2021). The integration of incidental exposure and intentional study facilitates both breadth of vocabulary—through the passive encounter of words—and depth of knowledge through active and focused engagement with those words.

A pertinent illustration of the link between incidental learning and cognitive mechanisms is the production effect. This phenomenon, whereby words spoken aloud are more likely remembered than those read silently, provides evidence of incidental learning in action (MacLeod et al., 2010). Although the production effect involves a conscious activity—speaking aloud—it often results in unintended improvements in word retention. Learners may not set out to memorize words by merely producing them. However, the cognitive engagement triggered by this action enhances memory retention, aligning it with the principles of incidental learning. Thus, the production effect exemplifies how incidental learning can occur naturally through active involvement with language without deliberate memorization efforts.

Balancing incidental and intentional learning strategies has been widely supported in the literature, as both methods contribute to overall vocabulary acquisition. While incidental learning fosters natural language development through repeated exposure in context, intentional learning promotes the conscious, structured acquisition of vocabulary. Effective language learners often employ both strategies, leveraging incidental exposure to reinforce vocabulary encountered in everyday interactions while using intentional study to ensure long-term retention and deeper lexical understanding (Schmitt, 2008; Sok & Han, 2020). The interaction between these approaches is pivotal for facilitating optimal lexical development, particularly within the domain of second language (L2) acquisition, where both incidental exposure and intentional focus contribute to the successful internalization and retrieval of lexical items, where learners benefit from structured learning to navigate the complexities of a foreign language (Gasparini, 2004; Hunt & Beglar, 2005).

In conclusion, integrating incidental and intentional vocabulary learning strategies is paramount to achieving comprehensive vocabulary acquisition. The dynamic interplay between these approaches, where explicit instruction enhances incidental learning and incidental exposure supports intentional study, highlights the importance of a mixed-method approach. Learners who engage in both types of learning are likely to develop more robust language skills, as each method offers unique advantages that contribute to a well-rounded vocabulary acquisition process.

2.1.2. Factors Affecting Vocabulary Learning

2.1.2.1. Processing of Information

Information processing is crucial in vocabulary acquisition, and several theories provide insights into how learners can enhance their retention and recall of new words. A well-developed vocabulary is essential to language learning and is influenced by various cognitive and social factors (Gardner & Maclntyre, 1992; Block, 2003; Atkinson, 2011; Bai, 2018). Cognitive factors, in particular, play a key role in facilitating vocabulary acquisition (Yaqubi et al., 2012; Namkung & Fuchs, 2015; Schneider & Niklas, 2017; Hackle, 2018).

Memory is integral to recalling new words, while attention is necessary to focus on language input (Dagenbach & Carr, 1994; Williams, 1999; Ellis, 2001; Robinson, 2003; Zhang, 2019). Additionally, processing speed—the time a learner takes to process linguistic information—significantly impacts vocabulary learning (Kail & Salthouse, 1994; Deary, 2012). Efficient processing allows learners to quickly react to and understand auditory and visual information, facilitating vocabulary comprehension (Marchman & Fernald, 2008; Tunmar & Hoover, 2017). Research shows that as children's processing speed increases, it directly influences working memory capacity, which is essential for vocabulary learning (Fry & Hale, 1996). Working memory allows learners to temporarily store and manipulate new linguistic information, enabling them to process word meanings and integrate them into their knowledge base (Cowan, 2010, 2014; Cockcroft, 2015). Faster processing also supports more effective mnemonic strategies, such as imagery or semantic encoding, further enhancing vocabulary retention (Smith & Border, 2019; Blunt & VanArsdall, 2021).

The Depth of Processing Hypothesis, introduced by Craik and Lockhart (1972), posits that the depth at which information is processed affects memory retention. Deeper cognitive engagement with words—such as focusing on their meanings—leads to more muscular memory retention than shallow processing tasks (Craik & Lockhart, 1972). However, while this theory emphasizes deeper engagement, it has limitations. For example, it may not fully account for individual differences in learning, such as learners with slower processing speeds or less working memory capacity who may struggle with tasks requiring deep semantic engagement. Furthermore, this theory does not explicitly address the role of production in learning, which leads us to the Production Effect theory.

The Production Effect refers to the phenomenon whereby words read aloud are remembered better than those read silently (MacLeod et al., 2010). This theory emphasizes producing language, such as reading aloud or repeating words, as a method of encoding, which enhances recall. When linked with the Depth of Processing Hypothesis, producing a word aloud can be seen as a form of deeper engagement, as it involves both semantic and motor processing. This additional processing layer creates more substantial memory traces, as the cognitive effort involved in production adds another dimension to learning (Ozubko et al., 2012). This aligns with the view that engaging multiple cognitive pathways, such as auditory, visual, and motor, optimizes vocabulary retention.

Furthermore, the Transfer Appropriate Processing (TAP) theory builds on depth by asserting that memory retrieval is most successful when the processes used during learning are aligned with those required during recall (Morris et al., 1977). For example, if semantic processing is employed during word learning, the same type of processing will be more effective during recall. This theory directly relates to the Production Effect, as producing a word during encoding involves a specific form of cognitive and motor engagement, which can be mirrored during retrieval. Thus, the Production Effect enhances retrieval when the recall task involves spoken or active word reproduction.

The TOPRA model (Barcroft, 2015) provides further insight into lexical input processing by emphasizing a balance between semantic, structural, and mapping processes in vocabulary acquisition. According to TOPRA, focusing heavily on one type of processing, such as semantic tasks, may limit cognitive resources available to others, potentially affecting overall learning efficiency (Barcroft, 2015). This model highlights the importance of varied and context-rich learning experiences, which simultaneously engage multiple processing layers to optimize vocabulary acquisition. Here again, the Production Effect plays a critical role, as it suggests that engaging in production activities (such as reading aloud or speaking) facilitates a balance between processing semantic meaning and other cognitive tasks, like auditory and motor engagement.

Processing speed also plays a critical role in these cognitive frameworks. Faster cognitive processing allows learners to engage more efficiently with deep processing tasks, such as those emphasized in the Depth of Processing Hypothesis and TAP, and to retrieve information more effectively. Similarly, rapid processing enables learners to balance different cognitive tasks, as the TOPRA model proposes, facilitating semantic engagement and structural understanding. In the context of the Production Effect, faster processing speeds may enhance learners' ability to engage in repeated production tasks, thereby improving memory retention and recall of vocabulary.

In summary, understanding vocabulary acquisition through the Depth of Processing Hypothesis, TAP, the Production Effect, and the TOPRA model provides valuable insights into optimizing learning strategies. While engaging deeply with meanings improves retention, aligning cognitive processes during encoding and retrieval maximizes recall efficiency. Although the Production Effect enhances encoding and recall through active production, it does not inherently focus on meaning and, in some respects, represents a more surface-level processing activity. In contrast, the TOPRA model advocates for a balanced approach, integrating different processing types to maximize learning outcomes. Together, these frameworks illustrate how various cognitive processes, including processing speed and the level of engagement with content, shape vocabulary acquisition outcomes.

2.1.2.2. Word Frequency and Familiarity

Another factor influencing L2 word learning is word frequency. Frequency represents the usage rate of a word in a language. This affects a language user's ability to recognize, understand, or use the word, its pronunciation, and its meanings (Ellis, 2002). Consequently, high-frequency words are encountered more frequently, resulting in a quicker build-up of memory traces and facilitating easier recall and comprehension. It is important to note that word frequency is not a fixed characteristic but can vary significantly across different contexts and language domains. This variation in frequency can impact how words are learned and processed in various situations. For example, there is an apparent contrast between words in spoken language and words in writing (Durrant et al., 2022). Frequency also can vary considerably based on the specific text being examined, and it can be interpreted differently depending on the nature of the texts being analyzed. For instance, there is a contrast between words used in fiction and those found in academic research (McCreath et al., 2017; Ludewig et al., 2023).

Given the vast vocabulary of most languages, learners face the challenge of prioritizing which words to focus on when acquiring a new language (Nation & Meara, 2013). Since it is impossible to master every word, this prioritization often considers frequency. High-frequency words are typically encountered more often in various contexts, leading to increased familiarity and repeated practice. As a result of this extensive exposure and usage, these words are generally processed more efficiently by language learners, making them a good starting point for language acquisition (Nation & Meara, 2013). This increased processing efficiency is primarily due to the learner's repeated exposure and practice. However, it might also be influenced by the inherent qualities of the words themselves, such as phonological simplicity or semantic transparency. The more a word is encountered and used, the stronger the neural pathways associated with its recognition and retrieval become, leading to faster and more automatic processing.

In psycholinguistics, word frequency has been shown to affect the ease and speed of lexical access (i.e., the process of retrieving words and their meanings) (Balota & Chumbley, 1984; Rayner & Duffy, 1986). Forster (1976) proposed that frequency is a critical factor in organizing the mental lexicon, the internal mental dictionary. This phenomenon, known as the frequency effect, translates to faster processing of higher-frequency words compared to lower-frequency words, observed in both native (L1) and second languages (L2) (e.g., Dahan et al., 2001). However, proficiency in the L2 can modulate the magnitude of this effect. While high-frequency words generally maintain an advantage, increased proficiency can lessen the processing speed difference between high- and low-frequency words (Dahan et al., 2001; Yi and Ma, 2017).

Familiarity, an individual's understanding of words or phrases (Harley, 2014), complements the role of frequency in vocabulary learning. Hallin and Van Lancker-Sidtis (2017) found that highly familiar words are processed more quickly than less familiar words, regardless of frequency. Familiarity is rooted in personal language experience rather than solely in the frequency of public language use (Wang et al., 2021). This personal experience with language plays a vital role in vocabulary growth by providing access to the meanings of morphologically complex words through smaller, familiar parts (Chen et al., 2008). Familiarity influences various aspects of language processing. For instance, listeners recognize talkers better when they speak a familiar language, indicating the impact of language familiarity on talker recognition (Fecher & Johnson, 2018). Additionally, speech recognition is enhanced when the speaker is familiar, leading to better understanding and retention of speech (Fleming et al., 2014). This highlights the importance of considering the learner's existing knowledge alongside frequency when creating learning materials.

2.1.2.3. Prior Knowledge

Prior knowledge helps learners to learn better and acquire more vocabulary (Hattie and Yates, 2014). Prior knowledge or pre-knowledge refers to the skills, expertise, or aptitude learners bring to the learning process (Jonassen & Gabrowski, 1993; Dochy & Alexander, 1995). For example, when learning new vocabulary in a second language, a learner with a strong foundation in their native language can more easily grasp new words by drawing parallels between the languages. This comparative method can enhance understanding and retention of vocabulary. Prior knowledge plays a significant role in forming new cognitive schemas for acquiring new information (Bartlett, 1995). It is shown that prior knowledge reduces cognitive load, resulting in better learning engagement (Myhill and Brackley, 2004; Mihalca et al., 2011). In vocabulary learning, this might mean that a student familiar with Latin roots can more efficiently learn English words derived from Latin, as they can break down and understand new words more quickly, thereby reducing the mental effort required to know them.

The cognitive load of students with low prior knowledge requires greater support with learning, whereas that of students with high prior knowledge is perceived as lower by the student (Myhill and Brackley, 2004; van Riesen et al., 2019). According to Witherby and Carpenter (2021), students with more existing knowledge about a topic tend to learn more new information related to that topic, as demonstrated by their better scores on knowledge tests. Coyne et al. (2019) noted that students who possess a wider range of words at the beginning of their learning journey are more efficient in acquiring new vocabulary as they are exposed to more oral and reading materials. They can also integrate the newly learned words into their knowledge more effectively. Thus, successful integration of prior knowledge into new information depends upon the capacity of existing knowledge to expand through the introduction of new ideas and concepts.

2.1.2.4. Item simplicity

The simplicity of words plays a crucial role in memory performance, particularly in recognition memory, where simpler items are more readily distinguishable and require less cognitive effort for recognition (Humphreys et al., 2000). Simpler items also demand fewer cognitive resources and occupy less space in working memory, facilitating their maintenance and retrieval without interference (Azizian & Polich, 2007). In the context of vocabulary learning, even when attention is diverted, prioritizing specific words within the visual field can enhance their subsequent recall. This demonstrates that the brain can allocate resources for future retrieval without direct attention (LaRocque et al., 2014). For instance, when learning new vocabulary, if certain words are subconsciously marked as necessary, they are more likely to be remembered later, even if they were not the primary focus during the initial learning session. This indicates that implicit prioritization can significantly aid in retaining new vocabulary.

In essence, simplicity optimizes cognitive resource allocation, enabling more efficient recognition, improved working memory maintenance, and enhanced prioritization during word learning (Chen et al., 2022; Galdo et al., 2022). As explained earlier, working memory can be likened to a limited-capacity workspace for processing new information. Simpler items, such as words with fewer syllables or easily distinguishable phonemes, necessitate less processing effort for recognition. The reduced cognitive effort required for recognizing simpler items allows for deeper processing of the item's meaning and associative connections. Additionally, simpler words occupy less cognitive space, akin to how a concise definition is less mentally taxing than a complex, multi-layered explanation (Feldman, 2016; Camos & Portart, 2014).

This facilitates better retention of the word form and its meaning within working memory, which is essential for effective encoding and later recall (Carter & Vitányi, 2003).

Furthermore, simplicity enhances the prioritization of word learning. Encountering complex stimuli, such as words with intricate morphology or unfamiliar phonetics, divides attention, complicating the focus on core elements necessary for encoding. In contrast, simpler words are more readily salient, enabling learners to direct their attention and cognitive resources more efficiently toward the word's form and meaning. This, in turn, leads to more effective word learning and vocabulary acquisition (Wong & Perrachione, 2007; Glavaš & Štajner, 2015).

2.1.2.5. Social Factors

Social factors significantly impact vocabulary acquisition, especially when learning a second language (Adwani and Shrivastava, 2017; Tawfiq, 2020; Vyas and Sharma, 2022). One of the powerful methods for developing vocabulary is to be immersed in a language environment that speaks the target language. Immersion in a second language is effective due to its constant and contextualized exposure, leading to enhanced understanding and retention (Linck et al., 2009). This exposure in an immersive environment allows learners to continually encounter new words and phrases in various contexts, aiding in grasping meanings through context clues and repeated usage (Linck et al., 2009). As a result of immersion, speaking and listening skills can be improved, vocabulary can be expanded, and the language barrier may be removed.

In social psychology, the social factors that affect vocabulary acquisition include motivation, attitudes toward language learning, and the sociocultural context in which the language is learned (Siegel, 2003; Gardner, 2005). There has been a shift in second language acquisition towards a context-oriented perspective, which emphasizes the role of social factors in L2 processing research (Adwani & Shrivastava, 2017). According to Gardner and Lambert (1972), second languages (L2) are intermediaries within multicultural environments that bridge different ethnolinguistic communities. It was noted that the desire to acquire the language spoken by another community plays an essential role in either facilitating or impeding intercultural communication. Therefore, it is beneficial for learners to engage in conversations and dialogues with peers, teachers, and native speakers to understand various cultural contexts, idiomatic expressions, and word usage (Illés and Akcan, 2016). In this manner, learners can acquire new words and gain a deeper understanding of their cultural and contextual meaning, thus improving their ability to comprehend the language (Davis, 2003).

2.1.3. Methods of learning vocabulary:

Learning new vocabulary can be accomplished through various methods. This section will explore some of the most common techniques used in teaching vocabulary, tailored to the needs of the current thesis. These methods include memorization, multimedia resources, digital tools, flashcards, and word production. Memorization is a traditional approach to vocabulary learning that involves repeated exposure to new words and their meanings, which enhances memory retention (Nemati, 2009; Deeb, 2017). This method can be associated with various language skill areas, such as speaking and writing (He and Shi, 2008; Chen et al., 2016). Memorizing vocabulary often involves learning the grammatical roles of words, such as nouns, verbs, and adjectives. This understanding helps learners structure clear and grammatically correct sentences in speaking and writing. By memorizing new words and their meanings, learners expand their ability to express themselves and understand others.

Memorization strategies involve techniques for committing language material to memory, with many closely linked to vocabulary acquisition (Gu, 2010; Sinhaneti & Kyaw, 2012). This has led to the term "vocabulary memorization strategy," which explicitly describes methods for memorizing individual words. Recent studies have distinguished between vocabulary and text memorization strategies to understand better how English as a foreign language learners handle longer texts during memorization to enhance their language learning (Yu, 2017; Wang, 2023).

Word production is a learning/teaching method that requires learners to actively engage with new words by verbalizing them aloud, creating a deeper memory trace by requiring active cognitive processing (Levelt, 1992; Bock et al., 2002; Greenwood, 2010; Kormos, 2014). Studies of language production have historically been less extensive than studies of language comprehension. This applies to research studying word and sentence levels (Vigliocco et al., 2012). Producing and repeating words through reading or verbalization has, however, emerged as a pivotal method that deserves further investigation (Icht & Mama, 2022). Studies indicate that this approach can improve vocabulary learning and later acquisition (Punar-Özçelik and Uzun, 2021). Among the many methods and strategies available, this thesis focuses on production as a method that reinforces vocabulary retention, as will be discussed extensively in section (2.3).

2.1.4. The Bilingual Learner

This section provides an overview of bilingualism with a specific focus on children who grow up in bilingual households, as distinct from children learning a second language solely in formal educational settings. Unlike learners who encounter a second language in controlled environments (e.g., Arab children learning English at a British school), children in bilingual homes interact with two languages regularly, impacting their vocabulary acquisition and cognitive processing in unique ways. This section defines bilingualism, examines the characteristics of Modern Standard Arabic (MSA) learners within bilingual environments, and explores current theories on bilingual lexical processing to shed light on how these experiences shape vocabulary learning. Bilingualism refers to the ability to speak and understand two languages. According to Dewaele (2015), a bilingual individual is competent in at least one of the four language skills (speaking, listening, reading, and writing) in a language that is not their first language or someone proficient in two or more languages. Similarly, Prior and van Hell (2021) consider any person who uses more than one language daily to be bilingual. In this view, bilinguals can have more fluency in all topics in both languages as they use their languages for different purposes, with various interlocutors in other domains of life (Grosjean, 2010; Cook and Bassetti, 2010). The proficiency level in a language can vary based on its necessity and domain of use.

When discussing bilingualism in this research, defining the term in the context of the participants' linguistic experiences is crucial. In this study, all participants are bilingual, speaking both English and a Colloquial Arabic dialect (e.g., Saudi, Cairene, Levantine, Yemeni). However, their experience with Modern Standard Arabic (MSA) is limited, primarily in formal educational settings. The critical question is how differences in L1 (first language) might affect L2 (second language) vocabulary acquisition, particularly in the case of bilinguals whose L1 (Colloquial Arabic) is similar to L2 (MSA) versus learners whose dominant language (English) is typologically different from L2. Given the overlap between Colloquial Arabic and MSA, these participants may benefit from shared linguistic structures, which could facilitate vocabulary acquisition.

In contrast, the influence of English as a dominant language may introduce additional challenges due to the greater linguistic distance from MSA. These bilingual differences are crucial in the production effect (PE) theory. PE suggests that vocalizing words during learning improves recall and recognition, but how this effect might vary between participants with a shared linguistic base in Colloquial Arabic and MSA versus those for whom English dominates is a crucial area of investigation. This study explores whether participants with different linguistic backgrounds experience PE differently and whether the similarities or differences between L1 and L2 influence the effectiveness of vocal production as a vocabulary learning strategy. Thus, while bilingualism is not the primary focus of this research, it serves as a valuable framework for investigating the role of L1 in L2 vocabulary acquisition. Understanding these dynamics can offer insight into how bilingual learners respond to learning strategies like the Production Effect when acquiring MSA vocabulary.

English learners of Modern Standard Arabic (MSA) are unequivocally categorized as L2 learners. Yet, a pivotal debate among MSA scholars is whether to classify MSA learners with an Arabic background as second language learners and to define those proficient in colloquial Arabic (of any variety) and MSA as bilinguals. These controversial views appeared as these two “languages” have the same grammatical roots, but they differ in their phonology, morphology, and lexicon as a result of their historical development (Al-Sughayer, 1990; Holes, 2004; Laks & Berman, 2014; Al-Omari, 2019).

Arabic native speakers learn their language in a unique linguistic context called diglossia (Ferguson, 1959). This means that different spoken dialects of Arabic coexist with a more formal variety known as MSA (Abu-Rabia, 2000; Maamouri, 1998). There are some differences and similarities between the two at all levels of linguistic description. MSA is used in formal situations like public occasions, religious contexts, media, and the press. It is not any native speaker's first language, and Arabs usually learn it in school. MSA is the language used for Arabic literacy, i.e., reading and writing.

On the other hand, colloquial Arabic (CA) varieties are acquired early and used in everyday situations. Reading and writing in CA varieties is not shared and is limited to aspects of social media like text messages, Twitter, and Facebook (Al-Khatib & Sabbah, 2008; Mostari, 2009). While all native Arabic speakers utilize the same Modern Standard Arabic (MSA) variety, their spoken dialects can vary between states and within different regions of the same state. According to Cook's theory (Cook, 1991, 2002, 2003), those who speak colloquial Arabic and MSA are regarded as “L2 users.”

The relationship between languages in the mind of an L2 user has been a subject of ongoing debate in second language acquisition. Some researchers argue that languages interconnect to form a cohesive multicompetence, while others maintain that languages exist as distinct systems with specific interaction points. Cook (2002) defines multicompetence as "the knowledge of more than one language in the same mind" (p. 10) and later expands this concept to encompass all languages within a single mind or community, emphasizing the potential interconnectedness of L1, L2, and additional languages (Cook & Wei, 2020). This perspective challenges the notion of monolingual native speaker superiority and suggests that multicompetence impacts the entire cognitive system, not just language abilities. However, alternative views exist. White (2011) proposes the Interface Hypothesis, which suggests that distinct linguistic systems may interact at specific interfaces (such as syntax-semantics) in bilingual individuals. This hypothesis implies that while different languages may operate independently to some degree, they interact at particular points, influencing language processing and production. Cook (1999) advocates recognizing L2 users as multicompetent language users rather than deficient native speakers, challenging the traditional hierarchy that places native speakers above L2 learners. This perspective emphasizes the unique abilities of L2 users and their potential to leverage multiple linguistic systems. The debate between these perspectives highlights the complexity of bilingual language representation and processing. Factors such as proficiency level, age of acquisition, and language similarity may influence the degree of interconnection or separation between languages. Given this ongoing discussion, an intriguing question arises: How are different languages activated and processed in the bilingual brain, and to what extent do they interact or remain distinct?

Current theories of bilingualism assume that there is a combined lexical store. In this view, lexical items are connected to several linguistic features in a common lexical store (Dijkstra and Van Heuven, 2002; Kroll and Sunderman, 2003). Research on bilingual language processing provides evidence for the non-selective view of linguistic processing. That is, both languages of a bilingual individual are activated simultaneously rather than accessing only the target language (i.e., the speaker L2) (Durrant et al., 2022). In other words, research on bilingual lexical access and language co-activation suggests that the activation of homographs in a target language can trigger their activation in a non-target language (i.e., speakers L1) (Wu and Thierry, 2010; Friesen and Jared, 2012; Durlik et al., 2016; Chen et al., 2017; Poort and Rodd, 2017; Woumans et al., 2021). Homographs, by definition, are words that share spelling but possess distinct meanings across languages (e.g., /sin/ in MSA means “age” or “tooth,” while “sin” in English means an immoral act).

As bilingual individuals gain more exposure to the non-target language, the degree of co-activation with the target language during spoken word comprehension increases (Chen et al., 2017). This heightened activation of the non-target language can lead to competition during processing, potentially causing ambiguity and slowing down the identification of the intended meaning of interlingual stimuli like homographs in the target language (Wu and Thierry, 2010). Interestingly, in tasks demanding the activation of lexical representations solely in one language, the non-target language's activation might be too subdued to activate homograph representations through its lexical connections with the target language (Titone et al., 2011).

Similarly, the cognate effect is related to the activation dynamics of homographs in the non-target language. Cognates are words that have similar forms and meanings across different languages. For example, the word /ʕayn/ directly means "eye" in both Colloquial Arabic (CA) and Modern Standard Arabic (MSA). However, in MSA, it also carries additional meanings such as "the heart of the thing" or "the water fountain," and the phrase /ʕayn ʔlħaqiːqa/conveys "the eye of truth." For a bilingual speaker, encountering this word may trigger its direct meaning in CA and English and the deeper connotations associated with its various meanings in MSA, posing challenges for MSA learners. The cognate facilitation effect underscores that bilinguals can read, hear, and articulate cognates faster than their non-cognate counterparts (Wu & Thierry, 2010). Two contrasting versions of the nonselective access hypothesis exist: a milder version positing that bilinguals cannot disregard an extraneous language in the environment or input, and a more assertive version suggesting that bilinguals' languages are perpetually active, potentially influencing their linguistic performance even when faced with monolingual tasks and input (Jiang, 2023). Given the significance of the cognate effect and the nonselective access hypothesis in bilingual language processing, understanding their interrelation and impact on cognitive functions is crucial. In the context of the current thesis, it is particularly vital to investigate how exposure to a second language influences activation dynamics in the first language. This exploration is essential for comprehending bilingual language acquisition mechanisms and potential cross-linguistic interference.

2.1.5. Children and Vocabulary Learning

Learning vocabulary is crucial for children's language development and cognitive growth. It significantly influences their ability to communicate effectively, understand complex concepts, and achieve academic success. This section explores the fundamentals of children's vocabulary learning, including the critical period hypothesis and age of acquisition, and highlights the importance of language input and usage patterns in shaping language acquisition.

2.1.5.1. The Critical Period Hypothesis and Age of Acquisition

Research indicates that there is a critical period for language learning. This cognitive ability peaks in early childhood and declines with age, making language acquisition more difficult as one age (Johnson & Newport, 1989; Lenneberg, 1967; Singleton, 1995). The notion of a critical period also called a 'sensitive period,' is deeply rooted in developmental psychology and neurology. It suggests that the human brain is exceptionally receptive to language learning only during an early phase of life. This period is generally considered to extend from birth to the onset of puberty, with some variability across individuals (Knudsen, 2004).

The Critical Period Hypothesis posits that there is an optimal period in a person's life during which language can be acquired with native-like proficiency and after which language learning becomes increasingly difficult and ultimately less successful. Lenneberg (1967) suggests that the human brain is more malleable and receptive to language input during this "critical" window, which typically spans from infancy to early adolescence. This seminal work was pivotal in outlining the biological foundations of the critical period for language acquisition. He argued that the lateralization of the brain's language functions (i.e., the process by which one hemisphere, usually the left, becomes specialized in language skills) solidifies around puberty (10-12 years old), thereby reducing the capacity for language learning thereafter. This aligns with findings that the ability to learn new words declines after the first few years of life, suggesting a critical/sensitive period for vocabulary acquisition in the first language (Jiang, 2021). During early infancy, there is an initial bias to attend to specific linguistic cues (e.g., frequent frames like "a dog" or "the car"), which facilitates rapid word learning. However, this bias declines after the first year of life (Friedmann & Rusou, 2015), indicating that vocabulary acquisition and language learning capacities are most pronounced in early childhood.

While the Critical Period Hypothesis proposes a strict window of opportunity for language acquisition, the Age of Acquisition (AoA) effect offers a more gradient perspective. This theory suggests that the age at which a language is learned can influence the proficiency and neural representation of that language (Perani et al., 2003; Cargnelutti et al., 2019; Oh et al., 2019; Malaia et al., 2020; Elsherief et al., 2023). The earlier a language is acquired, the more likely the learner will achieve higher proficiency and more native-like neural processing.

The impact of the Age of Acquisition (AoA) effect on lexical retrieval is determined by the influence of individual learning experiences on the progressive formation of mental representations and the connections between these representations (Ellis and Morrison, 1998; Navarrete et al., 2015; Elsherief et al., 2023). Words learned earlier in life are often more quickly and easily recalled than those learned later. This phenomenon can be attributed to the more deeply entrenched neural connections formed during the critical periods of language acquisition when the brain is more malleable and sensitive to linguistic input. For example, a native speaker of English who learns the word "apple" in early childhood is likely to retrieve it more readily than a complex technical term learned in adulthood due to the foundational placement and frequency of use of "apple" in their linguistic repertoire. This emphasizes that the Age of Acquisition (AoA) impact is relevant to acquiring knowledge at all stages of development. In essence, the Age of Acquisition (AoA) effect is a fundamental aspect of cognitive learning that will likely play a role in developing a comprehensive theory of lexical retrieval throughout one's lifespan.

These research findings have enhanced the understanding of language development and have had a practical impact on societal approaches to language education. They have influenced the methodologies adopted by educators who teach languages to young children, promoting techniques that mimic the natural language acquisition processes during the critical period. Additionally, the AoA effect has essential ramifications for diagnosing and treating language disorders, suggesting that early intervention is crucial for effective remediation (Elsherief et al., 2023).

2.1.5.2. The Process of Early Language Acquisition in Children

As children approach the latter half of their first year, they start to comprehend the common words and simple phrases employed by their caregivers, as demonstrated by their reactions and gestures (de Boysson-Bardies, 2001; Athari, 2021). Generally, by the end of the first year, children articulate their first recognizable words, which include terms typically used by parents, siblings, or everyday items such as "bottle" and "dog." These early words signify a significant shift from preverbal to verbal communication. Historically, Skinner (1957) proposed that positive reinforcement is the driving force behind language acquisition; however, Chomsky (1965) challenged this theory by asserting that inherent biological mechanisms are a crucial factor in language learning. Nonetheless, the usage-based theory adds another dimension to the study of children's language acquisition (Lieven & Tomasello, 2008). The theory emphasizes the role of language use itself in shaping how children learn. Children are not just passive recipients of language input and reinforcement. They actively participate in communication, using language in different contexts. This constant exposure and practice, a reassuring and confidence-boosting aspect, helps them refine their understanding of grammar and vocabulary. The child's environment, particularly the role of caregivers who name objects, respond to vocalizations, and model simple language structures, is pivotal during this stage (Kuhl, 2010; Rosselli et al., 2014). Therefore, the process of word learning is significantly influenced by the interaction between the child and their environment, which shapes the path for lifelong language use.

Enhancing vocabulary is crucial for language proficiency and cognitive growth, particularly in children. A rich and nuanced vocabulary significantly shapes a child's ability to communicate effectively, grasp intricate concepts, and succeed academically (Cummings et al., 2018; Ramsook et al., 2020). Children absorb new words during the language acquisition stages, laying the foundation for sophisticated linguistic abilities. This process starts with the native language, effortlessly picked up from their surroundings and rapidly advanced by caregiver interactions and environmental stimuli (Cook & Cook, 2010; Guerrout et al., 2019; Swanson, 2020; Masek et al., 2021). Additionally, children may learn a second language or multiple languages as they mature through formal education or as part of a multilingual setting (i.e., being raised in a diverse community or a multilingual household). For example, a caregiver might teach the word "rabbit" by pointing toward it. However, this method has limitations, especially when teaching abstract terms such as "happiness" or complex grammatical forms like prepositions, verbs, or articles (Tomasello, 2003). This complexity suggests that the direct teaching of vocabulary is not always straightforward.

2.1.6. Learning Modern Standard Arabic

Arabic is one of the most widely spoken languages, with more than 300 million native speakers in the world. It is the official language of 27 countries, ranked fourth in the number of people who speak it as a first language (Habash, 2010). The main reason for its wide use amongst Muslims is that the holy book of Islam (i.e., the Quran) is written in Arabic but translated into other languages. Non-Arab Muslims and non-Muslims also learn it as an additional language to enable them to speak it and use its writing system in their political or educational work or to learn about Arabic culture (Aladdin, 2010).

Arabic has several varieties; the main variety is Classical Arabic, the old Arabic, or “al-Arabiya al-fuṣḥā” (the purest) used in the media and literature. It is usually distinguished from Modern Standard Arabic (MSA), which is used in daily communication and education. Although both styles follow the same syntactic and phonological rules, MSA differs in that it allows loanwords and borrowings from other languages to follow the rapid growth in different fields of life, including technology, science, or general social media terms (Ordan et al., 2010). For instance, The Arabic Language Academy in Cairo, Egypt (established in 1932) added the word "trend" to the Arabic language dictionary (Alarabiya, 2023). The term "trend" is commonly employed in English to refer to popular topics rapidly gaining attention on social media platforms. These topics are disseminated widely within a short period, leading to frequent discussion and circulation among the public.

Within Modern Standard Arabic (MSA), there are regional dialects called Colloquial Arabic, which are used for informal communication in geographic areas (e.g., Cairene in Egypt). Native speakers of MSA might use it in formal contexts such as school or work emails, but they typically use their local dialects for spoken and informal written communication. This creates the phenomenon of diglossia, as described by Hudson (2002), where two levels of a language—MSA as the "high" variety and the local dialect as the "low" variety—are used simultaneously, with MSA for formal writing and local dialects for everyday speech. However, MSA is vital in mediating interactions between these high and low varieties, creating a dynamic linguistic system (Parkinson, 2005). The degree of MSA use varies across regions; for instance, in countries like Lebanon and Morocco, local dialects are sometimes used even in formal public speeches. By contrast, President Abdel Fattah El-Sisi prefers to use the Cairene dialect when addressing the public in Egypt. In contrast, in Saudi Arabia, MSA is consistently used for formal speeches, with the Saudi dialect reserved for informal interviews (e.g., on TV shows). This dialectal variation has been considered when examining the participants' language backgrounds and selecting the vocabulary for the current study.

Research on language acquisition in diglossic contexts, particularly in Arabic, highlights several challenges learners face navigating between the formal (high) and informal (low) varieties (Abu-Rabia, 2004; Khamis-Darkwar, 2010). In Arabic-speaking countries, children are typically exposed to Colloquial Arabic as their first language (L1), acquiring it naturally at home. However, their exposure to MSA begins later, usually when they enter school, where it is used as the medium of instruction and in written materials (Maamouri, 1998). This delayed and formal introduction to MSA often leads to uneven proficiency between the two varieties, with children usually becoming more fluent in their local dialect and less comfortable using MSA (Eisele, 2002).

Studies show that learning MSA in diglossic contexts can result in what some scholars call "compartmentalized bilingualism," where speakers develop separate competencies in the formal and colloquial varieties (Ferguson, 1959; Albirini, 2014). This can pose cognitive and linguistic challenges, particularly in vocabulary acquisition, as learners must continuously shift between the two varieties, sometimes treating them as distinct linguistic systems. Ayari (1996) suggests that the complexity of this diglossic situation may hinder literacy development, as children may struggle to transfer their spoken language skills (in Colloquial Arabic) to the formal written system of MSA. Similarly, in other diglossic languages, such as Swiss, German, or Greek, research has found that learners face challenges in acquiring the high variety due to limited exposure to everyday communication (Holmes, 1992). In these contexts, children may develop more vital oral skills in the low variety while encountering difficulties in mastering the formal grammar and vocabulary of the high variety, especially when they are introduced to it later in their educational journey. This has important implications for vocabulary learning in diglossic contexts. In Arabic, for instance, research by Ibrahim (2009) highlights the potential for confusion when children encounter words in MSA that are either different from or more formal versions of their colloquial equivalents. For example, the word for "book" in MSA is "kitāb," while the colloquial pronunciation or word may differ slightly in many dialects. This lexical distance can create an additional cognitive load during vocabulary learning, particularly for younger learners.

In the context of this thesis, these findings are particularly relevant. Since the participants are bilingual in Colloquial Arabic and English, their acquisition of MSA vocabulary must account for the cognitive and linguistic demands of navigating diglossia (between MSA and Colloquial Arabic) and their bilingualism with English. The PE, which this study investigates, may interact with these linguistic challenges. The role of vocal production in enhancing memory may vary depending on how familiar or distinct the MSA vocabulary is from the learners' everyday Colloquial Arabic, further complicating L2 acquisition in this diglossic context.

2.1.6.1. Arabic Phonology and Morphology

Regarding the segments, MSA has 28 consonants; however, MSA differs from English in that it lacks the plosive /p/, the voiced /v/, the affricates /tʃ/ and /ʤ/, and the nasal /ŋ/, with additional glottal fricatives /ʔ/, /ħ/, and /ʕ/. The phonology of MSA has the distinctive feature of emphasis, which is indicated by using the symbol /ˤ/ from the IPA chart, in which four consonants are considered emphatic (i.e. /t/, /d/, /ð/, /s/). The vowel quality and the phonetic form of vowel production is affected by the consonant environment surrounding the vowels. MSA contains limited vowels compared to the 12 vowels in English. Arabic has three vowels that can be long or short: /a/, /i/, /u/, /aː/, /iː/, /uː/. Regarding the syllable structure, MSA disallows both onset-less syllables and onset clusters and prohibits other syllable structures. The allowed syllable sequences are the light syllable, CV; the heavy syllables, CVC and CVV; and the super-heavy syllables, CVVC and CVCC (Holes, 1995; Al-Ani, 2014).

To understand word formation in Arabic, it is necessary to understand the Arabic morphology similar to other Semitic languages and has a root-based approach. The roots in Arabic words mostly have three consonants, but adding more consonants and vowels is possible. The root patterns determine the semantic meaning of a word and its phonological categories. However, grammarians believe that MSA is a derivational language in which nouns are derived from verbs or other nouns to give a semantically related meaning or to produce a new word (Ibrahim, 2010). For instance, the word "library", in Arabic "Ma<u>ktb</u>a" /mkˈtabæ/, is derived from the verb /ktb/, "he wrote". Nevertheless, nouns in MSA are also formed using the compounding technique to produce one of three types of compound nouns: a genitive noun, or "Iḍāfah" (e.g. "ˈ*Kitabu Al-Arabi*" /kitabu ˈʔlʕra:bi/ "The book of Arabic"); a predicative noun, which combines definite and indefinite nouns (e.g. "rajulun qawi" /rad͡ʒulun ˈqawi:un/ "a strong man"); or a syntactic compound noun (e.g. "Maliku Alghaba" /maliku ˈʔlɣa:bæ/ "The king of jungle" or "the lion"). The final type of noun formation in MSA is noun borrowing, which is used by the general public and scientists to "Arabize" a word or export a word and adapt it according to the phonological rules of MSA. For instance, it is likely to hear, "I need 'iskreem' /ˈʔiskiriːm/ to beat the heat of this city!" in which "iskreem" /ˈʔiskiriːm/ is a lexical term that has undergone phonological and morphological changes to be identified as an English loanword in Arabic to describe the popular frozen dairy product (Holes, 1995; Ibrahim, 2010; Hazem & Meteab, 2019).

Understanding Arabic phonology and morphology is essential for learners acquiring MSA vocabulary. Phonological differences, such as emphasis and vowel-consonant interactions, affect how new words are pronounced and processed. Additionally, Arabic's root-based morphological system offers a powerful tool for vocabulary learning. By mastering root patterns, learners can infer the meaning of unfamiliar words, leading to more effective vocabulary expansion. The derivational nature of Arabic allows learners to build semantic networks around core roots, enhancing both word retention and recognition. Learners' familiarity with these linguistic features can significantly impact how they process, learn, and recall new Arabic words, making phonological and morphological knowledge fundamental to effective vocabulary learning in Arabic.

2.1.6.2. Learning and Acquiring Arabic

In terms of acquiring Arabic as a first language, children are exposed to MSA prior to attending school, mainly through TV programmes (e.g., watching cartoons or educational shows, such as the Arabic version of Sesame Street); they are then formally exposed to it at school to learn its rules (Albirini, 2016). The acquisition of MSA consonants sounds was analysed by Dyson and Amayreh (2000) in typically developing children between the ages of 2;0 and 8;4 years. The authors classified the developmental stages of acquisition to three main categories: early, intermediate, and late (Table 2.1). In their study, they noticed it is easier for children to acquire medial consonants than initial or final consonants. For example, when analysing data on children learning the Arabic consonant /t/, it was easier for children to learn it in /ˈmatˤar/ "rain" (word medial) rather than learning it in an initial position /talaˈfawn/ "a telephone", or in a final position /baˈnat/ "girls".

Table 2.1: The acquisition of MSA consonants as adapted from Omar (1973) and Dyson and Amayreh (2000)

Manner of articulation	Age		
	2;0–3;0 years	4;0–6;0 years	7;0–8;4 years
Stops	b,d,ʔ,a	g,q, t, ʕ, d͡ʒ,k	tˤ, dˤ
Fricatives/affricates	h, ħ, f	s,χ,ð,γ,θ	Z, ʃ, sˤ
Liquids	m	n, l, r	
Glides	w,y		

With regards to learning Arabic as an additional language, there has been an ongoing controversial debate about the teaching of MSA or its varieties as a conversational language to non-native learners. Because MSA is used in formal settings, scholars are calling for it to be taught as a lingua franca (Jaradat and Al-Khawaldeh, 2015). However, another factor that should be considered when learning MSA is the difficulties that face second-language learners. In a study by Dijani et al., (2014), the authors noted that, due to the lack of existence of some MSA consonants (e.g., /tˤ/, dˤ) in other languages, learners of MSA as a second language struggle to produce words that have these phonemes. Another difficulty is the recognition of different noun classifications, especially genitive nouns or "Iḍāfah", and using the dictionary to understand the meaning of unfamiliar words in Arabic, as the language depends on the root and its derivatives.

Modern Standard Arabic (MSA) uniquely serves heritage language learners, particularly in multicultural and multilingual communities like the United Kingdom. While it differs from the spoken Arabic dialects used at home, MSA serves as a unifying form of communication and a representation of cultural identity for Arabic speakers (Park, 2013; Eghbaria-Ghanamah et al., 2022). MSA functions as a second "Arabic" language for many heritage learners, distinct from their colloquial dialects. It acts as a flexible lingua franca [1]among Arabic speakers, facilitating communication across different dialectal backgrounds (Solimando, 2019). This is especially important because spoken dialects can sometimes be difficult for speakers of other Arabic varieties to understand.

2.1.6.3. Children Learners of MSA

Since this thesis focuses on learning MSA as an additional language, the participating children must possess prior knowledge of MSA phonology. This prerequisite ensures that they understand the sounds used in Modern Standard Arabic, which is critical for accurately learning and repeating new vocabulary. Prior exposure to MSA phonology, whether through family,

[1] A **lingua franca** is a language used to communicate with people who do not speak a native language. It is a standard communication method, particularly in multicultural and multilingual settings. This term encompasses various languages facilitating interaction among speakers of different mother tongues, often emerging in trade, diplomatic, or cultural contexts (Merriam-Webster, 2024).

education, or cultural background, supports the ability to produce words correctly, a key factor in studying the Production Effect (PE). Accurate vocal production is necessary for the PE to be fully effective, as incorrect pronunciation could hinder learning and memory retention. This section provides an overview of children learning Modern Standard Arabic (MSA) as an additional language and examines the crucial role of phonological awareness in their MSA acquisition process.

A notable illustration of sociocultural context within the field of sociolinguistics is the teaching of Arabic as a second language (L2) to immigrant children of Arabic descent. These children are considered heritage language learners (Dhahir, 2015; Alabd, 2016; Albirini, 2016). By definition, heritage speakers are generally bilingual individuals from immigrant or minority backgrounds. They often grow up speaking their family's native language at home while simultaneously or subsequently learning the dominant language of the wider society and social environment. As a result of such a sudden or gradual shift, L1 development may be limited and different features of L1 may need to be adequately acquired (Valdes, 2005; Montrul, 2008, 2016). Language exposure to L1 in a majority language-dominant environment varies based on the languages spoken at home and outside the home (i.e., only L1 or both L1 and L2) and how often they are used. As opposed to monolingual language environments where exposure to L1 prevails, heritage speakers are exposed to different languages simultaneously, and the proportion of this exposure is generally not in favour of L1 (Thordardottir, 2011; Hoff et al., 2012, Montrul, 2016). The L1 proficiency levels of these learners vary where some have limited reading, writing or speaking skills while others show native-like literacy skills (Valdes, 2005; Montrul, 2010).

In the case of Arabic learners, it is often questioned whether linguistic integration, the maintenance of cultural heritage, and the potential impacts on their bilingual proficiency would be appropriate to immerse these learners in Arabic-centric educational environments when they reside in regions where Arabic is the nondominant or a minority language (Surrain, 2018; Aldawood et al., 2023). The majority of these children are likely to be exposed to Arabic as a minority language in three settings: their homes, schools that teach the Arabic language, and events held by members of Arabic-speaking communities (Ferguson, 2013; Al-Sahafi, 2015; Said and Zhu, 2017; Yazan and Ali, 2018; Bahhari, 2020). Being immersed in a target language environment, such as language teaching institutes, provides unparalleled advantages in vocabulary development, and the daily exposure to the language forces learners to adapt and engage with a broader lexical range (Van Lier, 2004; Mroz, 2014; White et al., 2013). As this case involves linguistic heritage, identity formation, and a sociocultural context in which these individuals navigate their educational trajectories, it deserves in-depth consideration. Hence, among the many groups of learners, this thesis focuses on child learners of Arabic as a heritage additional language.

Research suggests that child learners of Modern Standard Arabic (MSA) residing in minority Arabic-speaking countries, such as English-dominant nations (e.g., Australia, the United Kingdom, or the United States), often face challenges in achieving proficiency and exploring the language beyond the classroom setting (Abuhakema, 2012; Labanieh, 2019). This limitation is likely due to the dominance of the surrounding language (e.g., English) in their daily lives. This dominance may restrict opportunities for MSA practice outside of formal instruction, potentially hindering fluency development. Research indicates that child learners of Modern Standard Arabic (MSA) in minority Arabic-speaking environments, even those with a dominant Arabic dialect background, may encounter challenges in demonstrating proficiency in using MSA vocabulary for everyday conversations.

This difficulty can be attributed to the linguistic disparities between MSA and local Arabic dialects, a phenomenon known as diglossia (Saiegh–Haddad, 2003). Diglossia, as explained earlier, presents a particular obstacle because vocabulary knowledge is a critical factor in speaking proficiency (Koizumi & In'nami, 2013). The distinctions between the written MSA and spoken regional Arabic dialects can create confusion and hinder learners' ability to readily access and apply MSA vocabulary in everyday contexts (Zaidan & Callison-Burch, 2014). To explain, although the syntactic structure of MSA sentence is similar to the dialectal Arabic, the choice of words would be challenging for MSA learners. For example, the learner needs good command of MSA (i.e., the grammatical structure and precise words) to be able to express their intention of going to school the following day even if they speak a dialectal Arabic (Hijazi Arabic in the following example):

Hijazi Arabic: /ħɾuːħ llmadrasa bukra/

MSA: /saˀaðhab ʔilaː almadrasah ɣadan/

English: (I am going to school tomorrow)

In this study, children who are learning MSA are considered heritage language learners. They are also defined as children who were born in, migrated to, or are residing in the U.K., with caregiver(s) who are originally from an Arabic speaking country, and who are exposed to the

Arabic language either at home or at Arabic Institutes (i.e., weekend schools). Typically, these children have not reached the literacy level of native Arab speakers of MSA (i.e., ability to read and understand complex sentences), but some of them are using dialectal Arabic (i.e., Saudi Arabic, Levantine Arabic, Iraqi Arabic, etc) as a dominant speaking language without having the full literacy in it (i.e., the ability to read passages and write complex structured sentences in the spoken language). It is also worth mentioning that these children do not have the ability to converse in MSA as they use dialectal Arabic (some participants in study 1) or English (participants in studies 2 and 3) as a way of communication. In other words, the participated children in the study have the basic level of MSA to understands its alphabets, differentiate the variety sounds of letters, and have the ability to write simple MSA words.

It is also important to consider the purpose for learning a language. For the participants in the current study, different factors contribute to each family's decision to send their children to weekend schools. Some families from this study reported that it is important to link the child with their identity through learning Arabic as a mother tongue or a second language in order to understand the Qur'an and Islamic literature. Many parents assumed that it will be easier for their children to learn Arabic at an early age, as it is a difficult language to learn due to having a system that is unlike their first language (i.e., English). A parent in this study mentioned that they have tried to home-school their children to teach them Arabic by using expressions in place of English throughout the day and by watching films and cartoons, but that was not effective, so they decided to send the children to the weekend school.

Within the context of the current study, MSA is used as a tool to investigate the PE theory as children had the phonological awareness of Arabic (i.e., the basic level to distinguish the specific sounds in a spoken word) which is considered as an important element when acquiring a target language similar to the background language. Phonological awareness is the ability to analyse the phonetic characteristics of words regardless of their meaning (Hatcher et al., 1994; Stackhouse and Wells, 1997). It involves modifying the sounds found within spoken words (Goswami and Bryant, 1990). It also recognizes smaller units of speech (phonemes) and includes significant elements such as syllables and rhymes. In addition to identifying similarities between words, PA skills include manipulating words (blending and segmenting them into phonemes and syllables), recognizing rhyming words, recognizing words that share initial sounds, and understanding word components, such as phonemes and syllables (Alcock et al., 2010). In other words, phonological awareness is hearing and consciously breaking words into syllables, onset, and rhyme, and individual sounds or phonemes. It is a foundational skill that helps to develop later word recognition abilities and is a crucial predictor of future literacy acquisition. According to Bennett and Arrow (2023), phonological awareness is the most important emergent literacy skill to develop in preschool-aged children and is the best early predictor for later reading success. Therefore, it is important to develop phonological awareness skills before formal literacy instruction.

Language acquisition involves the creation of mental representations, which evolve based on prior knowledge, shaping the assimilation of new linguistic concepts. A notable example of this process is observed in phonotactic probabilities (i.e., the likelihood of specific phonological segments or sequences occurring in a language). Through exposure to a language, individuals glean information about the frequency of specific sound patterns, aiding in acquiring new vocabulary. Research indicates that learners better recall unfamiliar words containing sounds commonly found in their native language (Thorn & Frankish, 2005). Familiarity with sounds is pivotal in vocabulary learning, particularly in early language acquisition.

2.2. The Production Effect

2.2.1. The Production Effect in Adults

Apart from linguistic elements, various factors, including the training methods employed, have also been identified as influential in the learning process. One of the methods investigated in vocabulary learning is producing words aloud (i.e., reading or repeating words aloud rather than silent reading or listening). The idea that producing vocabulary aloud enhances memory has been well-known in cognitive psychology. The effect of production was first described by Gates (1917), who had participants report all tactics they use when learning lists of words in an experiment that investigated memory strategies. Participants reported that remembering and recalling the words was greatly improved by pronouncing the items. As Gates (1917) noted, recalling words reinforced the physical (motor) and acoustic components. The production effect was explored indirectly in a study conducted by Hopkins and Edwards (1972) to test the frequency effect theory (Ekstrand et al., 1966). This theory posits that recognising produced words should be better than recognising unproduced words because pronouncing a word increases its apparent frequency in memory. The mechanism is that vocalisation creates an additional encoding event, enhancing the word's situational frequency (i.e., the number of times it is experienced in the experimental context). In their study, Hopkins and Edwards implemented study phases where participants read half of the items on a mixed list aloud and the other half silently. The results showed that produced words were recognised approximately 10% better than unpronounced items, supporting the hypothesis that pronunciation increases a word's distinctiveness in memory, making it more readily identifiable during recognition tests. This finding supported the frequency effect theory and laid the groundwork for future research on the production effect in memory. Similarly, Conway and Gathercole (1987) conducted a study to test the effect of modality on long-term memory. The modality principle (Crowder & Morton, 1969; Low & Sweller, 2005) suggests that information retention is enhanced by using a presentation format that combines auditory and visual modes, which is more effective than when information is presented either visually or auditorily. The dual-modality approach enriches the encoding process by engaging both auditory and visual sensory pathways, leading to more robust memory traces.

Conway and Gathercole (1987) conducted a series of experiments investigating modality effects in long-term memory using an incidental learning paradigm. The study comprised four experiments, each employing mixed-mode presentations featuring three different modes of presentation. The baseline presentation mode required subjects to read words silently in all experiments. The other two presentation modes varied across experiments: Vocalise or read and hear in Experiment 1; Read and hear or mouth in Experiment 2; Vocalize or write in Experiment 3; Vocalize or write without seeing the written word in Experiment 4. The researchers also conducted a fifth experiment where separate groups of subjects were presented with pure-mode lists. In this experiment, participants were asked to read silently, write without seeing, write and see, mouth, hear, read and hear, or vocalise. After the input phase in each experiment, subjects completed a delayed unexpected memory test.

The study results indicated that read-and-heard words had the highest recognition rate at 76%, followed by mouthed words with an intermediate recognition rate of 69% and read-only (silent) words with the lowest recognition rate of 62%. The benefits of mouthing words can be attributed to the additional motor component it introduces to the encoding process. When people mouth a word, they engage in articulatory movements without producing sound. This motor activity creates a distinctive cue that can aid in later retrieval. However, it is essential to note that the advantage of mouthing was less robust than vocalisation (reading aloud) across Conway and Gathercole's experiments. Vocalisation consistently led to the best retention among the auditory presentation procedures.

In summary, while mouthing offers some memory benefits over silent reading, its effects are less pronounced than full vocalisation or hearing the word. The additional motor component of mouthing provides an extra encoding cue but lacks the auditory component, making vocalisation particularly effective for memory enhancement. Conway and Gathercole's study showed that reading a word aloud significantly improved memory performance compared to reading silently. Speaking aloud engages motor processes, enhancing the distinctiveness of the memory trace by adding an auditory attribute to the visual input. This dual encoding strengthens the memory, making it more resistant to forgetting. Silent reading, relying solely on visual processing, lacks this additional reinforcement, resulting in weaker memory retention. Thus, the benefits of production in memory retention are grounded in the enhanced activation and integration of multiple sensory and motor pathways, leading to more durable and accessible memory traces.

More recently, MacLeod et al. (2010) named this phenomenon the production effect (PE hereafter), highlighting the memory advantage of explicitly remembering items read aloud compared to those read silently. This vocalization differs from the silent, inner pronunciation of words, known as subvocal rehearsal. Subvocal rehearsal, a component of Baddeley's working memory model's phonological loop, involves the silent repeating of words for temporary storage and manipulation. The production effect, however, emphasizes actively saying the words aloud, creating a more robust memory trace than the more passive process (i.e., subvocal rehearsal or listening). MacLeod et al. (2010) conducted a series of experiments to investigate the production effect principle. Experiment 1 was designed to establish the primary production effect by comparing memory performance for words read aloud versus words read silently. The researchers employed a within-subject design where participants studied a list of words, with half of the words read aloud and the other half read silently. Following the study phase, participants completed a surprise recognition test. The results indicated a significant production effect: words read aloud were remembered better than those read silently. This finding demonstrated that vocalizing words during the study phase enhances memory retention compared to silent reading. The researchers attributed this effect to the distinctiveness of the produced items, which likely receive more robust encoding due to the additional articulatory and acoustic processing involved in vocalization. This initial experiment provided a foundational understanding of the production effect, highlighting the memory benefits of reading words aloud and setting the stage for further investigations into alternative forms of word production and the conditions under which the production effect occurs.

In Experiment 2, which comprised three sub-experiments (2A, 2B, and 2C), the researchers examined whether the production effect would apply to alternative production modes such as writing, mouthing, and whispering. The results demonstrated a consistent pattern across all three sub-experiments: words read aloud were remembered significantly better than those produced by writing, mouthing, or whispering, which were remembered better than words read silently. Notably, while the production effect was robust in within-subject designs, it was not observed in between-subject designs. This finding suggests that the effect relies on the distinctiveness of produced items relative to non-produced items within the same study list. The lack of significant interaction between study conditions and specific experiments indicated that this pattern remained consistent across the sub-experiments. These results extend our understanding of the production effect, showing that while alternative production forms offer some memory benefits compared to silent reading, they do not match the effectiveness of reading aloud. The researchers concluded that the production effect is primarily tied to the distinctive articulatory and acoustic properties of reading words aloud. Other production forms, such as writing, mouthing, or whispering, provide some unique information during the encoding process, but not to the same extent as reading aloud. This gradation of effectiveness in different production methods highlights the specific importance of vocalization in enhancing memory retention through the production effect.

In Experiment 3, MacLeod et al. (2010) aimed to further explore the production effect by investigating whether it extends to non-words and persists over a longer retention interval. The researchers used a within-subject design where participants studied both words and non-words, with half of each type read aloud and half read silently. The study phase was followed by a recognition test administered immediately or after a two-day delay. Results showed that the production effect was present for both words and non-words, with items read aloud being recognized more accurately than those read silently. This effect was observed in both immediate and delayed testing conditions, demonstrating the robustness of the production effect over time. Importantly, the magnitude of the production effect did not differ significantly between words and non-words, suggesting that the benefit of production is not dependent on the meaningfulness or familiarity of the stimuli. The persistence of the effect over a two-day delay indicates that the memory advantage conferred by production is not merely a short-term phenomenon but can enhance long-term retention. These findings extend our understanding of the production effect by showing its applicability to non-lexical stimuli and its durability over time, further emphasizing the decisive role of active vocalization in memory encoding and retrieval processes.

In experiments 4 and 5, the researchers examined whether different types of responses to items (i.e., a key press in experiment 4 and saying "yes" in experiment 5) contributed to better memory performance. Results did show a production effect. This implies that it is not merely the vocalization of the item itself but also the distinctiveness of the response that contributes to the production effect. In experiment 6, the researchers compared memory performance for words and nonwords that were orally vocalized (read aloud) to those that were read silently. The observed memory benefit was restricted to aloud-produced words and nonwords, where orally vocalized items were better remembered than their silently read counterparts. This finding further reinforced the importance of oral vocalization in the production effect, demonstrating that reading aloud enhances memory performance compared to silent reading for both words and nonwords.

The last two experiments revealed further dimensions of the production effect, which examined additional encoding conditions impacting memory performance. In Experiment 7, the production effect was tested by incorporating an elaborate encoding task known to improve memory performance: the generation effect, which involves the benefits of self-generated words on memory (Slamecka & Graf, 1978; Bertsch et al., 2007). Surprisingly, even with this more efficient encoding task, the production effect persisted, suggesting that vocalization provides an additional memory benefit beyond the generation effect. This indicates that vocalizing items enhances the retention of previously learned material, independent of the advantages conferred by self-generation.

Lastly, experiment 8 explored the possibility that the production effect arises from inferior encoding of silently read items due to "lazy reading" (Begg & Snider, 1987). This hypothesis posits that the benefit for produced items does not stem from superior encoding but rather from the inferior encoding of silent items due to reduced attention. Participants engaged in a semantic judgment task (abstract/concrete decision) for orally vocalized and silently read items to test this. The essential conditions involved participants reading words aloud while making a semantic judgment and reading words silently while making the same judgment. The results showed an apparent production effect: items read aloud were remembered better than items read silently, even when both items were processed with the same level of semantic judgment. This finding refutes the notion that the production effect arises solely from lazy reading practices. Instead, it suggests that the production effect is not merely a byproduct of inferior encoding of silent items but is influenced by the distinctiveness and additional cognitive processing involved in producing the words aloud.

Taken together, the series of eight experiments by MacLeod et al. (2010) (summarised in Table 2.2) delved into the production effect, revealing several key findings. They demonstrated that reading words aloud significantly enhances memory compared to reading silently. This benefit, which applied to both words and pronounceable nonwords, did not extend to other forms of production, such as writing, mouthing, or whispering. The production effect remained robust even when participants made semantic judgments on all items. Generation tasks further enhanced it, indicating that it benefits from distinctive articulatory information and deeper semantic processing. These results suggest that the production effect arises from the combined influences of unique vocalization features and meaningful semantic analysis, enhancing the memorability of produced items.

Table 2.2: Summary of MacLeod et al. (2010) experiments

Experiments	Design	Conditions	Items	Test	Main Findings
Experiment 1A: 23 students Experiment 1B: 21 students	Within-subject, mixed-list	1A: read aloud, read silently 1B: read aloud quickly, read silently	120 words	Yes/no recognition test.	Positive PE
Experiment 2A: 15 students Experiment 2B: 15 students Experiment 2C: 15 students	Between-subjects, pure-list	Between-subject: 2A: Read aloud vs. writing vs. read silently 2B: Read aloud vs. mouthing vs. read silently 2C: Read aloud vs. whispering vs. read silently	80 words	Yes/no recognition test.	No PE
Experiment 3A (within): 17 students Experiment 3B (between): 15 students per condition	**3A**: Within-subject **3B**: Between-subjects	3A: read aloud, read silently 3B: read aloud quickly, read silently	3A: 88 words 3B:80 words	Two-alternative recognition test.	3A: Positive PE 3B: No PE
Experiment 4A: 24 students Experiment 4B: 24 students	Within-subject	4A: keypress, read silently 4B: say yes, read silently	120 words	Yes/no recognition test.	No PE
Experiment 5: 24 students	Within-subject	read silently and mouth the words	120 words	Yes/no recognition test.	Positive PE
Experiment 6: 21 students	Within-subject	read aloud, read silently	160 pseudowords	Yes/no recognition test.	Positive PE
Experiment 7: 35 students	Within-subject	Generate aloud, generate silently	120 words	Yes/no recognition test.	Positive PE
Experiment 8: 27 students	Within-subject	Semantic decision (living vs. non-living), read aloud, read silently	120 words	Yes/no recognition test.	Positive PE

2.2.1.1 Distinctiveness and Cognitive Processing

As explained earlier, the production effect (PE) refers to the phenomenon where saying or producing words aloud leads to better memory for those items than words read silently. According to Lin and MacLeod (2012), this advantage occurs because producing words aloud creates distinctive cues that enhance later memory retrieval. The Lin and MacLeod study found that these distinctive cues are generated by vocalisation and silently mouthing words, suggesting that motor processes contribute to the effect. This implies that both vocalisation and related motor activities produce unique information that aids in memory retrieval. The study supports the distinctiveness account of the production effect, indicating that the enhanced memory performance is due to the creation and utilisation of distinctive cues during the encoding process. Individuals monitor their memory for these distinctive cues linked to the words they vocalised or mouthed, leading to a memory advantage for these items over those processed silently. The persistence of the production effect in older adults suggests that the ability to use these distinctive cues for memory enhancement remains intact with ageing, even though the benefit might be somewhat reduced compared to younger individuals. Specifically, research by Ozubko et al. (2012) demonstrated that while both younger and older adults experience a production benefit in memory, the magnitude of this benefit is more diminutive in older adults. This finding indicates that although older adults can still utilise distinctive cues generated by vocalisation, their efficiency in doing so is reduced compared to younger individuals.

The production effect in memory is primarily attributed to the interaction between the phonological loop and distinctiveness. The phonological loop, a component of the working memory model proposed by Baddeley and Hitch (1974), is responsible for temporarily storing and rehearsing verbal and phonological information. When words are read aloud, they receive direct phonological encoding, making them easier to store and rehearse, resulting in a more robust phonological representation than words read silently. Additionally, the distinctiveness account suggests that words read aloud become more distinctive relative to silently read words, also enhancing memory retention. This concept aligns with the von Restorff effect, which posits that memories of an event are more vivid when the event stands out within its context (von Restorff, 1933; Hunt, 1995). According to Hunt (1995, 2006), distinctiveness triggers specialized cognitive processing, leading to better memory.

The association between distinctiveness and enhanced cognitive processing is not confined to specific items. Instead, it occurs whenever differentiation exists within a set of items. This has been supported by studies from Acheson and MacDonald (2009), Icht and Mama (2015), and Gionet et al. (2022). Hunt (2006) asserts that distinctiveness is a fundamental theoretical mechanism underlying various memory phenomena. Distinctiveness contributes to the production effect, which arises from the combined influence of the phonological loop's robust phonological representation and the enhanced distinctiveness of read-aloud words. This interaction between the phonological loop and distinctiveness processes provides a comprehensive explanation for the memory advantage observed for produced words over silently read words.

Lockhart et al. (1976) asserted that richer encodings are more unique and distinct than shallow ones. A substantial body of research confirms that distinctiveness enhances memory, as demonstrated by studies from Conway and Gathercole (1987), Gathercole and Conway (1988), Dodson and Schachter (2001), Hunt and Worthen (2006), Hourihan and MacLeod (2008), Ozubko and MacLeod (2010), and Ozubko et al. (2011). This evidence supports the distinctiveness account, suggesting that during testing, individuals retrieve unique information stored in memory linked to the items they have vocalized, resulting in an advantage over items merely processed silently. The distinctiveness advantage can also be understood through dual coding theory, which posits that richer encodings involve verbal and visual elements, leading to more robust memory traces. The benefit of vocalization is attributed to the distinctiveness of the encoding and the presence of multiple codes (verbal and possibly auditory), which facilitate better recall. For instance, a study by Lin and MacLeod (2012) examined the memory abilities of older adults, who often experience challenges when utilizing distinct information. Interestingly, despite these challenges, older participants exhibited a notable production effect, with the magnitude of this advantage being similar to that observed in younger adults. These finding highlights that the production effect's benefits persist across age groups, indicating that the distinctiveness of vocalized information aids memory performance irrespective of age.

2.2.1.2 Memory Mechanisms, Retrieval Processes, and Their Role in the Production Effect for Vocabulary Learning

As explained earlier, the Production Effect (PE) has emerged as a significant phenomenon in cognitive psychology and educational research, particularly concerning memory and vocabulary retention. The PE refers to the enhanced recall and recognition of words when spoken aloud during learning, compared to reading them silently (MacLeod, 2011). This effect can be understood within broader frameworks of memory encoding, rehearsal, and retrieval, which are fundamental components of cognitive functioning.

Memory encoding is pivotal in determining how effectively information is retained and later recalled. Carpenter & Olson (2012) and Carpenter & Geller (2020) note that encoding involves the initial reception of sensory information, whether semantic, visual, or acoustic, before it is stored for future retrieval. The Production Effect directly enhances this encoding process by simultaneously engaging multiple sensory and cognitive systems. Learners who speak words aloud activate their visual processing systems (as in silent reading) and their auditory and motor systems. This multi-sensory engagement aligns with Paivio's Dual-Coding Theory (1971), which posits that information processed through verbal and non-verbal channels produces more robust memory encoding.

2.2.1.2.1 Retrieval

One of the fundamental functions of memory is to comprehend, encode and retain new learning materials and retrieve them to enhance learning (Karpicke et al., 2009; Karpicke and Blunt, 2011). During learning, information is transferred from short-term memory to long-term, a process strengthened by actively retrieving and rehearsing the information. Cues in the environment (e.g., visual or auditory) trigger the retrieval of memory traces associated with recalled contexts. This process utilizes multifaceted methods, such as **recall** (i.e., generating information without prompts), **recognition** (i.e., identifying previously encountered information), and **relearning** (i.e., re-encountering and strengthening previously learned information), to access and retrieve stored knowledge from the long-term memory system. (MacLeod & Dunbar, 1988; Carpenter, 2011; Roediger and Butler, 2011; Rowland, 2014; Antony et al., 2017).

Retrieval cues are affected by various aspects, including an individual's cognitive abilities and the learning environment that aids the recall process, either explicitly provided (e.g., when teachers ask students to memorise new vocabulary), self-generated (e.g., linking learning to the environment), or encountered incidentally (e.g., learning a new word from a song) (Pansky et al., 2005). In other words, visual images, key terms, verbal prompts, or any other stimulus aid in accessing and retrieving long-term memory information can be used as cues.

The process of retrieving information from memory may be complex. However, it becomes more straightforward in situations similar to the original context in which the information was encoded. This phenomenon, known as context-dependent memory, occurs because retrieval cues activate memory networks established during encoding, making these cues more effective. When the context or sensory details of the retrieval situation match those of the encoding situation, the brain has a more robust pathway to access the stored information (Cowan, 2008; Divjak, 2019). This principle, termed encoding specificity by Tulving and Thomson (1973), suggests that the effectiveness of a retrieval cue depends on how well it matches the cues present during the encoding process. According to Roediger and Butler (2011), for retrieval to be efficient, effective cues must be developed that will allow the encoded information to be remembered. These cues can include environmental factors, emotional states, or even internal cognitive states that were present during the initial learning experience.

The contextual retrieval hypotheses emphasize the role of environmental factors, emotional states, and attitudes in facilitating memory retrieval, indicating that when encoding and retrieval contexts are aligned, recall is more effective (Tulving & Thomson, 1971, 1973; Godden and Baddeley, 1975; Gruneberg and Morris, 1994). An example of this proposition can be found in language acquisition, where the context of the learning has a significant impact on the retention of vocabulary and grammar rules (Long et al., 2001; De Keyser, 1998; Norris and Ortega, 2000; Ellis, 2001; Laufer, 2005, 2006; Schmitt, 2010). Due to the congruence between the encoding and recall settings, immersive language learning environments are more effective for second language learners (Wilkinson, 1998; Cummins, 2009; Kinginger, 2011). That is, immersive language learning environments are more effective for second language learners due to the congruence between encoding and recall settings. In traditional learning environments, learners often encode information in contexts far removed from real-world language use. Conversely, immersion places learners in situations where the target language is constantly used. This reinforces the encoding process by creating strong associations between the language and the context in which it is used. As a result, retrieval of the learned language becomes more efficient and accurate because the cues in the immersive environment directly activate the relevant memory networks established during encoding. This fosters more substantial memory traces and ultimately improves retention and fluency.

In this view, the learning environment closely matches the use context, providing learners with multiple relevant retrieval cues specific to language learning. These cues can include visual cues (e.g., seeing a restaurant menu with pictures and corresponding vocabulary in the target language), auditory cues (e.g., hearing a native speaker order food at a restaurant and using specific grammar structures), and situational cues (e.g., needing to ask for directions on the street and using learned phrases for navigating). This reinforces the encoding process by strengthening the associations between the learned information (vocabulary, grammar) and the context it encounters (ordering food, asking for directions). Consequently, memorization and recall are improved because reencountering the information in a similar context triggers the retrieval of multiple memory traces linked together during encoding.

Many factors can influence memory performance in storing and retrieving words or vocabulary. For instance, studying words accompanied by pictures has a greater effect on memory performance than studying vocabulary (Carpenter & Olson, 2012; Paisart & Suriyatham, 2017; Andrä et al., 2020; Carpenter & Geller, 2020). Also, semantically related words (e.g., bread and butter) are better remembered than words studied in isolation from other words (Hunt, 1995, 2006). The earliest studies of memory and storing information found that when relating the environment of study to that of the test, memory performance improves, as the match between the participant's state or condition during the study and during the test causes them to recall information more quickly (e.g. Goodwin et al., 1969; Rogers et al., 1977). In essence, these findings highlight the importance of interconnectedness in memory. When information is presented with relevant pictures, related words, or studied under similar conditions, it creates stronger connections in the brain (i.e., retrieval cues), leading to better recall of information (as discussed in the following section). Building on these foundational insights, subsequent research expanded to explore a range of factors affecting the encoding and retrieval of information. Amongst these, the effect of production on memory encoding and recall is particularly noteworthy area in the current thesis.

2.2.1.2.2 Recall

In vocabulary studies, recall and recognition are essential for assessing vocabulary knowledge and retention (Spencer et al., 2016; Nation, 2016; Schmitt et al., 2020). Recall refers to the ability to retrieve and produce words from memory, while recognition involves identifying or acknowledging known words when presented with them (Krishnan et al., 2017). In order to indicate the memory performance, retained information or the amount of encoded information, memory is tested through various techniques including recalling and recognition. One of the primary ways to measure retention involves recalling information immediately after learning or at specified retention intervals (i.e., one week or a few weeks after learning) (Roediger, 1990; McKinney and Woodward, 2004; Lei and Shuyuan; 2005; Levine et al., 2009). According to Hollingworth (1913), "Recall is that aspect of memory process in which a setting, a background or association-cluster, is present in clear consciousness, but a desired focal element is missing" (p.533). This "focal element" refers to the specific information sought during recall but remains elusive. For instance, an individual might vividly recollect the context surrounding a birthday party (i.e., setting), including visual details like balloons and cake (i.e., association cluster), yet be unable to retrieve the name of the person whose birthday it was (i.e., focal element). Retention can be achieved through specific recall tests, such as memorizing a set of words and asking the learner to recall words they remember. When prompted, recall accuracy determines the retained amount, and the response time of accurately recalled items also serves as an indicator (i.e., the shorter the interval, the stronger the retention) (Carrier & Pashler, 1992; Amin & Malik, 2014).

There are three main types of recall: free recall, cued recall, and serial recall. Free recall involves recalling information without cues or hints (e.g., recalling newly learned words in any order), while cued recall, as explained earlier, uses cues or hints to aid in the recall process (e.g., using colours or semantic features between words to remember an item). Serial recall, on the other hand, involves recalling information in the sequence of its occurrence (e.g., recalling the sequential occurrences within a novel or recalling a list of words in the order they were presented, such as “dog, cat, apple, house”) (Carpenter et al., 2006; Nieuwenhuis-Mark, 2012; Wang and Gennari, 2019; Beyon and Coc-Boyd, 2020).

These forms of recall (e.g., cued recall, serial recall) have been studied in both humans and animals (e.g., Botvinick et al., 2009) to shed light on the underlying mechanisms of memory retrieval. Researchers can make inferences about the cognitive processes involved by observing how these different species perform on recall tasks. For instance, if humans and animals exhibit similar patterns of retrieval difficulty for specific tasks, it suggests a more fundamental mechanism shared across species. The recall mechanism involves a dual-stage mechanism. In the first stage, retrieval cues trigger a search through stored information. The second stage involves a decision-making process, where the retrieved information is evaluated to identify the correct memory among competing candidates. This complex process relies heavily on cues and contextual factors. The more prosperous and specific the available cues, the more efficient and accurate the retrieval process becomes, ultimately enhancing the success of subsequent recall.

2.2.1.2.3 Recognition

Recognition, on the other hand, refers to the ability to remember and distinguish an encountered item (e.g., people, places, words, sounds, and objects) as having been presented previously (Standing, 1973; Gardiner and Parkin, 1990; Medina, 2008). The ability to recognize past events is attributed to their representations being encoded in the memory. When an event is reexperienced, environmental information is matched with stored memory representations, which results in matching signals being elicited (Mandler, 1980; Wixted, 2007; Wang et al., 2017). Studies have implemented various methods to assess recognition in memory, including yes/no tests, multiple-choice assessments, and association tasks (Bayley et al., 2008). Two main underlying mechanisms explain recognition in memory: recollection of previous experiences and item familiarity (Yonelinas, 2002; McKenzie and Tiberghien, 2004; Yonelinas et al., 2005; Pratte and Rouder, 2012).

Recollection is referred to as “remembering” as it involves retrieving details associated with previously experienced events, while familiarity is referred to as “knowing” as it represents the sensation of having previously experienced the event (Jacoby, 1991; Yonelinas et al., 2022). This distinction aligns with the broader categories of declarative and procedural memory. Declarative memory refers to the conscious recollection of facts and events, similar to the conscious retrieval of details involved in recollection. Procedural memory, on the other hand, deals with the implicit knowledge of skills and habits, which aligns with the feeling of familiarity without specific details. This suggests that recollection and familiarity rely on different brain regions and retrieval processes, further supporting their distinct nature (Squire & Dede, 2015).

While studies agree that these processes or memory assessment mechanisms have contributed to describing recognition in memory, they have used the terms "recollection" and "familiarity" in varying ways. Occasionally, some describe familiarity and recollection as singular and distinct processes that occur separately (e.g., Squire et al., 2007; Yassa and Stark, 2011), while others define the familiarity-recollection process as a dual-process that occurs simultaneously but along a continuum, where recognition involves a continuous interplay between the feelings of familiarity and the conscious recollection of specific details or contextual information (e.g., Yonelinas, 1994; Yonelinas et al., 1998). However, recollection is fundamentally a slow, controlled search process. In contrast, familiarity is a fast, automatic process that identifies mnemonic information (i.e., previously encoded information that familiarity can detect without retrieving specific details) from non-mnemonic information (Tulving, 1984; Ingram et al., 2012; Yonelinas et al., 2022).

The role of specific detail retention in language learning has received growing research attention. Studies have examined how memory processes, particularly recognition and recollection, influence language acquisition in children. For instance, Mirandola et al. (2011) investigated these processes in ninth and tenth graders with and without reading difficulties. Students listened to a narrated story in a classroom setting, mimicking a real lesson. After exposure, researchers assessed memory through a recognition task where students distinguished previously encountered sentences from new ones.

Additionally, participants provided "Remember-Familiar" judgments for recognized sentences, indicating their subjective experience of recollection or simple familiarity. The findings revealed a link between students' reading abilities and their judgments of story familiarity. Notably, students across both groups exhibited weak recollection, suggesting that retrieving specific details from the text posed a challenge (Mirandola et al., 2011). This aligns with other research highlighting the importance of studying memory functions like recall and recognition in children's language development (e.g., Blankenship et al., 2018; Pelegrina et al., 2023).

There is a degree of commonality between recall and recognition. This connection was found when testing recall and recognition, for example, for famous surnames (Tulving and Wiseman, 1975; Muter, 1984), when testing patients with amnesia (Giovanello and Verfaellie, 2001), and when testing word retrieval (Watkins and Todres, 1987; MacLeod and Kampe, 1996). In other words, recall and recognition use the same retrieval process, with differences primarily related to cue availability. Recognition tasks typically offer more specific cues (e.g., multiple-choice options) than free recall, which relies on internally generated cues (e.g., phonological and semantic cues).

The continuity hypothesis (Tulving and Watkins, 1973) challenges the notion that recall and recognition are fundamentally separate memory processes by suggesting that they result from past and present information. In this context, Uner and Roediger (2022) found that continuity was reflected in the experience of recollection, but familiarity was not linearly related to cue condition. However, recall and recognition differ in how they retrieve information from memory. Recall primarily depends on recollection, the ability to consciously retrieve specific details from memory. Conversely, recognition relies on a confluence of recollection and familiarity. It arises from a sense of having encountered something before, even if the specific details remain elusive.

2.2.1.2.4 Distinguishing Recall and Recognition in Vocabulary Learning

Memory and learning are deeply interconnected, especially in acquiring vocabulary. Recall and recognition are two fundamental processes in memory retrieval that play distinct roles in vocabulary learning. While recall requires actively retrieving words from memory without cues, recognition involves identifying familiar words when they are presented again (Krishnan et al., 2017). According to Tulving and Pearlstone (1966), memory can be categorised as either available or accessible—while all encoded information is theoretically available, only some can be actively retrieved. This distinction is crucial in vocabulary learning, where students often find it easier to recognise vocabulary in multiple-choice formats but struggle with open-ended tasks that require active recall.

The difference in retrieval efficiency between recall and recognition can be explained by the encoding specificity principle (Tulving & Thomson, 1973), which suggests that memory retrieval is most successful when the retrieval cues match how the information was initially encoded. In vocabulary learning, this means that if a word was learned in a specific context or through a particular method, retrieval will be easier if similar cues are present during testing. However, interference can occur when multiple words or similar contexts are stored in memory, complicating retrieval (Nairne, 2002; Watkins, 1975). This is particularly relevant for second language learners or students learning vocabulary across different subjects.

2.2.1.2.5 The Role of the Production Effect (PE) in Vocabulary Recall and Recognition

The Production Effect (PE) plays a significant role in recall and recognition tasks (Ozubko & MacLeod, 2010). In vocabulary acquisition, the PE enhances memory by engaging multiple sensory modalities, auditory, visual, and motor, when words are vocalized during encoding. This multisensory engagement creates more distinctive memory traces, making retrieval more efficient and improving recall and recognition.

In the context of recall, the PE leads to a deeper processing of the vocabulary being learned, as vocalization acts as a form of rehearsal that strengthens memory traces. According to Bodner and Taikh (2012), when learners produce words aloud during encoding, they are more likely to recall them in free-recall tasks because production makes the memory trace more robust and more accessible. This effect is especially useful for vocabulary learning, where active recall plays a critical role in consolidating word knowledge. For example, when students are asked to produce newly learned words during study sessions, they are more likely to recall them when prompted in later tests, improving long-term retention.

The PE also offers significant benefits in recognition tasks, mainly when words are presented in mixed-list designs, where some words are vocalized, and others are read silently (Forrin et al., 2012; Jones & Pyc, 2014). In such designs, the vocalized words stand out due to their distinctiveness, making them easier to recognize later. The distinctiveness hypothesis suggests that vocalization enhances recognition by creating more robust and unique memory representations, distinguishing produced words from silently read ones (Bodner et al., 2014). This is particularly valuable in vocabulary tests where learners must recognize correct word forms, definitions, or translations among distractors.

However, in pure-list designs, where all words are processed the same way (either all vocalized or all silently read), the PE is diminished because the distinctiveness between items is lost (Forrin & MacLeod, 2016). This suggests that the PE is most effective when learners alternate between vocalizing and silently processing words during study sessions, maximizing the benefits of distinctiveness for recognition.

2.2.1.2.6 Implications for Vocabulary Learning

The Production Effect offers practical implications for vocabulary acquisition strategies. By incorporating vocal production into vocabulary study routines, learners can leverage vocalization's distinctiveness and sensory reinforcement to improve both recall and recognition. For example, a student who reads a vocabulary list aloud during study sessions will not only encode the words more deeply but also create distinct memory traces that make both active recall (e.g., writing the word from memory) and recognition (e.g., selecting the correct word from a multiple-choice list) more efficient.

The connection between recall, recognition, and the Production Effect highlights the importance of varied practice in vocabulary learning. Encouraging learners to engage in both vocal production and silent reading during study sessions can optimize their ability to recall words in open-ended tasks and recognize them in recognition-based assessments. By aligning encoding strategies (such as vocalization) with the expected retrieval conditions, learners might benefit from the encoding specificity principle, further enhancing their vocabulary retention and retrieval success.

2.2.1.3. Variability in Experimental Designs

The Production Effect (PE) has been observed through various experimental designs, but its impact tends to be more pronounced and consistent in within-subjects designs than in between-subjects designs. In within-subject studies, such as those conducted by MacLeod et al. (2010), Forrin et al. (2012), and Quinlan and Taylor (2013), participants experience both conditions (e.g., reading aloud and reading silently), allowing for a direct comparison of the two. This design reliably demonstrates a significant recognition advantage for words produced aloud. For instance, Hopkins and Edwards (1972) found a strong recognition effect in within-subjects designs, where participants more easily recognized words they had vocalized. In contrast, between-subjects designs divide participants into separate groups, each experiencing only one condition (e.g., reading aloud or reading silently). This setup can obscure the detection of PE due to individual differences among participants that may influence memory performance. The Production Effect is less consistently observed in these designs, highlighting the challenges of detecting PE when participants cannot directly compare the two conditions. Roediger and McDermott (1995) support the idea that within-subject designs demonstrate a more substantial memory effect for produced words, particularly when words are semantically related. Their study reinforced that produced words (e.g., spoken aloud) benefit from deeper encoding processes and are better recognized than silently read words. Other research, such as studies by Gathercole and Conway (1988) and Ozubko and MacLeod (2010), further confirms that words produced aloud are better recalled in within-subjects mixed-list designs, where the distinctiveness of vocalization is a crucial factor. These findings suggest that within-subject designs are particularly effective for detecting PE because participants can directly contrast the distinctiveness of produced and non-produced items. Pure-List vs. Mixed-List Presentation

To clarify, pure-list presentations involve participants being exposed to lists containing only one type of word frequency (high-frequency or low-frequency) leading to apparent performance differences based on familiarity. In contrast, mixed-list presentations combine word frequencies within the same list. This intermixing can mitigate the exaggerated performance differences typically seen in pure-list designs by simultaneously forcing participants to process items of varying familiarity. Despite the difficulty in detecting PE in between-subjects designs, the effect has still been observed under certain conditions. For example, Bodner and Taikh (2012) and Fawcett (2013) challenged the idea that PE is exclusively a within-subject phenomenon. Fawcett's research advanced the understanding of PE by demonstrating that it can persist in between-subject paradigms. In these studies, participants who vocalized words still showed a memory advantage compared to those who silently read words despite not having the opportunity to contrast the two conditions directly. This suggests that the Production Effect extends beyond just distinctiveness and may involve deeper memory encoding processes that are effective across different experimental contexts. Icht et al. (2014) proposed that the strength of the Production Effect in between-subject designs might be due to more than just word distinctiveness. They argued that producing words aloud introduces an additional sensory element—auditory feedback—that enhances memory encoding. This multimodal encoding process engages both visual and auditory stimuli, creating a richer memory trace that aids in both recall and recognition. During a recognition test, participants not only remember the word itself but also recall the act of vocalizing it, which provides an additional retrieval cue. This multifaceted approach to encoding strengthens memory retention even in experimental designs where direct comparison of produced and non-produced words is impossible.

In summary, while the Production Effect is more reliably observed in within-subjects designs due to direct comparisons between conditions, it is not limited to these designs. Research has shown that even in between-subject paradigms, where participants experience only one condition, the multimodal nature of producing words aloud enhances memory encoding, leading to better recall and recognition. This underscores the broader applicability of the Production Effect across different types of experimental designs, highlighting its significance in memory research.

2.2.1.4. Familiarity and the Production Effect in Word Learning

Contrary to the distinctiveness account, the production effect varies with the level of familiarity. The PE is observed in adults only when familiar words and native phonemes are used (Fawcett, 2013; Kaushanskaya and Yoo, 2011; MacLeod, 2011; Mama & Icht, 2016; Ozubko and Macleod, 2010; Roediger and McDermott, 1995), other studies show a disruptive effect (i.e., a reversed PE: Better recall for words learned silently) that seems to appear when the stimuli presented to participants are novel words with non-native phonemes (Kaushanskaya and Yoo, 2011; Cho and Feldman, 2013, 2016; Dahlen and Caldwell-Harris, 2013; Baese-Berk and Samuel, 2016; Zamuner et al., 2016). For example, Kaushanskaya and Yoo (2011) conducted a study involving adult participants, focusing on teaching them pseudowords that include native or non-native phonemes. The participants were trained under two conditions: repeating the words aloud or silently. The results indicated that participants exhibited higher proficiency in recalling pseudowords with non-native phonemes when learning silently. The study identified that the lack of phonological familiarity disrupted the learning process. Repeating the words aloud prompted learners to allocate more attention to the phonological structure of the word. When studying words with native sounds, vocal production highlighted similarities between the novel words and the participant's native language. This allowed learners to use their L1 existing language knowledge, facilitating word learning and establishing robust, distinct mental representations. Vocal production supported the learning process in this context.

Conversely, the learning process regarding non-native sounds could have been improved. Participants lacked relevant information to aid in the learning task, making it inherently more challenging to grasp and retain these non-native phonemes compared to the familiar native sounds. In simpler terms, introducing non-native sounds means participants have more to learn. They need to learn not only new words but also the new phonemes that make up those words. This dual challenge of acquiring unfamiliar sounds and new vocabulary simultaneously increases the cognitive load on learners, potentially impacting their ability to process and retain information effectively. Baese-Berk and Samuel (2016) investigated the effects of production on perceptual learning of non-native speech sounds. They trained native Spanish speakers to discriminate a Basque sound contrast (i.e., a non-native phonetic distinction), with participants divided into two groups: perception-only and perception + production. The perception-only group showed significant improvement in discrimination, while the perception + production group did not show reliable improvement, and their production accuracy did not correlate with discrimination performance. The study concluded that overt production during perceptual training impaired perceptual learning, indicating an antagonistic relationship between perception and production and highlighting that task difficulty and prior linguistic experience lead to different learning outcomes.

These studies, however, require evaluation for their generalizability. It should be considered that most word learning research relies heavily on immediate testing rather than delayed tests. Word acquisition unfolds over a developmental timescale rather than immediately (Twomey et al., 2022), necessitating consideration of word learning variations based on delay length. The production effect (PE) has primarily been investigated concerning its impact on memory, with a limited exploration of its effects on the durability word learning. While word learning inherently involves memory processes, the distinction lies in the focus on the durability of learning. Many of the studies discussed above indeed pertain to learning new words; however, they primarily examine immediate memory tests rather than the long-term retention of learned words. Therefore, to determine the extent to which production can be attributed to these factors, it is essential to include delayed tests to assess the durability of learning.

In this vein, Krishnan et al. (2017) conducted a study examining the effects of recall (i.e., testing oneself on newly learned word associations), reproduction (saying aloud), and restudy (further exposure) on the learning and retention of novel spoken words. Initially, both recall and reproduction methods demonstrated superior learning compared to restudy, consistent with established phenomena known as the "testing effect" and "production effect". However, a week later, there was no discernible difference in retention among the three conditions, contradicting the anticipated recall advantage for long-term retention. The authors suggested potential factors contributing to this unexpected outcome, such as the study's design and using pseudowords rather than familiar vocabulary. Despite expectations, the recall did not exhibit a long-term advantage over restudy, challenging prior research findings. Specifically, while recall initially facilitated better learning, it failed to confer lasting benefits compared to reproduction. Statistical analyses confirmed the absence of a disparity in long-term retention between recall and reproduction conditions. In conclusion, although recall and reproduction yielded initial advantages, they did not diverge significantly in their effectiveness for retaining novel spoken words over time in this study. Cued recall, which refers to recalling information with cues or prompts, did not demonstrate superior long-term retention compared to reproduction.

In a recent study by Icht and Mama (2022), conventional patterns of the production effect were observed across immediate, one-week delayed, and two-week delayed tests. While participants learning a second language demonstrated greater accuracy in immediate testing, a more pronounced production effect was noted during the two-week delayed test. This production effect refers to the enhanced memory performance for words that were spoken aloud compared to those that were read silently. Specifically, participants showed a greater difference in their ability to recall words spoken aloud versus silently after two weeks, suggesting that they not only learned the new vocabulary initially but also retained it effectively over an extended period. This indicates the long-term benefits of active engagement in language learning, such as speaking, on memory retention.

Similarly, Kapnoula and Samuel (2022) investigated the effects of production on word learning, aiming to reconcile contradictory findings in the literature. Their key findings are that production initially facilitates word learning but later impairs it compared to just listening. Delaying production practice until after a consolidation period can alleviate the detrimental effect of production on word learning. Specifically, they found that producing novel words immediately after exposure helped learning initially. However, after a delay, those who only listened showed better retention of the novel words than those who produced them. A follow-up study showed that delaying production practice until after a 24-hour consolidation period eliminated the long-term detrimental effect of production on word learning. Their findings suggest a time-dependent shift in the role of production, where it is beneficial for initial encoding but later becomes a liability for long-term retention of novel words. Delaying production until after consolidation can mitigate this detrimental effect. These results have implications for optimal word-learning strategies and shed light on the dynamic interplay between the production, consolidation, and long-term retention of novel linguistic information. Thus, the aim of the current thesis to investigate the durability of learning new words with L2 children.

2.2.1.5. Modality and the Production Effect in Vocabulary Learning

Modality is pivotal in how learners process, store, and retrieve lexical information. The two primary modalities, oral (speaking and listening) and written (reading and writing), engage distinct cognitive processes and learning strategies, influencing vocabulary acquisition differently. In this context, the Production Effect (PE), which emphasizes the advantages of active word production, particularly in oral tasks, provides a valuable framework for examining the role of modality in vocabulary acquisition.

Oral vocabulary learning typically involves speaking and listening tasks, where learners benefit from auditory and articulatory processes. These tasks have been found to enhance phonological encoding, improve listening comprehension, and provide real-time feedback, supporting more efficient vocabulary acquisition (Ellis, 2008). Learners in oral modality environments often experience better retention of pronunciation and word meaning because auditory input helps to create more robust memory traces (Vidal, 2011).

Conversely, written vocabulary learning focuses on visual processing and engages reading and writing skills. Written tasks allow learners to process words more slowly and thoroughly, which can be particularly beneficial for learning word spelling, orthography, and the semantic features (Hulstijn & Laufer, 2001). The visual nature of the written modality strengthens orthographic memory, which is essential for vocabulary recall.

The interaction between these modalities can also impact learning. Studies have shown that combining oral and written tasks may enhance vocabulary learning more effectively than either modality alone. For example, learners who read and speak new vocabulary tend to achieve better long-term retention (Barcroft, 2004). This suggests that multimodal engagement creates more robust memory traces through visual, auditory, and motor processes, enhancing vocabulary acquisition.

The Production Effect (PE) theory offers critical insights into the role of active oral production in enhancing memory for new information. The PE refers to the improved memory retention of words that are actively produced (spoken aloud) compared to those that are passively processed (MacLeod et al., 2010). In vocabulary learning, the PE highlights how speaking a word aloud creates distinct memory traces, contributing to better recall than silent reading or listening alone.

PE is closely tied to oral modality in vocabulary learning. Oral production activates multiple cognitive pathways (i.e., motor, auditory, and articulatory) that are not fully engaged during silent reading or written tasks. As a result, learners who practice vocabulary orally, particularly by speaking words aloud, are likely to experience the benefits of the PE.

Moreover, integrating written tasks with oral production may enhance the effectiveness of vocabulary learning. While written tasks focus on visual encoding, adding oral production reinforces auditory and motor encoding, creating a more comprehensive learning experience. For instance, a study by Ozubko and MacLeod (2010) demonstrated that learners who read words silently and spoke them aloud had superior recall compared to those who only read silently. This indicates that multimodal engagement, through writing and speaking, leads to more substantial memory traces and better vocabulary retention.

Despite the assumed benefits of oral production, written tasks also play an important role in vocabulary learning, particularly in literacy contexts where orthographic knowledge is critical. Learners may benefit from a balanced approach that combines the strengths of both modalities. Barcroft (2007) suggests that providing learners with opportunities to engage in both written and oral production can maximize vocabulary retention and retrieval, mainly when tasks involve multiple exposures to the target vocabulary in different contexts.

Thus, oral and written modalities offer unique advantages in vocabulary learning, with oral tasks enhancing phonological processing and written tasks improving orthographic memory. The Production Effect, which emphasizes the benefits of active oral production, aligns closely with the oral modality, suggesting that learners who speak words aloud during vocabulary learning will experience improved memory retention. However, integrating both modalities may offer the most robust learning outcomes, as it engages a broader range of cognitive processes. The current thesis explores how combining modalities in vocabulary instruction can enhance learning, especially in multilingual or second-language contexts.

In summary, the production effect underscores the interaction between distinctiveness and memory processing, emphasizing the benefits of vocalization for enhancing recall. Although distinctiveness and memory strength are key contributors to the production effect, the context and modality of production also play significant roles in its manifestation. The following section explores the production effect and its applications in children. Table 2.3 provides a summary of significant studies on the production effect in adults.

Table 2.3: Summary of Key Studies in Adult PE Literature

Source	Participants	Conditions	Methods	Test	Main Findings
Kaushanskaya and Yoo (2011)	Experiment 1: 22 adults Experiment 2: 24 adults	Read aloud, read silently	Experiment 1: 48 novel words English translations Experiment 2: 48 monosyllabic and disyllabic nonwords with 8 unique phonemes	Recall + Recognition	Experiment 1: Immediate: Positive PE Delayed: No PE Experiment 2: Immediate: Reversed PE Delayed: No PE
Zamuner et al. (2016)	Experiment 1: 29 undergraduate students Experiment 2: 30 undergraduate students	Heard; heard and produced 2: Corrected pronounced and mispronounced-heard and	16 CVC nonwords	Recognition	Positive PE
Krishnan et al. (2017)	36 undergraduates	Restudy; Reproduce; Recall; Immediate after training and a week after training	27 pseudowords with 27 unfamiliar pictures	Recognition + Recall	No PE
Icht and Mama (2022)	75 undergraduates	No production; and production	80 familiar bisyllabic Hebrew words	Recognition	Positive PE
Kapnoula and Samuel (2022)	Experiment 1: 40 students Experiment 2: 41 students Experiment 3: 41 students	Perception; production	8 CVC nonwords	Recognition	Experiments 1 and 2: Positive PE Experiment 3: No PE

2.2.2. The production effect in children

To date, only four studies have examined PE in early childhood (Icht & Mama, 2015; Zamuner et al., 2017; Pritchard et al., 2019; López López Assef et al., 2021). Similar to the adult literature, these studies show contradictory results on memory while none have looked at the effect of PE on L2 learning in children in the long term. The study by Icht and Mama (2015) was the first to investigate the production effect (PE) in childhood (5 years old) and found a positive impact of PE on memory for both familiar and unfamiliar L1 words. The study used a mixed-list design with three conditions: look-only, where children looked at the picture, and the word; look–listen, where children looked at the picture and heard the word; and look–say, where children looked at the picture and were asked to say the word they recognized. Results showed that the highest recall rate occurred in the produced condition (look–say), with 88% of children able to recall familiar words, compared to lower recall rates in the look-only and look-listen conditions. In a second experiment, researchers examined the PE in novel word learning by teaching children 30 low-frequency words rarely seen in children's vocabulary. Children were taught these words either by hearing them said by the experimenter or by hearing them and then producing them. The results demonstrated that recognition was higher for words that were both heard and produced, with 66% of children recalling unfamiliar words in the hearing and producing condition, compared to lower recognition rates for words that were only heard. This study highlights the significant role of active word production in enhancing memory recall in children for both familiar and unfamiliar words.

In contrast, the study of Zamuner et al. (2018) found a reversed effect of PE on memory for produced nonwords using two conditions in a mixed-list design: heard and heard then produced. The results indicate that the children's looking time to correct answers was longer in the heard condition on a preferential looking test. During the test, children's eye fixations were recorded while looking at stimuli from a mixed-list design that combines pictures and words from the two conditions. In the second experiment, the number of nonwords was increased, and a free-recall task was included in the preferential looking test; the results showed better performance in the heard condition only. Similar results were found in a follow-up study using a blocked-list design (Zamuner et al., 2017). These findings are similar to López Assef et al. (2021), who found that the PE was affected by age, such that older children (5 and 6 years old) showed a typical PE in a recalling task for familiar monosyllabic English words studied in three conditions: seen, heard, or produced. Conversely, a reversed PE was noticed in younger children (2-4 years) where they showed a better recall for words studied in the hearing condition. López Assef et al. (2021) explained the reversed effect in younger children as being due to their cognitive and linguistic development. Younger children may have less developed speech production abilities and memory integration processes, making it more challenging for them to benefit from the production effect. Instead, they rely more on auditory information, which is simpler and more consistent with their developmental stage. As children grow older and their cognitive and linguistic skills mature, they can better utilize the benefits of vocal production for memory recall.

The studies mentioned above combined spoken and illustrative stimuli in which reading was not included. However, Pritchard et al. (2019) took a different approach in which they used mixed- and blocked-list stimuli in within-subjects design experiments. In the first experiment, children between 7 and 10 years old were presented with familiar words on a computer screen to either read out loud or read silently. In the second experiment, another group of children was presented with longer lists of unfamiliar words using the same design as in the first experiment. The results demonstrated that recognition was better in read-aloud words for both words and nonwords in both the mixed-list and blocked-list designs.

Childhood PE studies have systematically focused on short-term memory (i.e., immediate recall or immediate recognition test. In studies examined the PE in children (Table 3), learning lists varied from a study to another to include familiar words or novel non-words (Zamuner et al., 2007; Icht & Mama, 2015; Zamuner et al., 2018; Pritchard et al., 2020; López Assef et al., 2021). Thus, while previous work on the PE has shown that the strength and direction of the PE depends on the linguistic characteristic of the stimuli (familiar native-sound, or non- words), the scope of the work has been restricted to the manipulation of unfamiliar lexical items with native sounds and the use of extensive learning lists with children. These approaches aim to assess memory retention rather than evaluating the durability of newly learned words.

Table 2.4: Summary of studies that have investigated the PE in children

Source	Participants	Conditions	Methods	Test	Main Findings
Icht and Mama (2015)	30 (5 years old)	Look (observation); Look and listen; Look and say.	30 pictures of Hebrew bisyllabic nouns (e.g., tiger, closet).	Free recall	Positive PE
Zamuner et al. (2018)	16 (4.5- 6.0 years old)	Heard; Produced	8 CVC novel words (e.g., wis, zel, jig, mig).	Recognition	Reversed PE for the younger age group and a positive PE for 6 years old children
Pritchard et al. (2019)	Experiment (1): 41 (7- 10 years old) Experiment (2): 40 children.	Read aloud- read silently	Experiment (1): 40 familiar printed words. Experiment (2): 80 monosyllabic novel words (e.g., hest, prench).	Recognition	Positive PE
Assef et al. (2021)	Experiment (1): 120 (3- 6 years old) (n=30 for each age group). Experiment (2): 30 (2 years old).	Look, Listen, Say	30 monosyllabic familiar English words (e.g., boat, bath, bird).	Free recall	Reversed PE for the younger age group (2-4 years old) and a positive PE for 5-6 years old children

2.2.3. Objectives of the Study

As outlined in Chapter 1, this thesis aims to address a significant gap in understanding the Production Effect (PE) in the context of second language acquisition among children, particularly concerning learning Modern Standard Arabic (MSA). While research on second language acquisition (SLA) in adults has established a connection between production and enhanced language learning, more is needed to know how production influences children's ability to retain newly learned vocabulary. The available studies exploring PE in child language learners are limited in scope, primarily focusing on short-term effects. Only four studies have examined PE in children, and none have assessed its long-term impact on vocabulary retention. This leaves a critical gap in the literature, particularly concerning how PE supports durable, long-term language acquisition in young learners.

This thesis seeks to address this gap by incorporating delayed tests, combining recall and recognition assessments, and investigating the long-term retention of MSA vocabulary. The study explores various learning strategies, including listening, repeating, and writing, to understand how these modalities influence memory retention and children's learning depth. By doing so, the research aims to provide new insights into how production facilitates immediate and sustained vocabulary learning in a second language context.

2.2.4. The Research Gap

Although PE has been studied in adult SLA and children's first language (L1) learning, the unique context of MSA acquisition in children still needs to be explored. Existing research has focused on short-term recall of familiar L1 words, with little attention given to the long-term retention of L2 vocabulary acquired through production-based learning. In particular, the specific linguistic environment of MSA presents a diglossic challenge, where children learn the formal variety (MSA) as a second language. At the same time, their colloquial dialect serves as the first language. Psycholinguistic research has yet to thoroughly investigate how young learners navigate this linguistic landscape and how production strategies might enhance vocabulary retention in such a context.

Additionally, the comparative impact of different production modalities (i.e., spoken versus written production) on memory retention in children has not been adequately addressed. The durability of PE effects beyond immediate testing remains an open question, especially in a multilingual environment where children may manage multiple languages simultaneously. This thesis responds to these gaps by introducing delayed recall and recognition tests to measure long-term retention and exploring the effects of various production methods. The results will provide a clearer understanding of how production strategies contribute to language learning and retention in child learners of MSA.

2.2.5. Research Questions

To guide this investigation, the following research questions have been addressed:

1. Does the Production Effect (PE) influence vocabulary recognition in Modern Standard Arabic (MSA) child learners?

2. What factors moderate the PE in this context, with a particular focus on age and language dominance?

3. compared to spoken production, does written production exert a similar influence on vocabulary recall in older child learners of MSA?

2.2.6. Hypotheses

Based on initial assumptions and the existing literature, this study proposes the following hypotheses:

1. Production Enhances Recognition: Children will exhibit better word recognition when actively repeating words than passive listening. This aligns with the established PE, which suggests that active production, such as vocal repetition, enhances memory encoding.

2. Age as a Moderator: Age may moderate production effects, with older children benefiting more from active repetition due to their more advanced phonological processing skills. Younger children, whose phonological loops are still developing, may not experience PE as intensely as their older counterparts.

3. Spoken vs. Written Production: Both spoken and written forms of production are expected to enhance recognition and recall, as production generally facilitates deeper information processing. This is consistent with research showing that active engagement with words—speaking or writing—strengthens memory traces.

4. Writing's Role for Older Children: Writing as a form of production may be particularly beneficial for older children. This may be due to their more developed motor skills and cognitive abilities, allowing them to fully engage with the word's written form. In contrast, younger children may find speaking a more natural and beneficial mode of production.

This thesis will test these hypotheses through experiments involving both recall and recognition tasks, focusing on how different production methods and age-related factors influence vocabulary retention in children learning MSA. Examining these variables will contribute to a deeper understanding of the cognitive processes underpinning second language learning in young learners, providing insights that could inform more effective language teaching strategies in diglossic and multilingual contexts.

Table 2.5: Overview of the Experiments

Study	Purpose of investigation	Conditions	Learning List	Participants	Test	Design
The pilot	To test the effect of the PE on various ages	Listen vs repeat	20 MSA words	80 students (4;0- 11;0)	recognition	Within-subjects mixed-list
Exp 1	A refined version of the pilot with longer learning list and controlling the age.	Listen vs repeat	64 MSA words	30 students (6;0- 6;9)	Free recall + recognition	
Exp 2	To test the effects of a different type of production (writing) and typical PE conditions (listening and repeating) when learning MSA as an additional language.	Listen vs repeat vs write	64 MSA words	30 students (8;0- 8;9 years old).	Free recall + recognition	

2.3. Conclusion

This chapter has explored extant research into memory and language learning, providing a comprehensive overview of cognitive processes involved in retaining and recalling new vocabulary. It also shows how different types of memory, such as sensory, short-term, and long-term memory, contribute to language acquisition. Moreover, it has delved into various models of memory, including the Working Memory Model and the Multi-Store Model, and their implications for understanding how learners process and store linguistic information. Additionally, the PE has been examined primarily in the context of memory, emphasizing its role in enhancing retention through active production. The discussion highlighted the importance of developmental stages in determining the effectiveness of PE, particularly concerning long-term retention. The remainder of the thesis reports the studies that were conducted to answer the research questions identified in this chapter. The following chapter begins by reporting the first experiment which aims to investigate the PE in L2 vocabulary learning by young learners of Arabic and to examine whether learning with and without active production is durable over time in these learners.

Chapter 3. A Pilot Study

The previous chapter has shown how research indicates that the effective use of a learned language depends on accurate word recall and that one of the vocabulary learning methods studied in recent years is the Production Effect (PE) in second language learning and memory. Under this umbrella, the PE has been explored as a strategy to improve explicit memory by distinguishing between words learned silently and spoken aloud (Andrä et al., 2020; MacLeod et al., 2010). Research on PE in children has demonstrated a typical production effect (i.e., better recall for words learned aloud) for older children (Icht & Mama, 2014; Zamuner et al., 2017; Pritchard et al., 2019; López et al. et al., 2021) and a reversed production effect (i.e., better recall for words learned silently) in children aged 2;0 to 6;0 years.

This chapter describes a pilot study designed to investigate the Production Effect (PE) in second language (L2) vocabulary learning and to examine the durability of learning with and without production over time. Given the exploratory nature of this pilot, the study's findings will inform the methodology and experimental design for subsequent, larger-scale research. The joint investigation of age and language background has been studied in young L2 learners as factors potentially affecting language acquisition. However, studies explicitly examining PE in contexts where young learners' native language closely aligns with the target language (i.e., the language they are learning) remain underexplored.

This pilot experiment compares word learning with and without production in child learners of Modern Standard Arabic (MSA). The participants, who already possess some phonological knowledge of MSA, provide an ideal foundation for assessing the impact of PE on vocabulary acquisition. The pilot study aims to identify potential patterns and effects in a smaller, controlled group, laying the groundwork to test the study design for the research for future research.

The pilot study addresses two primary questions:

1. To what extent does producing words aloud during word learning enhance L2 vocabulary acquisition in children compared to solely listening to the words being learned?

2. Do age and language dominance influence vocabulary learning under PE conditions (i.e., listening and repeating)?

3.1. Introduction

Understanding the impact of language dominance (i.e., the degree to which a bilingual or multilingual individual's proficiency, usage, and preference for one language surpasses their other languages) and age on language acquisition is crucial in bilingual and multilingual Research indicates that language dominance significantly influences cognitive processes and learning strategies (Bialystok & Craik, 2010; Kroll & Bialystok, 2013), while age affects the ease and manner in which new languages are learned (Cummins, 1980; Andreou & Karapetsas, 2004). Section 2.1.5 described how Modern Standard Arabic (MSA) serves as a heritage language (HL) for its speakers, representing their identity, particularly in multicultural and multilingual communities. Heritage language use unifies communication methods among speakers (Park, 2013). However, many child MSA learners need help to achieve proficiency or explore language uses beyond the classroom. They often find it challenging to use MSA vocabulary in simple conversations despite having an Arabic dialect as their dominant language.

In this study, children learning MSA are considered heritage language learners. These children, born in, migrated to, or residing in the U.K., have caregivers from Arabic-speaking countries and are exposed to Arabic at home. Typically, they have not reached the literacy level of native Arab speakers. Although some use dialectal Arabic (e.g., Saudi Arabic, Levantine Arabic, Iraqi Arabic) as their dominant spoken language, they often need full literacy, meaning they cannot read or write in their spoken dialect. Furthermore, these children struggle to comprehend or use MSA fluently.

Participants were grouped based on their language dominance (Arabic or English), a critical factor as it may affect the efficacy of different learning methodologies. Research suggests that language dominance influences cognitive processes, learning strategies, and educational outcomes. Individuals more proficient in their dominant language tend to perform better in tasks requiring complex language skills, significantly impacting instructional approaches. This proficiency may not always align with an individual's primary or first language (Bialystok et al., 2012; Kroll & Bialystok, 2013). While the primary language is often the first language learned, the dominant language reflects the individual's current linguistic abilities and usage patterns, which can evolve over time based on various factors such as environment and practice (Palma et al., 2023). Language dominance can shape learners' engagement with and comprehend new material, potentially necessitating different pedagogical techniques to optimize learning in bilingual or multilingual contexts (Cummins, 2017). Thus, understanding and accommodating language dominance is essential for developing effective educational practices catering to diverse learners' linguistic strengths and needs.

Considering the purpose of learning a language is also crucial. Different factors contribute to families' decision to send their children to weekend schools to learn MSA. Some families aim to link their children to their cultural and religious identity by learning Arabic to understand the Qur'an and Islamic literature. Many parents believe it is easier for children to learn Arabic early due to its complexity compared to English. One parent mentioned trying to home-school their children using Arabic expressions and watching Arabic films and cartoons but found this approach ineffective, leading to the decision to enrol their children in a weekend school.

This study uses MSA to investigate the production effect (PE) theory. Although the children in this study may not fully understand MSA, they possess phonemic awareness of Arabic, meaning they can recognize specific sounds in spoken words. Phonemic awareness is crucial when acquiring a target language similar to the background language. This awareness depends on familiarity with known phonemes, which helps learners effectively process and understand the new language. Prior knowledge, such as phonemic awareness, aids in learning and acquiring more vocabulary by leveraging these known phonemes to recognize and produce new words accurately. Language learning studies indicate that phonological awareness is a significant predictor of vocabulary acquisition, with children exhibiting better phonological awareness performing better in language learning tasks (Gillon, 2004; Snowling & Hulme, 2011). Additionally, students' curiosity about learning new information significantly influences information retention. The more prior knowledge they possess on a subject, the more curious they become to learn, leading to increased retention of new information (Stanovich, 2009).

As discussed in Section 2.2.5, age plays a crucial role in language acquisition. The Critical Period Hypothesis (CPH) posits that there is an optimal developmental window from birth to early adolescence during which the brain is especially receptive to language learning. During this period, children can achieve native-like proficiency. However, as neural lateralization solidifies around puberty, language acquisition becomes more challenging. The Age of

Acquisition (AoA) effect further supports this, indicating that early learners achieve higher proficiency and more native-like neural processing.

This progression in language acquisition is evident in the pronounced production effect observed in younger children. Perani et al. (2003) demonstrate that younger children exhibit a more significant production effect due to heightened neural plasticity and increased receptiveness to language input. Developmental shifts in the production effect are expected as children advance through different stages, potentially aligning with critical periods in language acquisition (Singleton, 1995). Furthermore, threshold effects are proposed, where the prominence of the production effect correlates with age thresholds and critical developmental milestones, including the onset of formal education (Stanovich, 2009).

This study investigates several considerations regarding age. First, it explores age-related variations in the production effect to determine how evolving cognitive abilities and memory strategies throughout childhood influence its manifestation across different age groups (Gathercole et al., 2004). Additionally, understanding how the production effect varies with age informs age-appropriate educational interventions, guiding the development of tailored learning strategies that optimize language acquisition and memory retention (Luz et al., 2022; Gogate, 2022). Lastly, this investigation contributes to the broader understanding of cognitive development by examining the relationship between age, language processing, and memory formation, thereby enhancing knowledge of developmental cognitive neuroscience (Cummins, 2017).

3.2. Methods and Design

This study explored how children learning Arabic as an additional language remember and recognize words when encountering unfamiliar MSA terms in a listen-and-repeat setting. A within-subject mixed-list design was employed. Consequently, the children were presented with the words in a randomized order. To ensure balanced exposure, the children heard the "listen" words three times and the "repeat" words twice, with the latter heard a third time during self-production. This approach ensured an equal number of exposures for both conditions.

The study included three phases: training, learning, and testing. Children were first trained on six known MSA words to familiarize them with the study procedure. Two steps were followed in the training: (a) familiarize the children with the conditions and (b) familiarize the children with the testing. Children learned new words under two conditions: Listen (children see the image and hear the associated word three times) and Repeat (children see the image, hear the associated word twice, and produce the word once). During the learning phase, children were introduced to 20 unknown MSA words. After the learning session, there was an immediate testing session followed by a 24-hour delayed recognition test and a 9-week delayed recognition test to evaluate word learning durability and any differences between conditions in the different testing intervals. After the first learning session, there was a short break, followed by a recognition test. The learning and testing sessions were conducted online using Gorilla Experiment Builder (https://app.gorilla.sc/admin/home), a web-based experiment builder and deployment platform. Children participated in the presence of a caregiver to assist whenever needed.

3.3. Participants

Participants were 73 students (34 males, 39 females) aged 4.1 to 10.8 years old (M = 6.0, SD = 1.5). According to parents' reports, all children were neurotypical with normal or corrected-to-normal vision and normal hearing and no background in any language impairment or delay (e.g., dyslexia). Children who reported having hearing loss or severe language disorders were excluded from the study. Seven participants were excluded due to their own request to withdraw or due to their having refused to complete the initial learning task. Refusal or delaying the 24-hour recognition task was another criterion for exclusion. A total of 73 participants were included in the analysis for the immediate and the 24-hour delayed test. Only 53 children (26 males, 27 females, aged 4.1-10.8 years old, M=6.1, SD=1.5, range=4.1-10.8) agreed to complete the 9-week delayed test, and the data was analysed according to the test (i.e., immediate post-test, one-week delayed test, and 9-week delayed test).

Children were recruited from the Arab society living in the UK. They were required to have prior knowledge of Arabic sounds and letters, which means they should be able to recognize and differentiate between the basic Arabic alphabet and their corresponding sounds. This foundational knowledge is necessary because it forms the basis for further language learning. It ensures that children can progress in the learning task without being hindered by a lack of familiarity with the alphabet. This level of knowledge does not necessarily include the ability to read or understand words but instead focuses on the recognition and auditory differentiation of individual letters and their sounds. All children had different Arabic backgrounds (i.e., they belonged to one of the Arab countries) and had been exposed to various accents or dialects of

MSA (i.e., Saudi, Omani, Kuwaiti, and Yemeni); however, only MSA words that are consistent across these dialects were chosen for the experimental word list.

Additionally, children were required to attend a British school and be exposed to or learn Modern Standard Arabic at an online- or weekend- Arabic school. This requirement ensures that the children are immersed in both English and Arabic speaking environments during their formal education. This dual exposure is crucial for mitigating the effect of language dominance. Children's language dominance was determined by the primary language used by family members (i.e. spoken language at home).

In order to be assigned to the appropriate language level, all children took a standardized test created and used by their Arabic school to determine their MSA language level as required by the school's registry. Each class contained students of different ages but of the same learning level. In these Arabic schools, the children were usually divided into six learning levels, from year 1 to year 6. Their level (i.e., year group) in the English school was provided by the parents on the participant information sheet, as the focus of the current study was on acquiring Modern Standard Arabic and how children's language ability could be affected by the method of production in acquiring a language.

Participants completed a language background questionnaire (LBQ) prior to taking part in the study (Appendix 4). LBQ was administered to the parents of the children to collect information regarding the children's language abilities. Additionally, the children's proficiency levels were determined through a standardized test that evaluated various language skills, such as reading and writing. While these standardized tests provided an overall measure of ability (Tremblay, 2011), the detailed LBQ enabled the researcher to compare participants' performance in the MSA. This approach facilitated the recruitment of appropriate participants for the study, ensuring that the children were not proficient in MSA and did not have any language impairments. All participated children were attending level 1 Arabic classes and were divided into two groups according to their dominant language (Arabic or English).

Finally, all participants and caregivers signed a consent form, which informed them that all trials would be audio recorded. This recording was necessary to evaluate their online production and ensure they followed the instructions correctly. By signing the consent form, participants acknowledged their understanding and agreement to these conditions, and they were informed that they could withdraw from the study at any time.

3.4. Materials

The stimuli were 20 unfamiliar animate (e.g., spider) and inanimate (e.g., rocket) standard Arabic nouns. Proper names, abstract nouns such as colours, and English loanwords or borrowed words with similar English pronunciation (e.g., 'octopus'/ʔo:tʕbu:tʕ/, 'ice cream'/ʔiskiriːm/) were excluded from the study. These exclusions were made because participants were likely already familiar with these words, which could influence recall over non-cognates (Jessen et al., 2000). Non-cognates are words that do not share a similar form or origin between languages, unlike cognates, which can often have recognizable similarities due to their common etymological roots. This distinction ensures that any recall or recognition is based on the novelty of the words rather than prior exposure or phonetic similarity.

Words were selected from 23 Arabic storybooks written and published by various Arab children's publishers (e.g., Kadi and Ramadi, Boustany's Publishing House). Some of the extracted words were taken from books that were translated from famous English stories into Arabic (e.g., The Gruffalo or Al-Gharfoul); they were then compared to the curriculum being studied by the children at the Arabic school in order to exclude all words mentioned in the textbooks and increase the likelihood that target items would be unfamiliar. Because the children were studying Arabic at different levels, a familiarity questionnaire was developed to include all selected words that were not in the textbooks, and the parents and teachers completed it. The form asked the caregivers to mark all words they thought the children knew from the class extracurricular activities or that they might have known from books they read at home. All nouns likely to already be familiar to the children were excluded from the study list.

Phonotactic probability was another controlled variable. This refers to the frequency of occurrence of a set of segments or sound sequences in a language; the higher the frequency, the greater the likelihood of the children encountering it on more than one occasion. Phonotactic probability was considered, as studies of speech processing among English-speaking preschool children have revealed that it influences spoken word production. For instance, children aged 3–6 years show faster recognition and better production for frequent sound sequences than less frequent sounds (Storkel, 2001; Auer & Luce, 2005). In the present study, words were carefully checked to exclude low-frequency sound patterns, such as specific combinations of vowels and consonants, or any word with a low-frequency syllable structure in Modern Standard Arabic (MSA). This was done by giving the selected list of words to a teacher of Arabic studies, who reviewed and confirmed the appropriateness of the chosen words. Also, to ensure that the words would be easy to pronounce by children in the age groups in this study, eight English-Arabic bilingual children between the ages of 4 and 11 who were not participating in the study or attending the Arabic school completed a trial run of the words. Word length was also considered, as long words are difficult for children to pick up and produce (Jackson et al., 2019). Hence, the stimuli included words 3–8 phonemes in length but varied in the number of syllables (e.g., /ʕænkæbu:t/"spider", /xuld/"mole"). After these checks had been completed, a final list of 20 items that best matched all of the criteria was chosen and randomly assigned to two sets of 10 words each according to the conditions: listen and repeat, in addition to six Arabic familiar and frequent words (e.g., rabbit /ʔarnab/) that were presented for training purposes (Table 3.1).

Table 3.1: List of words used in the experiment according to learning condition and experimental phase

Training	Learning	
apple /tuffa:ħ/	whale /ħu:t/	squirrel /sind͡ʒæb/
pen /qalam/	mole /xuld/	statue /timθa:l/
car /sayarra/	window /næfiðæ/	cottage /ko:x/
	air balloon /min͡tsæd/	grasshopper /d͡ʒundub/
	needle /ʔebræ/	owl /bu:mæ/
dog /kalab/	bat /wætˤwa:tˤ/	crocodile /timsæħ/
chair /kur.si/	deer /zˤæbi/	volcano /bu:rkæn/
rabbit /ʔar.nab/	tent /xæjmæ/	goose /ʔewazæ/
	rocket /sˤæro:x/	spider /ʕænkæbu:t/
	bag /ħaqebæ/	arrow /sæhm/

For the chosen stimuli, illustrations were produced to give a clear indication of the meaning associated with each word. The visual stimuli were digitally illustrated using Adobe Illustrator by a volunteer Saudi artist who was given the final list of stimuli words. All pictures were designed to be as simple and easy to recognise by children as possible based on children's storybooks (Figure 3.1). The colours of the pictures were selected to be bright and clear, with distinct black boundaries. Additionally, all pictures were evaluated by three elementary school teachers, a specialist in children's literature, and a storyteller to ensure their suitability for the age group of children in the study. The images were assessed for clarity and lack of ambiguity, ensuring that a child would be able to accurately name the target word based on each image. Finally, all stimuli were recorded by a native Arabic speaker to ensure accurate pronunciation and natural intonation.

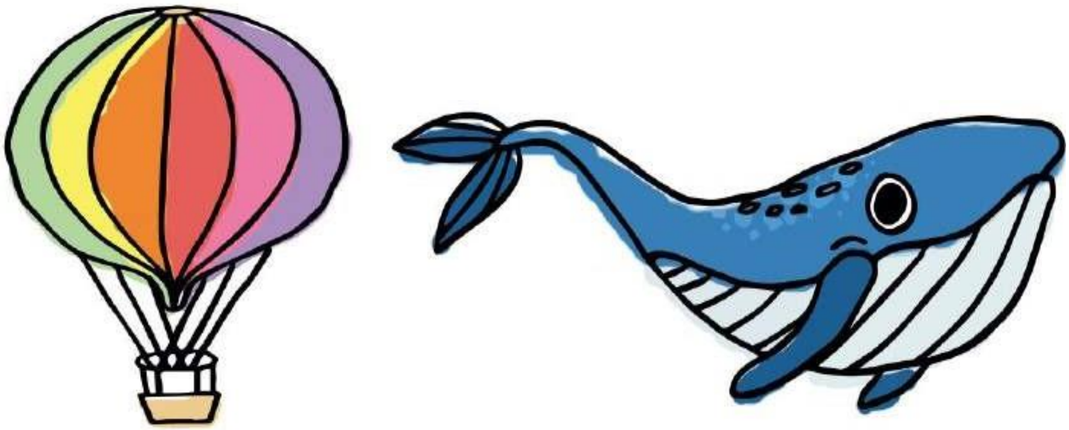


Figure 3.1: Example of visual stimuli:

"air balloon" / mintˤæd/ and "whale" / ħu:t/

3.5. Procedure: Training, Learning, and Testing

Testing was conducted online via the Gorilla platform where children were tested in three sessions that lasted 20–30 minutes in total. They were tested individually at home under the supervision of their caregiver. The experiment involved three online sessions that occurred over three days. The current experiment included 3 phases:

3.5.1. The training phase

Six known MSA words were randomly assigned to the conditions for training purposes (Table 3.1). Each word was accompanied by an instructional emoji representing its condition. The training screen began with the phrase "Let's Practice!". The children viewed instructional images that explained the experimental conditions (Figure 3.2). The listen condition was introduced with an image illustrating a finger over the mouth to indicate silence, and the repeat condition was introduced with a smiley face and two fingers, indicating that they would hear the word twice.

Each item was presented in auditory and visual formats for 10 seconds, with a 10-second gap between each item. The caregiver's role was to refer to the instructional image to remind the participants of the condition. After a 5-minute break, children were trained on the recognition test.

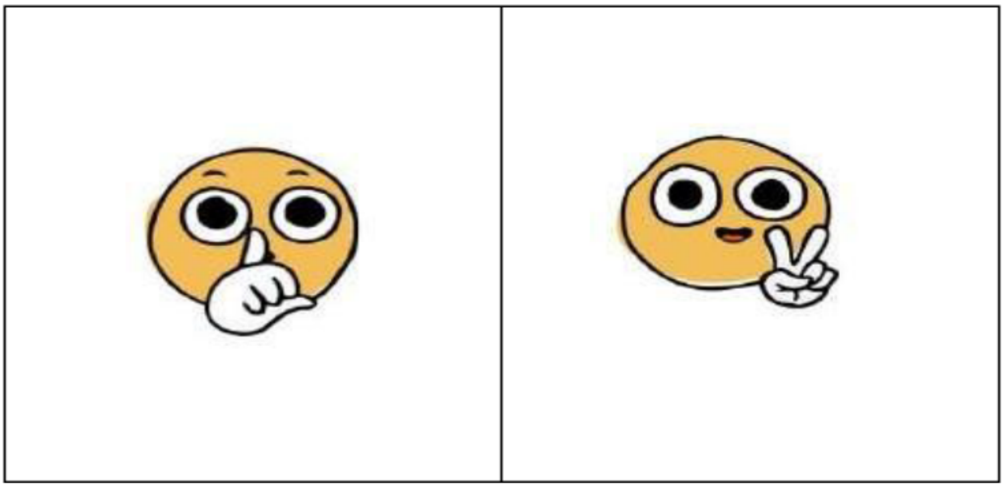


Figure 3.2: Visual instructional stimuli for the conditions

listen (left), repeat (right)

3.5.2. The learning phase

Learning trials had the same structure as training. They included 20 MSA words randomly divided into two lists of ten words according to the study conditions: listen or repeat. All children learned the same 20 words, but the conditions were counterbalanced.

3.5.2.1. The testing phase

Participants were tested using a recognition task administered immediately, after 24 hours, and after nine weeks. Due to time constraints, the immediate and 24-hour delayed tests consisted of ten trials (five per condition). This limitation was implemented to minimize testing time and reduce participant fatigue. However, the 9-week delayed test was adjusted to include all 20 learned items.

During the recognition test, children were instructed to click on their preferred image as quickly as possible. The initial screen displayed written instructions: 'Listen to the question, then select the answer.' Once participants understood the instructions, they clicked 'I'm ready!' to start the test. A blank screen appeared for 10 seconds, followed by four images and an auditory question (e.g., /ʔajna al-ʔebræ?/ "where is the needle?") (Figure 3.3).

The trials displayed four randomized images, including different combinations of target words, distractors, and competitors. Specifically, each trial consisted of four pictures representing: a word from each condition (i.e., listen or repeat), a competitor from the other condition (i.e., if the target was from the listen condition, the competitor was from the repeat condition, and vice versa), and two distractors chosen from a list of known Arabic words agreed upon by the parents and teachers of the participating children. The distractor items were not included in the learning list. Each testing session took about 15-20 minutes per participant, excluding training.

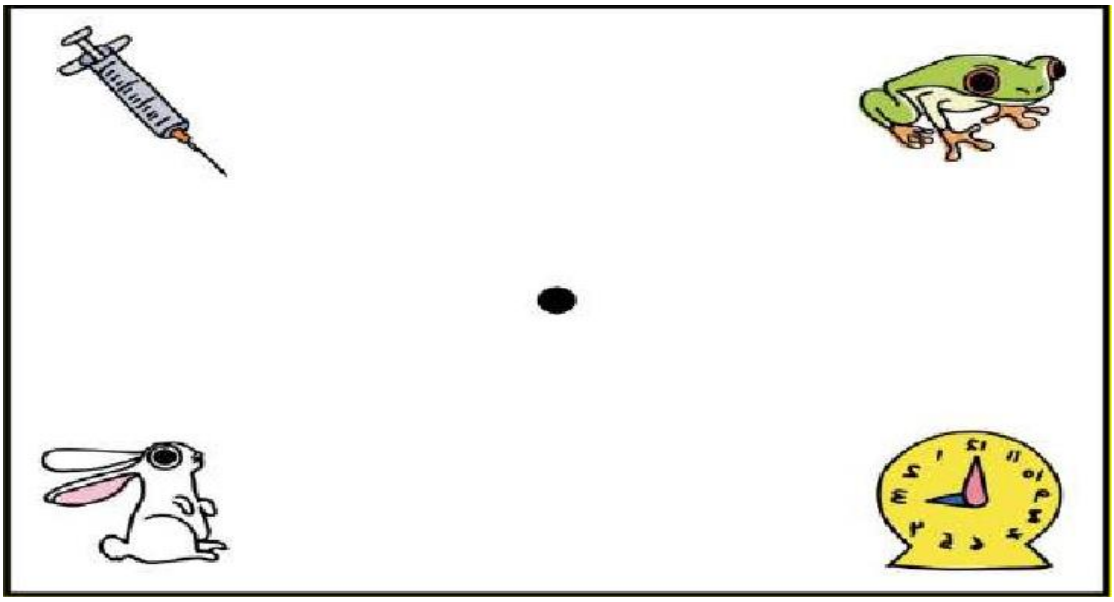


Figure 3.3: Example of a test trial demonstration

Target: /ʔebræ/ "needle", a word from the listen condition. Competitors: /dˤ u:fdaʕ/ "frog", a word from the repeat condition that was seen during the experimental session. Distractors: /ʔarnab/ "rabbit" and /sa:ʕæ/ "watch"

3.6. Analysis

Accuracy of answers and reaction time were the primary dependent variables in this study. The main effects of interest were whether items were repeated or observed silently (condition: listen and repeat) and the timing of the post-test (post-test: immediate, 24-hour delayed, and 9-week delayed). Age and language dominance were also included as additional factors.

3.6.1. Coding

Responses to the recognition test were coded in Gorilla. Testing trials were coded for accuracy, with "1" representing correct responses and "0" representing incorrect responses. For reaction time, only reaction times to correct answers were included in the analysis. As the age of participants varies in the current experiment, it was calculated in months to provide a more precise measure of age differences, which can be critical in understanding developmental changes over shorter time spans (Schjølberg et al., 2011). Language dominance was coded as either Arabic or English.

3.6.2. Statistical Analysis

Prior to analysis, data was extracted from Gorilla in spreadsheet form and cleaned in R version 4.0.5 (R Core Team, 2021). The data was organized to include age, language dominance (i.e., Arabic or English), reaction time to correct answers, and accuracy (correct/incorrect) in all testing phases (immediate post-test, 24-hour delayed test, and 9-week delayed test).

Participants who scored less than 50% in accuracy responses across the three tests were omitted, which removed 7.37% of the data. Items associated with words that emerged during the 'listen' condition or words not produced during the 'repeat' condition were excluded from the analysis, removing an additional 2.5% of data across the three tests, leaving 2034 data points for analysis. The final data sheet was saved in CSV format, and all statistical tests were performed using R (version 4.1.2, R Core Team, 2021) and RStudio.

Reaction times (RTs) were cleaned and trimmed in R to exclude all errors and remove outliers based on response times below 200 ms and responses ±2.5 SD from each participant's mean response time. This process removed 16.9% of the data, leaving 1616 data points for the RT analysis across the three tests. The RTs were then log transformed to reduce skewness[2]. Finally, although participants were informed to answer as quickly as possible, it is important to note that reaction time results are interpreted with caution as children take longer time to respond to questions due to their developing cognitive and motor skills (Wolfe & Bell, 2007).

A generalized linear mixed-effects model (GLMM) with a binomial distribution was employed to analyze the accuracy of data. The "glmer" function was used to fit the model, incorporating both random and fixed effects, with the "bobyqa" optimizer used for maximum likelihood estimation. For analyzing reaction times to correct answers, we utilized linear mixed-effects models, which were constructed and analyzed using the "lme4" (version 1.1-26; Bates et al., 2015) and “lmer Test” (version 3.1-3; Kuznetsova et al., 2017) packages. The maximal random structure model included the interaction of all conditions with tests, and a fully random model was applied whenever feasible. If convergence issues occurred, the random slopes were simplified. To identify significant differences among the groups, a multiple-comparisons test with Bonferroni correction was performed using the 'emmeans' package (Lenth, 2022).

3.6.3. Results

Overall recognition results are shown in Table 3.2, which presents the mean and standard deviation (in parentheses) of accuracy (%) and reaction times (RTs) in milliseconds for the two conditions (Listen, Repeat) across three tests (immediate, 24-hour, nine-week). The recognition test showed that the "Listen" condition had stable accuracy across all three tests, with a slight increase in accuracy observed in the 9-week test. In contrast, the "Repeat" condition demonstrated a noticeable improvement in accuracy in the 24-hour test but decreased back to the original level in the 9-week test, indicating a possible decline in retention over time.

[2] Reaction time (RT) data can be transformed either by using an inverse Gaussian distribution or by applying a logarithmic transformation. RT data are nearly always right skewed, which might dominate the outcome. Generalized linear models, unlike linear models, do not assume a normally distributed dependent variable. While an inverse Gaussian distribution is one approach to normalize raw RT data, log transformation is often preferred. Log transformation is recommended because it helps stabilize variance and make the data more closely conform to a normal distribution, which is essential for the assumptions underlying many statistical tests (Czamolewski, 1996; Keene, 1995). In the current study, log transformation is applied to normalize raw RT data using the glmer function from the lme4 package (Bates et al., 2015).

Regarding reaction times, the "Listen" condition exhibited a decrease from the immediate to the 24-hour test, followed by an increase in the 9-week test. The "Repeat" condition showed consistently lower reaction times compared to the "Listen" condition across all tests, with the fastest response times recorded in the 24-hour test.

Table 3.2: Mean and SD (in parentheses) of correct answers (%) and RTs (in milliseconds) based on condition and test intervals

Recognition		**immediate (n= 73)**	**24-hour (n= 53)**	**nine-week (n=53)**
Accuracy Mean (SD)	Listen	82.9 (37.6)	82.2 (38.2)	84.7 (35.9)
	Repeat	81.5 (38.9)	86.9 (33.8)	81.8 (38.5)
RT Mean (SD)	Listen	4930.4 (2543.6)	4725.9 (2348.5)	4835.1 (4168.7)
	Repeat	4708.9 (2362.3)	4562.1 (2206.8)	4928.3 (4279.5)

3.6.3.1. Accuracy of Answers

To analyze the accuracy across the three tests, a generalized linear mixed effect model (glmer) with binomial distribution analysis was performed on the data for the 53 participants who completed all three tests. The base model included fixed effects of condition, treatment coded with listen as the baseline, and test time, treatment coded with immediate test as a baseline, and their interaction. This model included random intercepts for subject and item, and an additional model was built to include by-subject random slopes for condition and test time, aiming for a maximal random effects structure. Model comparison was used to compare the models, and this showed that adding by-subject random slopes for test time significantly improved the fit ($\chi2$ (5) = 13, p =.02). Overall, there was no effect of condition (z = 0.37, p = .71), but there was a significant effect of test whereby the 9-week delayed test shows higher scores than the immediate test (z = 5.23, p < .001). There were no interactions between condition and test. The results of the best fit model are shown in Table 3.3.

Table 3.3: Mixed effects model output for the accuracy of answers interacting condition and test

Formula: glmer (correct ~ condition * test + (1 + condition*test | ID) + (1 | answer)

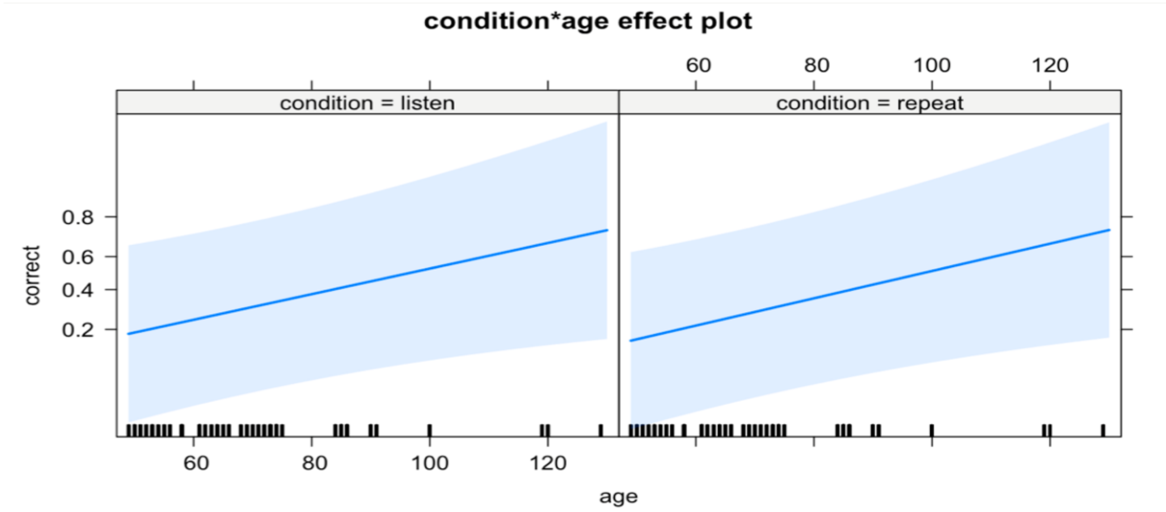
control = glmerControl(optimizer= "bobyqa"), family = binomial)

Fixed effects	**Estimate**	**SE**	**z-value**	**p-value**
(Intercept)	-3.43	2.10	-1.63	0.10
Condition: repeat	0.13	0.36	0.37	0.71
Test: 24-hour delayed	0.04	0.38	0.91	0.91
Test: 9-week delayed	2.32	0.44	5.23	<.001***
repeat*24-hour delayed test	0.25	0.52	0.49	0.63
repeat* 9-week delayed test	-0.64	0.46	-1.37	0.17
Random effects variance (subject = 53, item = 40)				

* $p < 0.05$; ** $p < 0.01$; *** $p < .001$

3.6.3.1.1. The Effect of Age on Accuracy of Answers

To evaluate the impact of age on model fit, it was incorporated into the base model as a fixed effect, as well as through its interactions with condition, test, and the combined interaction of



condition and test. The inclusion of age significantly improved the model fit as a fixed effect (Table 3.4), and was further improved as an interaction with condition (χ^2 (2) = 40.6, $p < 0.001$) (Figure 3.4), and as an interaction with test (χ^2 (1) = 6.42, $p < 0.001$) (Figure 3.5). However, the interaction of age with both condition and test did not yield further improvement (χ^2 (6) = 10.1, $p = .12$). These results suggest that the best model fit is achieved when age is included as a fixed effect, particularly due to the interactions illustrated in Figures 3.4 and 3.5. In Figure 3.4, the effects of age on answer accuracy are consistent across both conditions, supporting the rationale for treating age as a fixed effect.

Figure 3.4: The interaction of age and condition on accuracy of answers. y axis is log-odds of answer accuracy

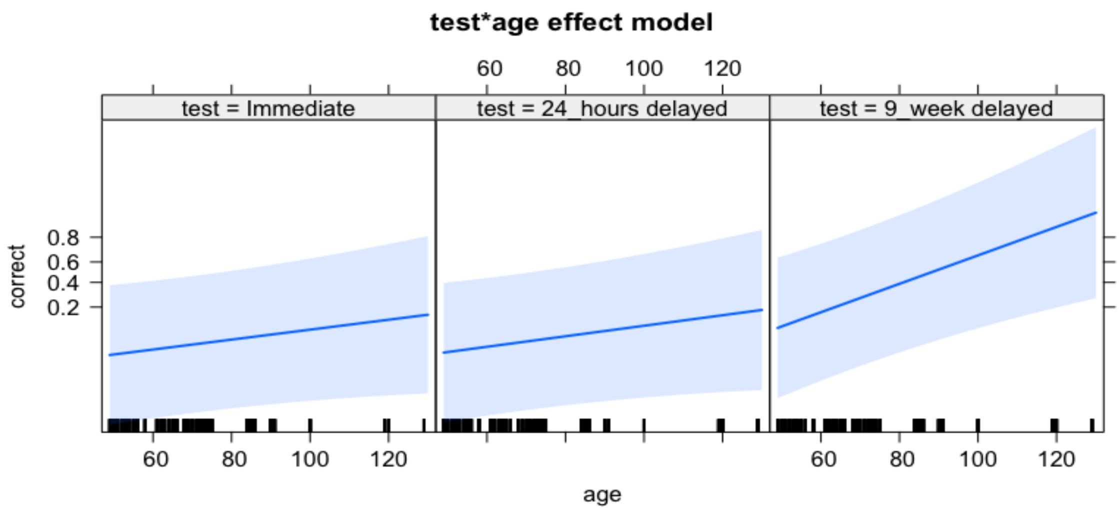


Figure 3.5: The interaction of age and test time on accuracy of answers. y axis is log-odds of answer accuracy

The influence of age on answer accuracy is illustrated in Figure 3.6. It is indicated that increased age (left) is associated with higher log-odds of answer accuracy, suggesting a positive correlation between age and accuracy. The figure also presents the interaction effects of condition and test time on accuracy (right), highlighting the differential impacts of these variables over time. Overall, the accuracy of answers is significantly enhanced by age, both as an independent variable and through its interactions with condition and test time.

Table 3.4: Mixed effects model output for the interaction between condition and test in accuracy including age

Formula: glmer(correct ~ condition * test + age + (1 + condition*test | ID) + (1 | answer) ,

control = glmerControl(optimizer= “bobyqa”), family = binomial)

Fixed effects	Estimate	SE	z-value	p-value
(Intercept)	-4.67	1.62	-2.87	<.001***
Condition: repeat	-0.124	0.36	0.34	0.73
age	0.02	0.01	2.03	0.04 *
Test: 24-hour delayed	0.02	0.38	0.07	0.94
Test: 9-week delayed	2.13	0.43	4.92	<.001***
repeat* 24-hour delayed test	0.27	0.53	0.52	0.60
repeat* 9-week delayed test	-0.62	0.46	-1.33	0.18
Random effects variance (subject = 53, item = 40)				

* $p < 0.05$; ** $p < 0.01$; *** $p < .001$

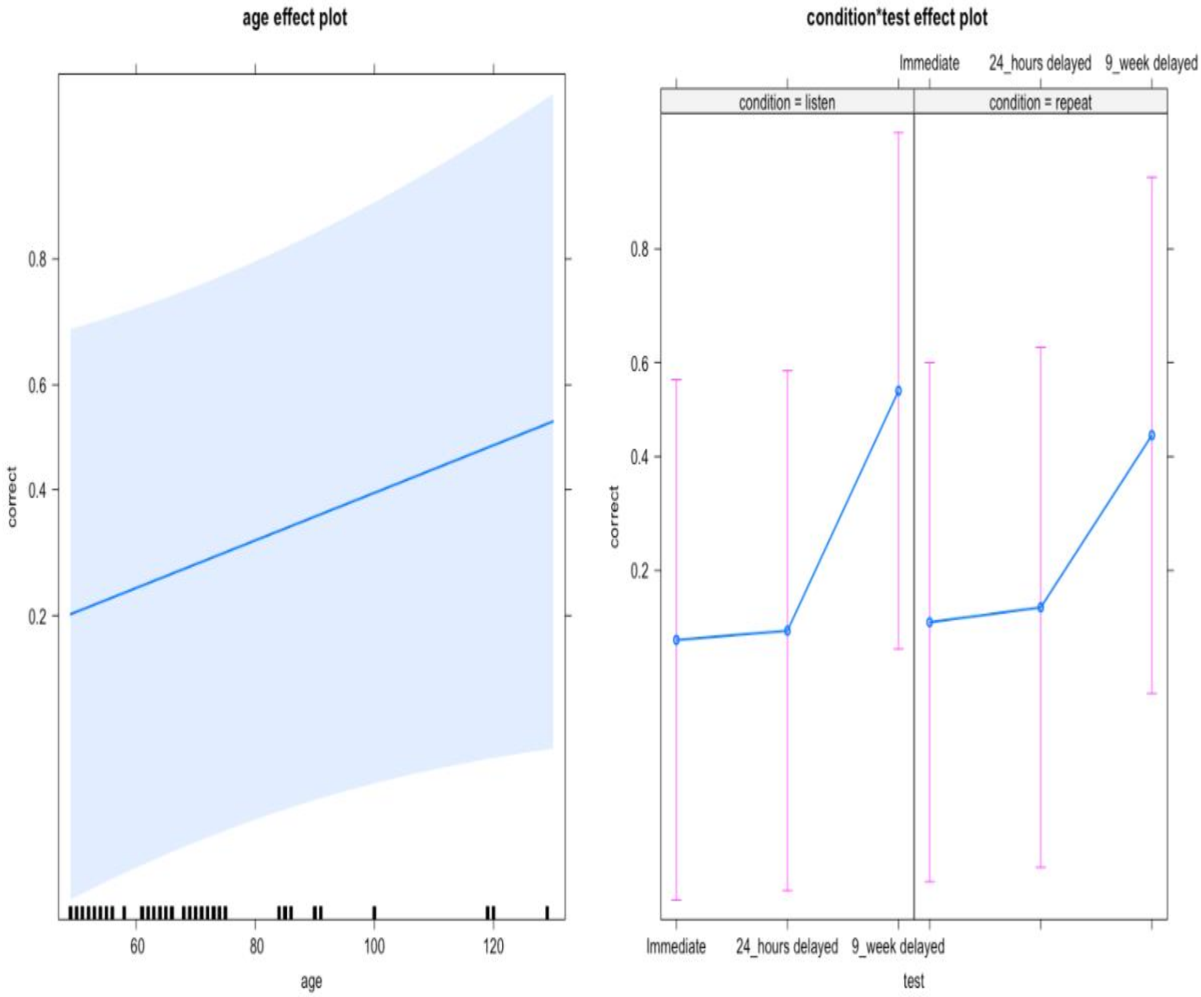


Figure 3.6: The effects of age (left) and the interaction of condition and test time on accuracy of answers (right). Y axis is log-odds of answer accuracy

3.6.3.1.2. The Effect of Language Dominance on Accuracy of Answers

Adding language dominance to the baseline model as a fixed effect caused an unavoidable convergence issues; a simplified model is used, but it did not improve the model fit (χ^2 (16) = 0, p = 1.0). Including language dominance as an interaction with test ($\chi^2(17) = 0.0$, p = 1.0), as an interaction with condition ($\chi^2(19) = 0.0$, p = 1.0), and as an interaction with both test and condition ($\chi^2(11) = 0$, p = 1.0) did not improve the model fit. The lack of improvement in model fit when including language dominance, either as a fixed effect or in interaction terms, suggests that language dominance is not a significant factor in predicting the variance in correct responses. That is, whether participants' dominant language is Arabic or English does not significantly affect their ability to correctly identify MSA words.

Finally, a posthoc test using the package "emmeans" (Lenth, 2022) was applied to the base model (i.e., condition*test) to compare the conditions (i.e., listen and repeat) at different test levels (i.e., immediate, 24-hour delayed, 9-week delayed). Table 3.5 presents the results of a posthoc test using the "emmeans" package (Lenth, 2022), applied to the base model to compare the conditions (listen vs. repeat) across different test levels (immediate, 24-hour delayed, 9-week delayed). The pairwise comparisons reveal no significant differences between the conditions at any of the test points. This indicates that the accuracy does not significantly differ between listening and repeating tasks across these time intervals, suggesting that the mode of condition (listening or repeating) does not substantially impact accuracy over immediate, 24-hour, or 9-week delayed tests.

Table 3.5: Post-hoc tests with Bonferroni correction for accuracy based on the interactions between condition and test model

Contrasts	**Estimate**	**SE**	**z-ratio**	**p-value**
(Immediate listen) - (Immediate repeat)	-0.13	0.36	-0.35	1.0000
(24-hour delayed listen) - (24-hour delayed repeat)	-0.39	0.37	-1.05	1.0000
(9-week delayed listen) - (9-week delayed repeat)	0.49	0.28	1.77	1.0000

* $p < 0.05$; ** $p < 0.01$; *** $p < .001$

3.6.3.2. Reaction Time to Correct Answers

A linear mixed-effects model (lmer) was fitted to evaluate the log-transformed reaction time (RT) to correct answers. The model included condition and test time, and their interaction as fixed effects, random intercepts for subject and item, and by-subject random slopes for condition and test time. Table 3.6 shows no significant effects of condition ($\beta = 0.06$, $t = 1.30$, $p = .19$) or the 24-hour delayed test ($\beta = 0.04$, $t = 0.87$, $p = .38$). The 9-week delayed test also showed no significant effect ($\beta = 0.05$, $t = 0.51$, $p = .61$). Furthermore, the interactions between condition and the 24-hour delayed test ($\beta = -0.09$, $t = -1.29$, $p = .19$) and between condition and the 9-week delayed test ($\beta = -0.02$, $t = -1.36$, $p = .70$) were not significant. These results suggest that neither the conditions nor the test times significantly influence the response time to correct answers. Additionally, the lack of significant interaction effects indicates that the condition does not affect reaction times at different test times.

Table 3.6: Mixed effects model output for reaction time (log-transformed)

Formula: lmer (RT_log ~ condition * test + (1 + condition * test | ID) + (1 + test | answer)

Fixed effects	Estimate	SE	t-value	p-value
(Intercept)	8.15	0.10	74.90	<.001***
Condition (repeat)	0.06	0.04	1.30	0.19
Test (24-hour delayed test)	0.04	0.04	0.87	0.38
Test (9-week delayed test)	0.05	0.10	0.51	0.61
repeat* 24-hour delayed test	-0.09	0.07	-1.29	0.19
repeat* 9-week delayed test	-0.02	0.06	-1.36	0.70
Random effects variance (subject = 53, item = 40)				

p-values are estimated using the *lmerTest* package in R (Kuznetsova et al. 2017). * $p < 0.05$; ** $p < 0.01$; *** $p < .001$

3.6.3.2.1. The Effect of Age on Response Time to Correct Answers

To assess the influence of age on model performance, it was added to the base model as a fixed effect, and through its interactions with condition, test, and the combined interaction of both condition and test. The results from this model showed that adding age as a fixed effect (χ^2 (1) = 87.8, p <.001), and as an interaction with test only (χ^2 (1)=5.19, p =0.02) (Figure 3.7) improved the model fit. However, adding its interaction with condition (χ^2 (3) = 4.05, p= 0.25) (Figure 3.8), and as an interaction with test and condition (χ^2 (1) = 2.82, p = 0.09) was not significant. The best model fit was achieved by adding age as a fixed effect (Table 3.7).

Table 3.7: Mixed effects model output for reaction time (log-transformed) including age

Formula: lmer (RT_log ~ condition * test *age + (1 + condition * test | ID) + (1 + test | answer)

Fixed effects	Estimate	SE	t-value	p-value
(Intercept)	8.76	0.18	47.6	<.001 ***
Condition: repeat	-0.17	0.14	-1.25	0.26
Test: Immediate	-0.09	0.24	-0.30	0.69
Test: 24-hour delayed	-0.06	0.22	-0.30	0.75
age	-0.00	0.00	-3.36	<.001 ***
repeat* Immediate test	0.40	0.25	1.60	0.11
repeat* 24-hour delayed test	0.13	0.26	0.50	0.61
repeat* Immediate test*age	-0.00	0.00	-1.56	0.12
repeat* 24-hour delayed test*age	-0.00	0.00	-1.78	0.43
Random effects variance (subject = 53, item = 40)				

p-values are estimated using the *lmerTest* package in R (Kuznetsova et al. 2017). * $p < 0.05$; ** $p < 0.01$; *** $p < .001$

Table 3.7 presents the mixed-effects model output for log-transformed reaction time. The model results indicate that age significantly impacts reaction time, as evidenced by the highly significant p-value (p <.001) for age. The negative coefficient (β = -0.00) suggests that as children age, their reaction times improve, indicating that older children tend to respond more quickly. The non-significant effects of condition, test, and their interactions, when age is included in the model, suggest that these factors do not independently influence reaction time to the same extent as age. This underscores the pivotal role of age in determining reaction time performance. Overall, Figures 3.6 and 3.7 underscore that as children grow older, their cognitive and motor functions likely mature, leading to faster response times. This illustrates the significance of age as a crucial variable in studies of reaction times in children.

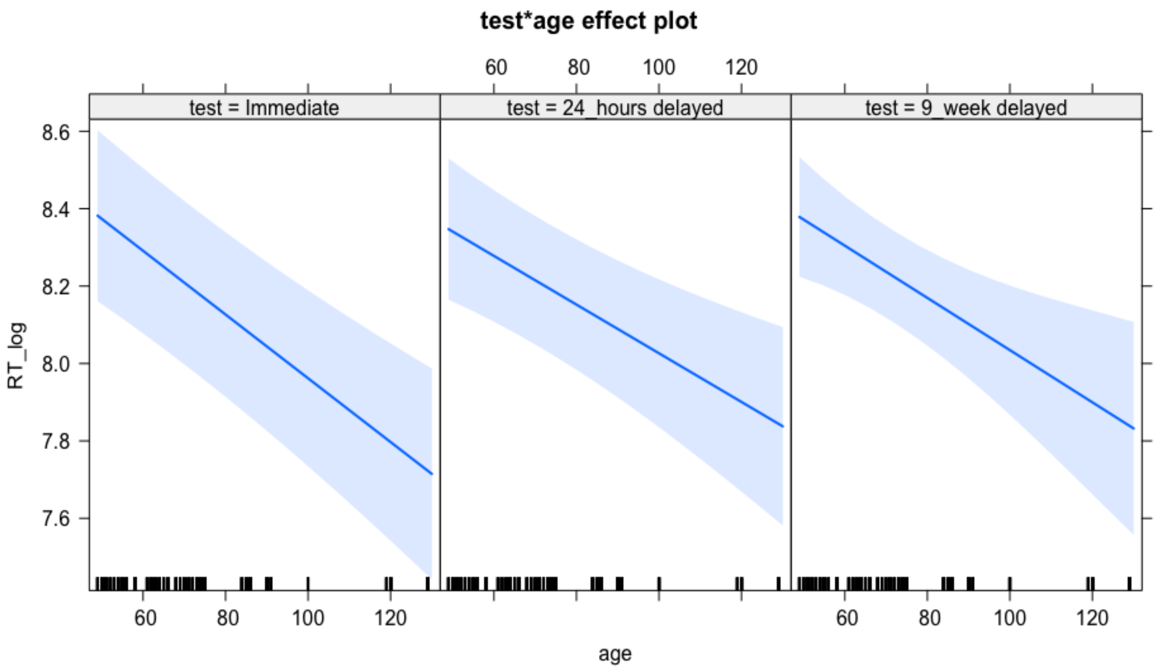


Figure 3.7: The interaction of age and test on response time to correct answers. y axis is log-odds of answer accuracy

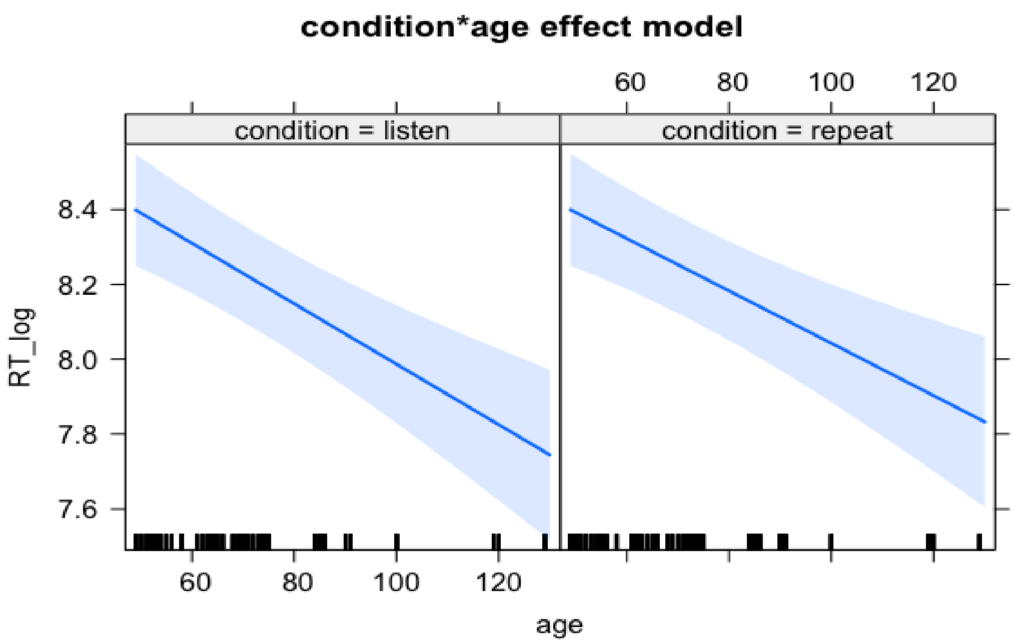


Figure 3.8: The interaction of age and condition on response time to correct answers. y axis is log-odds of answer accuracy

3.6.3.2.2. The Effect of Language Dominance on Response Time to Correct Answers

To evaluate the impact of language dominance on model performance, it was included in the base model as a fixed effect, but it did not improve the model fit (χ^2 (1) = 0.62, p = 0.42). Also, no significant improvement in the model fit was observed by adding language dominance interaction with test ((χ^2 (6) = 4.23, p = 0.64)), or its interaction with condition ((χ^2 (2) = 2.50, p = 0.28)). Adding its interaction with test and condition to the maximal model caused unavoidable convergence issues; however, a significant improvement was found in the simplified model (χ^2 (12) = 68.8, p < 0.001) (Table 3.8).

Table 3.8: Mixed effects model output for reaction time (log-transformed) including language dominance

Formula: lmer (RT_log ~ condition * test * dominant langauage + (1 | ID) + (1 + test | answer)

Fixed effects	Estimate	SE	t-value	p-value
(Intercept)	8.17	6.17	132.3	<.001 ***
Condition: repeat	4.95	3.54	1.39	0.16
Test: Immediate	2.24	9.38	0.02	0.98
Test: 24-hour delayed	1.04	7.48	0.14	0.89
Dominant langauge:Arabic	-1.98	1.04	-1.91	0.06
Dominant langauge:English	1.93	1.02	1.89	0.06
repeat* Immediate test	-2.63	6.34	-0.41	0.67
repeat* 24-hour delayed test	-5.04	6.27	-0.80	0.42
Immediate test*dominant Langauage: Arabic	0.22	0.11	2.01	0.04 *
24-hour delayed test*dominant Langauage: Arabic	0.10	0.09	1.10	0.26
Immediate test*dominant Langauage: English	-0.16	0.12	-1.40	0.16
24-hour delayed test*dominant Langauage: English	-0.14	0.10	-1.38	0.17
repeat* Immediate test*dominant Langauage: Arabic	-0.11	1.15	0.77	0.43
repeat* 24-hour delayed test*dominant Langauage: Arabic	0.05	0.15	0.33	0.73
repeat* Immediate test*dominant Langauage: English	1.11	1.50	0.74	0.45
repeat* 24-hour delayed test*dominant Langauage: English	-3.31	1.49	-0.22	0.82
Random effects variance (subject = 53, item = 40)				

p-values are estimated using the *lmerTest* package in R (Kuznetsova et al. 2017). * $p < 0.05$; ** $p < 0.01$; *** $p < .001$

Table 3.8 summarizes the mixed-effects model for log-transformed reaction time, examining the influence of condition (listen vs. repeat), test type (immediate vs. 24-hour delayed), and dominant language (Arabic vs. English). The main effects of condition ($p = 0.16$) and test type ($p = 0.89, 0.98$) were not statistically significant, suggesting that repetition and the testing time do not independently affect reaction time in this study. Dominant language exhibited a marginally non-significant trend ($p = 0.06$), with Arabic dominance potentially associated with slower reaction times ($\beta = -1.98$) and English dominance potentially associated with faster reaction times ($\beta = 1.93$). Notably, the interaction between immediate testing and dominant language (Arabic) was significant ($\beta = 0.22$, $p = 0.04$), indicating that Arabic-dominant participants displayed a differential response pattern depending on the test timing. Specifically, the interaction effect ($p = 0.04$) suggests a notable improvement in reaction times for Arabic speakers in the immediate test condition compared to the delayed test condition, and/or compared to speakers of other dominant languages under the same immediate test condition. The three-way interaction among repeat condition, immediate test, and dominant language was insignificant, implying that the combined effects of repetition, test timing, and dominant language did not significantly influence reaction times.

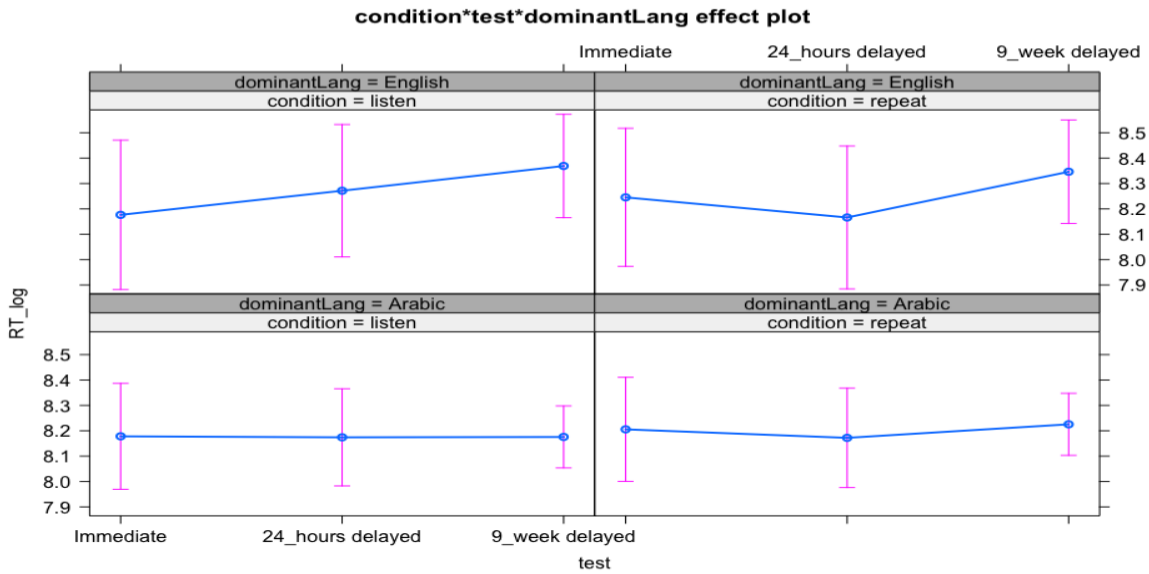


Figure 3.9: The interaction of language dominance, test, and condition on response time to correct answers. y axis is log-odds of answer accuracy

Overall, while the main effects of condition and test type were insignificant, the interaction between immediate testing and the dominant language (Arabic) revealed specific conditions where reaction times varied significantly (Figure 3.9). This finding underscores the importance of considering how dominant language interacts with testing conditions to influence reaction time.

As a final step in the analysis, a posthoc test was applied to compare reaction times based on the interactions between condition (listen vs. repeat) and test type (immediate, 24-hour delayed, 9-week delayed). Table 3.9 shows the results of these pairwise comparisons with Bonferroni correction. The results indicate no significant differences in reaction times between the conditions at any of the test points. This suggests that the mode of condition (listening or repeating) does not significantly affect reaction times across immediate, 24-hour, or 9-week delayed tests.

Table 3.9: Post-hoc tests with Bonferroni correction for reaction time based on the interactions between condition and test model

Contrasts	Estimate	SE	df	t-ratio	p-value
(Immediate listen) - (Immediate repeat)	-325.4	194	1421.2	-1.68	1.0000
(24-hour delayed listen) - (24-hour delayed repeat)	14.4	185	1214.6	0.07	1.0000
(9-week delayed listen) - (9-week delayed repeat)	-245.4	129	1.904	-1.90	1.0000

* $p < 0.05$; ** $p < 0.01$; *** $p < .001$

3.7. Discussion

This pilot study investigated the production effect in L2 word learning and its durability over time (i.e., short-term and long-term word learning) in 4–11-year-old children. Two questions were addressed to look at the effect. The first question explored whether saying the words aloud while learning unfamiliar Modern Standard Arabic (MSA) vocabulary helps recognize newly learned words and their corresponding better than simply listening to them during learning. The second question addressed whether the child’s age and language dominance (i.e., Arabic or English) affect learning new L2 words in listening or repeating conditions. To answer the questions, three recognition tests were used to examine the production effect hypothesized to impact word learning and memory. The first test was conducted immediately after learning the vocabulary, the second test was conducted 24 hours after the learning phase, and the third test occurred nine weeks after the learning phase. Here, all children learned the exact unknown twenty L2 words (i.e., items) under the two conditions (i.e., listen and repeat), but the tests included an unequal number of items. Due to time limitations, children were tested in 10 words in the first two tests, but the third test included all 20 learned items. Overall, there was no significant difference according to condition, either overall or at any of the test points.

To address the first research question, the present study's findings show that the production effect was not observed. Specifically, there was no significant difference between the two conditions (i.e., listen and repeat) in accuracy and reaction time results. As discussed earlier, research has identified conditions where no significant difference is observed between produce and no-produce conditions (i.e., read silently, read aloud, sing, hear, etc), leading to a null effect. Experiments involving different types of production, such as reading aloud versus silently or even singing versus reading aloud, have shown varied results. For instance, Quinlan and Taylor (2013) demonstrated that singing items produced greater subsequent recognition than reading aloud. However, such effects are not always consistent across different contexts, indicating that the mode of production can significantly influence memory outcomes. While the production effect is generally robust, its null manifestation under specific conditions highlights the complexity of memory processes. This discussion explores the circumstances and explanations for such null findings, emphasizing the complexities in memory processes.

3.7.1. How do listening and vocal production conditions differ in their impact on vocabulary recognition accuracy and reaction times in both short-term and long-term contexts?

A key finding in this study is that the accuracy of vocabulary recognition remains relatively stable across tests for the listening condition. The repeat (vocal production) condition shows some improvement in accuracy at the 24-hour test and a decline at the nine-week test. However, these variations were not statistically significant ($p = 0.71$ for the condition effect). This suggests that while vocal production might initially enhance memory encoding, these benefits do not persist over longer periods without ongoing reinforcement. The stability observed in the listening condition indicates that it may support consistent long-term retention. This could be attributed to the reinforcement effect of repeated exposure through listening, which helps maintain and strengthen memory traces over time. The significant improvement in recall over time, with the nine-week test showing higher accuracy than the immediate test across both conditions ($p < 0.001$), suggests that long-term memory consolidation can occur regardless of the initial encoding strategy used. These findings align with previous research indicating that passive listening can effectively support the consolidation phase of memory, facilitating the transition of information from short-term to long-term memory (Engelkamp & Zimmer, 1997). The consistent performance in the listening condition may reflect its role in reinforcing and stabilizing memory without requiring active engagement. While both listening and repeat conditions contribute to vocabulary retrieval, the listening condition demonstrates a trend toward stable retention over extended periods.

One notable observation was the reaction time to correct answers across different conditions and test times. Contrary to expectations, the analysis did not reveal significant differences in reaction times between the listen and repeat conditions. The absence of significant effects suggests that vocal production might not substantially impact the speed of recall as initially hypothesized. This finding implies that both conditions facilitate retrieval processes similarly, with no measurable advantage in response time attributable to the repeat condition. Additionally, the lack of significant interaction effects between condition and test times indicates that the impact of vocal production on reaction times does not vary significantly over time. Although a slight increase in reaction times was observed in the 9-week test for both conditions, this change was not statistically significant, suggesting that memory decay might not be as pronounced or measurable through reaction time as previously thought.

In conclusion, this study found no significant differences between vocal production and passive listening regarding vocabulary recall accuracy, reaction time, or long-term retention. Both methods appear to facilitate vocabulary learning similarly, without one showing a clear advantage. These findings suggest that the effectiveness of active and passive learning strategies may depend on other factors not captured in this study.

3.7.2. How Does Test Time Influence Accuracy in Retention?

Recent results indicate that test timing significantly affects accuracy and reaction time, with participants demonstrating varying performance levels across immediate, 24-hour, and 9-week delayed tests. Notably, accuracy rates improved over time in listening and repeating conditions, with a marked increase observed during the 9-week test. This trend contrasts with typical vocabulary research, where a decline in retention is often noted between immediate and delayed assessments.

One potential explanation for these findings is a ceiling effect in the earlier tests. The immediate and 24-hour delayed assessments involved a limited number of items, which may have resulted in scores clustering near the maximum achievable accuracy. This clustering could obscure differences in participants' learning and recall abilities, masking distinctions between the listening and repeating conditions. In contrast, the 9-week delayed test included all learned words, providing a broader range of items that may have offered a more accurate reflection of participants' retention capabilities, thus reducing the ceiling effect.

Another explanation is informed by research from Mama and Icht (2018), which explored the timing of vocal production. Their findings suggest immediate vocalization does not consistently enhance recall compared to conditions without vocal production. This indicates that task timing may heavily influence vocalization's cognitive benefits. Immediate vocalization might not allow sufficient processing time or could introduce interference from other cognitive tasks, thereby diminishing its effectiveness for memory retention. In this context, delayed vocal production might be more beneficial for retention as it facilitates better consolidation of information over time.

The observed increase in recognition scores across sessions also diverges from the expected decline typically reported in vocabulary studies. This counterintuitive trend may suggest that participants encountered the items outside the study context, perhaps driven by curiosity or incidental exposure. Alternatively, this improvement could reflect the natural consolidation of knowledge over time, showcasing the brain's capacity to process and solidify information beyond the initial learning phase. These possibilities highlight the complexity of retention processes and underscore the importance of caution when interpreting such results.

In summary, while findings indicate an overall improvement in accuracy over time, several factors warrant careful consideration: the potential ceiling effect, the timing of vocal production, and external influences on retention. This complexity emphasizes the need for further research to disentangle these factors and enhance our understanding of long-term vocabulary retention processes.

3.7.3. Can the Production Effect Enhance Memory When Combined with Other Cognitive Processes (e.g., Picture Superiority Effect)?

The interaction of the production effect with other cognitive processes can significantly influence its effectiveness in enhancing memory. As stated above, the production effect refers to the phenomenon where actively producing words (e.g., saying them aloud) improves memory recall compared to passive reading. However, when combined with other cognitive processes, such as the picture superiority effect (where images are remembered better than words), the benefits of the production effect might be less pronounced. Fawcett et al. (2012) investigated this interaction and found that the simultaneous application of the production and picture superiority effects did not result in additive memory benefits. This lack of additive effect implies that when multiple distinctive processes are engaged concurrently, the distinctiveness that typically enhances memory recall for produced items may be diluted. Consequently, produced items might not stand out as much, leading to no significant improvement in memory (Fawcett et al., 2012).

Further exploration into the interplay of these cognitive processes reveals that the competition for cognitive resources could be a critical factor. When multiple distinctive processes, such as production and visual imagery, are activated, they might compete for the same cognitive resources, thereby diminishing their overall effectiveness. This suggests that the human cognitive system has limitations in simultaneously processing and benefiting from multiple distinctiveness-enhancing strategies. Supporting this, Zormpa et al. (2018) found that both the production and picture superiority effects improve item memory in a picture naming task. However, they observed that the generation effect (i.e., better memory for generated than read words) also contributes significantly to the memory advantage in picture naming. Zormpa et al.'s study demonstrated that while the production and generation effects can coexist, their combined impact on memory enhancement is complex and not always straightforward. These findings underscore the importance of understanding the intricate dynamics between various cognitive processes to optimize learning and memory strategies (Zormpa et al., 2018).

3.7.4. Does the Experimental Design Impact the Effectiveness of the Production Effect?

The experimental design, specifically whether a between-subject or within-subject design is used, can influence the observed effects of the production effect. Fawcett's (2013) meta-analysis found that while the production effect is generally moderate, it tends to be more pronounced in within-subject designs. This suggests that the benefits of the production effect might be more noticeable when participants can directly compare produced and non-produced items within the same session due to increased relative distinctiveness. Despite this, the current study used a within-subject design and observed no significant difference between produced and non-produced items.

In contrast, Bodner and Taikh (2012) conducted a study involving 60 undergraduate students to reassess the basis of the production effect in memory, which refers to the memory advantage for items studied aloud compared to those studied silently. They examined this effect using within-subject and between-subject designs and found robust evidence of the production effect in recognition memory. Specifically, participants who studied items aloud showed better recognition performance than those who studied items silently, even when the manipulation was applied across different groups. Interestingly, their results challenged the distinctiveness account, which suggests that producing items makes them more distinctive and, thus, more memorable. They found no evidence that recognition of aloud items was enhanced more within-subjects than between-subjects.

Additionally, the study revealed potential costs associated with the production effect in mixed designs, where some items are read aloud and others silently within the same list. Recognition of silent items was impaired in mixed designs compared to blocked designs or pure silent conditions. Bodner and Taikh (2012) suggested that encoding all items aloud might be more effective for increasing overall memory accuracy than only a subset.

3.7.5. How Does Age Affect the Production Effect on Memory in Children? Specifically, how Does Age Influence Memory Accuracy and Reaction Times in the Production Effect?

Age has emerged as a significant variable in studying the production effect on memory, particularly in children. The production effect refers to the phenomenon where actively produced items (spoken or written) are better remembered than passively received items (heard or silently read). This effect, however, is not uniform across different age groups, exhibiting distinct patterns in older and younger children. This discussion critically examines the impact of age on the production effect, focusing on memory accuracy and reaction times across various conditions and test timings.

Previous research has demonstrated a typical or positive production effect in older children, specifically those aged six years and above. Icht and Mama (2014) observed that older children exhibited better memory for produced items compared to heard items, a finding corroborated by Pritchard et al. (2019). This positive production effect can be attributed to the developmental advancements in cognitive processes and verbal abilities that occur with age. These advancements likely enhance the encoding and retrieval processes involved in memory, thereby facilitating better recall of actively produced items. Conversely, younger children, particularly those aged four and below, have shown a reversed production effect, wherein they remember heard items better than produced ones. Studies by Zamuner et al. (2018) and López Assef et al. (2021) highlighted this phenomenon, suggesting that the cognitive and linguistic skills required to benefit from the production effect may not fully develop in younger children. This reversed effect underscores the importance of developmental considerations in understanding memory processes and suggests that different mechanisms may underlie memory performance at different developmental stages.

In the current study, age was a primary factor influencing accuracy rates across three distinct test points: immediate, 24-hour, and 9-week delayed. Older participants consistently demonstrated higher accuracy rates, indicating a positive correlation between age and memory accuracy. This finding aligns with the finding that working memory capacity increases with age during childhood. A study examining visual working memory in children aged 6-8 and 11-13, as well as college students, found that memory performance improved significantly from each age group to the next, even when controlling for factors like encoding ability and verbal strategies (Cowan et al., 2011).

Age also played a critical role in moderating reaction times in the study. Older children exhibited faster reaction times, reflecting the developmental progress in cognitive and motor functions that typically improve with age. These findings suggest that the neural and motor pathways involved in response execution become more efficient as children grow older, facilitating quicker responses. Age-related increases in speed on motor reaction tasks have been observed between ages 7-8 and late adolescence, attributed to maturation of neural networks in motor and cognitive brain systems (Alhamdan et al., 2022). Their finding aligns with the current results. Older children responded more swiftly across different conditions and test timings, indicating that developmental advancements contribute to overall cognitive efficiency. Notably, neither condition nor test timing independently influenced reaction times significantly, and no significant interactions between condition and test timing for accuracy were observed. This implies that age is a critical determinant of reaction time and overall cognitive performance, underscoring the importance of considering developmental stages in memory and cognitive research.

The findings of this study underscore the pivotal role of age in the production effect on memory. Older children demonstrated superior accuracy and reaction times, highlighting the developmental benefits of learning efficiency. These results suggest that age-related changes in cognitive functions significantly impact memory processes and should be carefully considered when designing interventions to enhance learning outcomes, particularly in language acquisition contexts.

3.7.6. How Does Language Dominance Influence Accuracy and Reaction Time in L2 Learning Across Listening and Repeating Conditions?

In the current experiment, participants were children from diverse Arab backgrounds exposed to various accents and dialects of Modern Standard Arabic (MSA) (e.g., Saudi, Levantine, Yemeni, etc.). They attended British schools while also learning MSA through online or weekend Arabic schools. Their language dominance was determined by the primary language spoken at home, and they were grouped into Arabic or English-dominant categories. All participants had prior phonological exposure to MSA sounds and knowledge of the Arabic alphabet, providing a foundational basis for examining how language dominance influences

MSA acquisition and production. The current experiment examined how language dominance (Arabic vs. English) impacts accuracy and reaction time in recognizing Modern Standard Arabic words among young learners across different learning conditions (listen vs. repeat) and testing intervals (immediate, 24-hour delayed, 9-week delayed). Language dominance did not significantly influence accuracy or reaction time in various conditions and intervals, except for a specific interaction where Arabic-dominant participants showed faster reaction times in immediate testing.

As discussed in sections 2.2.4 and 2.2.6, the linguistic landscape of the Arab world consists of Classical Arabic, Fus'ha, Modern Standard Arabic (MSA), and colloquial dialects. Classical Arabic, the language of the Quran, forms the basis for Fus'ha and MSA, used in formal documents, education, and media. Colloquial dialects, differing widely across regions, are the everyday spoken languages, with MSA unifying communication across Arabic-speaking communities. In this context, no statistical significance was expected in the language background when learning Modern Standard Arabic (MSA) as a second language, given its distinct nature compared to acquiring dialectal languages. This supports previous studies suggesting that learning MSA differs from acquiring colloquial or home-spoken languages, thus positioning MSA acquisition within the framework of second language (L2) studies (Alresaini, 2016).

Alresaini (2016) examined the acquisition of Modern Standard Arabic (MSA) among speakers of various Arabic colloquial dialects. The participants were 147 adolescents from diverse Arabic-speaking regions such as Egypt, the Levant, and the Gulf. The study aimed to determine whether MSA grammar among these speakers could be classified as native-like despite MSA not being a language of daily conversation and being acquired primarily through formal education. The research focused on how the age of first exposure (AoE) and the speakers' knowledge of their native dialects (L1) influenced their proficiency in MSA. The findings indicate that the diverse colloquial Arabic dialects spoken by the participants as their first language (L1) did not significantly impact their acquisition of Modern Standard Arabic (MSA). Notably, the initiation of MSA learning at age 6 suggests that participants were well within the sensitive period for acquiring critical syntactic structures. Furthermore, despite their varied dialectal backgrounds, the uniform performance across participants demonstrates their capability to reconfigure their distinct L1 linguistic frameworks to align with a standardized

MSA grammar. This adaptability underscores significant cognitive flexibility and linguistic competence in the face of language variation and learning.

This study found that language dominance (Arabic vs. English) did not significantly influence accuracy in recognizing MSA words across different learning conditions and testing intervals. This aligns with Albirini (2014), who observed that heritage speakers familiar with colloquial Arabic may initially benefit from their phonological and syntactic familiarity. However, this advantage does not persist at advanced learning stages. The study's results suggest that both Arabic- and English-dominant learners can achieve comparable levels of accuracy in MSA word recognition, irrespective of whether they learned through listening or repeating.

The study's findings on reaction time revealed no significant main effects of language dominance, learning condition, or testing interval. However, an interaction effect was observed where Arabic-dominant participants showed faster reaction times under immediate testing conditions. This suggests a role of language dominance in immediate cognitive processing, possibly due to the faster access to the phonological loop of MSA for Arabic-dominant learners. However, the decline in delayed tests is supported by Haddad (2006), who noted learners' cognitive challenges when switching between MSA and colloquial dialects. The participants' familiarity with the phonological structures of colloquial Arabic may expedite immediate processing in MSA, although this advantage does not translate to long-term improvements. Thomure et al. (2021) indicate that young Arabic speakers often face challenges in accessing MSA phonemes, which are linked to difficulties in the phonological encoding of MSA words. This suggests that the phonological structures of colloquial dialects can interfere with the accurate processing of MSA phonemes. The diglossic nature of Arabic, where MSA and colloquial dialects coexist, affects phoneme isolation tasks. According to Thomure et al. (2021), children often struggle with isolating initial phonemes and singleton phonemes in MSA, which can be attributed to the differences in phonological structures between MSA and their native dialects. Therefore, although the present study meticulously controlled for prior knowledge and the participants' proficiency level in Modern Standard Arabic (MSA), the observed advantage among Arabic-dominant learners was transient, suggesting a short-term processing benefit rather than a deep, enduring understanding of MSA. In other words, language dominance does not significantly affect long-term accuracy or reaction times in L2 learning of MSA which aligns with Alzu'bi et al. (2023). They found that current spoken Arabic dialects are an extension of ancient dialects. They are linguistic patterns that exist alongside

Classical Arabic, one of these dialects. However, for religious and economic reasons, they showed their dominance over other dialects and became the origin that was followed by others. In addition, contrary to the assumptions that the Arabic language is unified, it was divided into two levels: the Eloquent, which conveyed poetry and literature, and the dialects that are cognate to the dialect patterns prevalent in the past, with which people spoke and communicated among each other.

The multifaceted nature of Arabic's historical development underscores the complexities and challenges face its learners, irrespective of their dominant language when acquiring MSA. The current results underscore the importance of integrating effective teaching strategies that accommodate learners' cognitive and linguistic backgrounds. As Kharrat (2018) advocates for functional language use in textbooks incorporating practical and contextual learning experiences can enhance outcomes for both Arabic- and English-dominant learners. Moreover, teaching MSA as a lingua franca, as proposed by Jaradat and Al-Khawaldeh (2015), can bridge the gap between different dialects and support learners in achieving practical communication skills across formal and semi-formal contexts.

Thus, the current findings show that language dominance is limited in influencing accuracy and reaction time in L2 learning of MSA under listening and repeating conditions. While it does not significantly impact long-term accuracy or reaction times, immediate testing conditions reveal faster cognitive processing among Arabic-dominant learners. These findings emphasize the importance of understanding the cognitive processes and memory mechanisms involved in learning MSA. Investigating the dynamic relationship between language dominance and learning conditions can enhance the comprehension of MSA vocabulary acquisition. The historical and linguistic context provided by Alzu'bi et al. (2023) further underscores the complexity and resilience of Arabic language learning.

3.8. Conclusion and Limitations

The findings from this pilot study suggest no significant Production Effect (PE) in L2 word learning among children aged 4 to 11. No notable differences in accuracy or reaction times were found between the "listen" and "repeat" conditions, indicating that vocal production may not significantly enhance L2 vocabulary learning in this sample. However, given the exploratory nature of this study and its limitations, no definitive conclusions about the effectiveness of the Production Effect can be drawn.

This pilot study has highlighted several methodological areas that can be refined to enhance the sensitivity and validity of findings in subsequent research on the Production Effect (PE) in L2 vocabulary learning among children. The insights gained from the study have informed specific adjustments that will be implemented in future experiments to address these limitations more effectively.

1. Incorporating a Norming Task:

The absence of a norming task (such as an English naming task) made it difficult to control for variability in item familiarity or difficulty across participants. To address this, future studies will include a norming task with a comparable group of children to establish a baseline for item difficulty. This control will help standardize the experimental items and ensure more reliable measurements of vocabulary learning outcomes.

2. Adding a Pre-Test to Measure Prior Knowledge:

Without pre-testing for baseline familiarity with the target vocabulary items, the study lacked a precise measure of each participant's initial knowledge. Future studies will incorporate a pre-test to assess prior knowledge, ensuring that differences in learning can be more accurately attributed to the Production Effect rather than varying levels of familiarity with the words.

3. Increasing the Number of Trials in Early Tests:

In the pilot study, limiting the immediate and 24-hour delayed tests to ten trials helped maintain engagement but reduced data depth at these critical intervals. In future experiments, increasing the number of trials in early tests will enhance sensitivity, enabling a more detailed analysis of differences between conditions. This will allow for a more robust examination of the PE across different time points.

By implementing these methodological improvements, the following studies will better address the pilot study's limitations, providing a more robust framework to evaluate PE's impact on L2 vocabulary learning. These adjustments ensure a more accurate and comprehensive exploration of the research questions, leading to more precise and reliable findings.

Several limitations of the pilot study must be acknowledged, as they may have influenced the overall findings. First, this pilot did not include a norming task (naming task in English) with a comparable group of children. The absence of this control step means that item difficulty or familiarity could have varied between participants, potentially affecting their performance.

Moreover, prior knowledge of the vocabulary items was not assessed before the study. This lack of pre-testing introduces uncertainty regarding the baseline familiarity of the participants with the items, which may have impacted learning outcomes. A further limitation is the reduced number of trials in the immediate and 24-hour delayed tests. Due to time constraints and concerns about participant fatigue, these tests included only ten trials (five per condition). While this decision minimized testing time and maintained engagement, it also limited the depth of data collected during these critical periods. In contrast, the 9-week delayed test included all 20 learned items, providing a more comprehensive evaluation of retention at that stage. However, the limited number of trials in the earlier tests may have restricted the sensitivity to detect more subtle effects between conditions. In future research, including a control group that completes all tests without exposure to the target items would offer a valuable baseline for assessing performance. This would enhance the findings' reliability and help isolate learning effects from potential confounding variables such as practice effects or incidental exposure.

This pilot study provides a foundation for the development of subsequent experiments. The following experiments will incorporate longer lists to investigate the Production Effect under classic conditions (i.e., listen and repeat). Moreover, extending the time frame for memory retention assessments will offer a deeper understanding of the durability of the Production Effect in L2 word learning.

Chapter 4. Experiment One

In the pilot study (Chapter 3), the production effect (the PE hereafter) was investigated using a simple learning list (10 items per condition) to explore the effects of age, language background, and, most importantly, whether "listen" or "repeat" conditions would enhance memory when learning unknown MSA words. Age showed an effect, but no significant effect was found for the learning method (i.e., listen vs. repeat) or the language background. The rationale for the current study was to conduct a revised version of the pilot to provide a more comprehensive examination of the production effect. Specifically, the small sample number of items in the pilot and the simplicity of the learning list (only ten items per condition) limited the ability to detect more subtle effects of the learning conditions and language background. Additionally, the variability in participants' age and language background could have introduced confounding factors that obscured the potential effects of the production effect.

Hence, the current experiment controlled for participant variables by including only six-year-old, multilingual children whose dominant language is English, and used an extended learning list of 64 words divided into two lists based on the condition (listen and repeat) refined to increase the likelihood of unknown items. To better understand the long-term effects of the production effect, the study employs both a free recall test and a recognition test across three test times (immediate, one-week delayed, and two-week delayed). This approach allows for a comprehensive assessment of how the production effect influences not only immediate recall but also the retention of information over time. The current study also implemented lexical competitors (i.e., items that share semantic or phonological features with target words) during the recognition test. The study included phonological and semantic competitors as part of the design, but they were not manipulated to make claims about lexical processing, as the focus of the thesis is on vocabulary learning. These competitors were included in the testing phase to explore how different distractors might interact with vocabulary learning, providing insight into how closely related words could influence recognition and retrieval processes.

These revisions are designed to build on the findings of the pilot by providing a more robust test of the production effect, focusing on the durability of learning and the influence of controlled variables. The refined methodology aims to yield more definitive insights into how different learning conditions affect memory for unknown MSA words.

4.1. Introduction

As discussed in section 2.2, the production effect (PE) refers to the improved retention of information produced aloud compared to information read silently (MacLeod et al., 2010; Icht et al., 2014). This effect enhances memory recall, although its significance can vary depending on specific conditions, leading to mixed evidence in the literature. The PE has been investigated through recall and recognition tests to understand how words are remembered and learned (Bodner & Taikh, 2012; MacLeod & Bodner, 2017; Willoughby, 2020). Producing words aloud highlights their phonological structure, allowing learners to leverage prior knowledge and create distinct memory representations. This distinctiveness is believed to result from increased cognitive effort, enhancing memory consolidation and retrieval (MacLeod et al., 2010; Jamieson et al., 2016).

Positive effects of the PE have been observed in both L1 and L2 adult learners (Gathercole & Conway, 1988; MacLeod et al., 2010; Forrin et al., 2012; Zamuner et al., 2016; Icht et al., 2020; Ellis & Sinclair, 1996; Kaushanskaya & Yoo, 2011; Icht & Mama, 2019) and in children aged 5 to 10 (Icht & Mama, 2015; Zamuner et al., 2018; Pritchard et al., 2019; López López Assef et al., 2021). However, the effect may reverse or show no difference when non-native or unfamiliar sounds are involved, as seen in adult learners of non-native words (Kaushanskaya & Yoo, 2011; Thorin et al., 2018; Baese-Berk & Michaud, 2019) and children aged 2 to 4 (Zamuner et al., 2018; López López Assef et al., 2021). Few studies, including Experiment 1 of this thesis, have reported null effects where participants show similar performance across different production conditions (Thorin et al., 2018; Zamuner et al., 2018; Baese-Berk & Michaud, 2019; López López Assef et al., 2021).

Other studies show that production might only support the initial stage of learning but hinders the later stages. The study by Zamuner et al. (2016) identified a positive effect of production on word learning during the initial stages. However, this effect diminished in the later stages. Similarly, Kapnoula and Samuel employed eye-tracking experiments with adult participants to investigate the impact of production on word learning. Their main finding was that production influences different stages of word learning differently. Consistent with Zamuner et al. (2016), they found that production facilitates early stages of learning but impedes word learning in later stages. This indicates that the effect of production evolves throughout the learning process. Production is reported to disrupt the learning process, caused by a heavier cognitive load from several factors, including task structure and presented stimuli (Kaushanskaya & Yoo, 2011; Baese-Berk et al., 2016; Baese-Berk & Michaud, 2019). Moreover, Bodner and Taikh (2012) found that the distinctiveness account of the production effect might not always apply, as their results included negative production effects in certain list-discrimination tasks. These findings suggest that production may ultimately hinder lexical integration, the process of incorporating new words into the existing mental lexicon. Producing words during learning may increase cognitive load, potentially detracting from other aspects of word learning, such as understanding meaning or achieving correct pronunciation. These findings also indicate that the production effect can be context-dependent and may not universally enhance memory.

In studies examining the production effect (PE) in children, the learning lists varied between studies (Table 4.1), including familiar words or novel non-words (Zamuner et al., 2007; Icht & Mama, 2015; Zamuner et al., 2018; Pritchard et al., 2020; López Assef et al., 2021). Previous research has demonstrated that the strength and direction of the PE depend on the linguistic characteristics of the stimuli, whether they are familiar native-sounding words or non-words. However, this work has primarily focused on manipulating unfamiliar lexical items with native sounds and using extensive learning lists with children. These approaches aim to assess memory retention rather than evaluate the durability of newly learned words.

Table 4.1: List length and test used in the studies of the PE in children

Source	Participants	Conditions	List length	Test
Icht and Mama, 2015	30 (5 years old)	Look (observation); Look and listen; Look and say.	30 pictures of Hebrew bisyllabic nouns (e.g., tiger, closet).	Free recall
Zamuner et al., 2017	16 (4.5- 6.0 years old)	Heard; Produced	8 CVC novel words (e.g., wis, zel, jig, mig).	Recognition
Pritchard et al., 2019	Experiment (1): 41 (7- 10 years old) Experiment (2): 40 children.	Read aloud-read silently	Experiment (1): 40 familiar printed words. Experiment (2): 80 monosyllabic novel words (e.g., hest, prench).	Recognition
López Assef et al., 2021	Experiment (1): 120 (3- 6 years old) (n=30 for each age group). Experiment (2): 30 (2 years old).	Look, Listen, Say	30 monosyllabic familiar English words (e.g., boat, bath, bird).	Free recall

Childhood PE studies have systematically concentrated on short-term memory, using immediate recall or recognition tests. Only one study has explored the effect of PE on the durability of learning in adult second language (L2) learners (Icht & Mama, 2022). As previously mentioned, the current study employed an extended learning list of 64 words to test recall and recognition over three intervals: immediate, one-week delayed, and two-week delayed. This approach was designed to evaluate PE's effect on learning's durability. In this study, children learn words through hearing and associating them with images. Recognition is expected to be enhanced by this method due to the visual cues provided by the images. However, a positive effect on recall is also anticipated, as memory traces may be strengthened by the combination of auditory and visual learning, although possibly to a lesser extent than recognition due to the lack of immediate retrieval cues.

Word learning is believed to be closely associated with and supported by phonological short-term memory (Baddeley et al., 1988; Gathercole, 2006). The general concept is that producing a word requires the brief retention of its phonological sequence in phonological short-term memory, which, over time, facilitates longer-term learning. Thus, including phonological and semantic competitors in this study helps to evaluate the degree of lexical integration by observing whether newly learned words can compete with established lexical items, indicating successful integration into the mental lexicon. The relevance of phonological competitors lies in their ability to provide insight into the level of integration, as competition with similar-sounding words indicates that the newly learned words have been processed and stored alongside existing vocabulary, thus demonstrating the effectiveness of word-learning strategies used in the study.

While previous studies that examined PE in childhood have focused on immediate tests (i.e., direct and short learning effects on memory), the objective of this study is to explore how PE affects the durability of learning in L2 child learners during three testing times (immediate, one-week, two-week). The current study makes several predictions regarding the production effect (PE) and related factors. A null PE (similar performance across conditions) is anticipated, possibly due to a mismatch between target words and their lexical associations, increasing cognitive demand. This prediction is based on the idea that the production advantage may diminish when cognitive resources are taxed. Also, children are predicted to perform better on delayed tests than on immediate ones. This prediction is consistent with the results of the previous pilot study, where delayed post-tests revealed enhanced retention over time. The current study extends this by examining how the combination of auditory and visual learning impacts long-term retention compared to immediate recall. According to Roediger and Karpicke (2006), when an unfamiliar word form is associated with a visual referent, this referent is subsequently used during training or testing to prompt the production of its newly learned label. Retrieval practice is well-documented to enhance information retention, a phenomenon known as the testing effect. Supporting this hypothesis, Karpicke and Roediger (2008) trained English-speaking adults on Swahili-English word pairs using different training conditions: 1- involving repeated testing (which included recall) and 2- involving repeated studying (which did not include recall). Unlike repeated studying, which had no impact on delayed recall, repeated testing significantly facilitated retention.

4.2. Methods and Design

The current experiment investigates the production effect (PE) on learning unfamiliar Modern Standard Arabic (MSA) words under two conditions (listen and repeat) in multilingual children aged six whose dominant language is English. Specifically, the study addresses the research question: How does the production effect influence the learning of unfamiliar MSA vocabulary in multilingual children whose dominant language is English? A within-subject mixed-list design is employed to mitigate the frequency effect associated with exposure to conditions, which can exaggerate performance differences in a pure-list design (McDaniel & Bugg, 2008; MacLeod et al., 2010). By including both conditions within the same list, this design prevents participants from developing strategies tailored to a single condition, providing a more accurate assessment of the PE. The children learned the words in a mixed order, with conditions presented randomly on the screen. To equalize exposure to the words, children listened to the "listen" items twice and the "repeat" items once, since the word is heard a second time through self-production. This approach ensured that each item was heard twice across both conditions: either twice by the computer in the "listen" condition or once by the computer and once through self-production in the "repeat" condition. By reducing and equalizing the number of exposures, this method prevents frequency effects from skewing the results, allowing for a more precise evaluation of the PE's impact on learning. The study included three phases: Training, learning, and testing. Children were first trained on eight unfamiliar words. Two steps were followed in the training: (a) familiarize the children with the conditions; and (b) familiarize the children with the testing. Children learned new words under two conditions: Listen (children see the image and hear the associated word twice); and Repeat (children see the image, hear the associated word once, and produce the word once). After learning the conditions, there was a short break, followed by a free recall task and a forced choice recognition test. The learning items were presented using PowerPoint, and the testing was conducted online using Gorilla Experiment Builder (https://app.gorilla.sc/admin/home), a web-based experiment builder and deployment platform, but children participated in person with the presence of the experimenter at Amanah Muath Trust (Islamic and Arabic school based in Birmingham, the UK).

4.3. Participants

Participants were 30 students (12 males, 18 females, aged 6;0- 6;9 years old, M=6;10, SD= 10.38, range = 9 months) recruited from Amanah school. Age was controlled to address potential cognitive abilities across different age groups. Participants were grouped into A and B to control word conditions, as explained in the following section. Ethics approval was obtained from the University of Birmingham Research Ethics Board, and consent was obtained from participants' parents. Participants received a classic story book (i.e., "Guess how much I love you" by Sam McBratney) as compensation for their participation. Parents completed LBQ (i.e., Language Background Questionnaire) to ensure that the participants met the language requirements for the current experiment. According to parents' reports, all children were neurotypical with normal or corrected-to-normal vision and normal hearing and no background in any language impairment or delay (e.g., dyslexia). Participants were required to have been exposed to English from birth but have a phonological awareness of Modern Standard Arabic (i.e., English is their L1, but they know the sound of Arabic letters) to control any effect of language dominance and to minimize task difficulty on the children when learning (i.e., longer time in processing or recognizing speech sounds of learning materials due to unfamiliar phonemes).

All participants were recruited from the 1st grade classroom as it is the introductory level (i.e., having the basic phonemic knowledge to distinguish the phonemes but they do not have the ability to understand the language without translating it to English according to the school's entrance test and parents' report). This entrance test is designed and conducted by Amanah Muath Trust to determine the child's language level prior to enrolling in the school. The language levels are divided into six levels or grades (i.e., from grade one to grade six). The children are placed in the appropriate grade according to their language level, rather than their age. Ten children were additionally recruited but were excluded from the analysis as they either refused to take the test after the learning phase (n= 3), dropped from the study (n= 4), or failed to follow the instructions in the condition (i.e., spoke during “listen” condition or remined silent in the “repeat” condition) (n= 3). All 30 remaining children completed the recall and recognition tests.

4.4. Materials

The stimuli included 72 unfamiliar Modern Standard Arabic nouns. Eight words were assigned to the training list, and 64 words were assigned to the study list and divided according to the condition (i.e., 32 words in each condition). Word condition was counterbalanced, that is, if a word is under the “listen” condition in group A, the same word is assigned under the “repeat” condition in group B. The average word length for letters was 3.9 and the number of syllables was 2.22. Words were selected from various Arabic storybooks written and published by various Arab children’s publishers (e.g., Kadi and Ramadi, Boustany’s Publishing House). Some of the extracted words were taken from books that were translated from famous English stories into Arabic (e.g., The Gruffalo or Al-Gharfoul); they were then compared to the curriculum being studied by the children at the Arabic school to exclude all words mentioned in the textbooks that might increase the likelihood of word familiarity.

A familiarity questionnaire was administered to caretakers and teachers in Arabic and English, and it invited them to comment on how familiar they thought the items would be to the children in charge. All nouns likely to already be familiar to the children were excluded from the list of selected words. Moreover, before the experimental training, the experimenter asked the children (Do you recognize the object presented? Can you name it in Arabic?). Any familiar word would be excluded from the analysis. The final chosen list included 72 stimuli that a female native speaker of MSA pre-recorded. The visual stimuli consisted of colored images digitally illustrated using Adobe Illustrator (Appendix 5). The volunteer artist carefully drew the visual stimuli to ensure clarity and relevance to the words they represented. All images were high-quality, age-appropriate, and easily identifiable by children. To verify that the children recognized and correctly identified the images, a pre-study familiarization session was conducted with children of the same age as the participants (note that these children were not participants in the study). During this session, the children were shown the images and asked to name or describe them. Only the images that were consistently recognized and correctly identified by most of the children were included in the final study set. If necessary, images were amended for clarity before being included. This process ensured the children could easily associate the images with the corresponding words during the learning tasks.

Table 4.2: List of words used in the study. Words were learned in isolation but were selected to include lexical associations

MSA	Translation	MSA	Translation	MSA	Translation	MSA	Translation
/ˈtu:t/	blueberry	/hu:t/	whale	/karaz/	cherry	/fara:ʃa/	butterfly
/kalb/	dog	/qalb/	heart	/ðajl/	tail	/ʃaʒra/	tree
/qalam/	pen	/ʕalam/	flag	/daftar/	copybook	/ʔarnab/	rabbit
/naħla/	bee	/naxla/	palm_tree	/ˈʕəsəl/	honey	/mursa:m/	pencil
/mo:z/	banana	/lo:z/	almond	/tuffa:ħ/	apple	/ʔasad/	lion
/qitˤtˤa/	cat	/batˤtˤa/	duck	/ħali:b/	milk	/kursij/	chair
/miθqa:b/	brace drill	/miħra:b/	sanctuary	/minʃa:r/	saw	/qunfuð/	hedgehog
/ʔtˤa:r/	frame	/ʔza:r/	garment	/lawħa/	picture	/sˤunbu:r/	tap
/watar/	cord	/wabar/	fur	/ʕu:d/	oud	/ta:d͡ʒ/	crown
/dabu:s/	pin	/dabu:r/	wasp	/xajtˤ/	string	/kawkab/	planet
/ʕiʒil/	calf	/fiʒil/	radish	/θawr/	ox	/ʔunbu:b/	tube
/ʤiðr/	root	/ʤamur/	coals	/nabtah/	plant	/maɣna:tˤi:s/	magnate
/sˤafi:ħa/	metal plate	/sˤaħi:fa/	newspaper	/ɣawwa:sˤa/	submarine	/jaʕˈsu:b/	dragonfly
/tˤi:n/	mud	/ti:n/	fig	/faxxa:r/	pottery	/zˤarf/	envelope
/kaʔs/	cup	/faʔs/	axe	/ʔibri:q/	pitcher	/tˤa:wu:s/	peacock
/raff/	shelf	/daff/	tambourine	/xiza:na/	closet	/sˤabbar/	cactus
/ʤanzi:r/	chain	/xanzi:r/	pig	/ˈʃa:ħinah/	truck	/midfaʔah/	fireplace
/wari:d/	vein	/bari:d/	postbox	/ðira:ʕ/	arm	/dˤabʕ/	hyena

4.5. Procedure: Training, Learning, and Testing

The first session included three stages: (a) participants were trained at the beginning with 8 MSA words so as to be familiarized with the study procedure as mentioned earlier; (b) the learning phase where children learned 64 words; and (c) the testing phase that includes a free-recall task and a recognition test. Prior to the experiment, a picture-naming task was administered to determine word familiarity in the participating children, including words appearing in the training list. At the beginning of the task, the children were asked by the experimenter (Do you recognize the object presented? Can you name it in Arabic?). Any familiar word would be excluded from the analysis. However, no children in the current study could name the words in MSA, even though the list contained highly frequent words and images (e.g., banana /mo:z/, duck /batˤtˤa/). Each child spent around one hour with the experimenter in the 1st week for the learning and immediate test tasks, then 30 minutes for the one-week delayed test, and 30 minutes for the two-week delayed test. The whole experiment took around two hours with each participant.

4.5.1. The training phase

Eight frequent but unfamiliar MSA words were randomly assigned to the two conditions for training purposes. Words were presented with an instructional emoji that represented each condition. The training slides started with an encouraging smiley emoji and a "Let's Practice!" phrase followed by the training slides. Both training and learning phases were done using PowerPoint. Each child learned about the experimental conditions using instructional images (Figures 4.1 and 4.2). If the image presented an ear and a cross over the mouth, they were instructed to listen to the word twice and to remain silent. The words were to be repeated once when the image presented an ear and a smiley mouth. During this phase, the experimenter pointed to the instructional image to remind the child of the condition. Each item was presented auditorily and as an image that appeared on the screen for 10 seconds, and there were 10 seconds between the items.

After a short break, all children were trained to perform the free recall task, then a forced choice recognition test. Firstly, the experimenter asked the children to tell the assistant teacher all the words that they remember (i.e., "can you tell Ms. Iman all the words that you learned today?" and after that they were prepared for the recognition test.

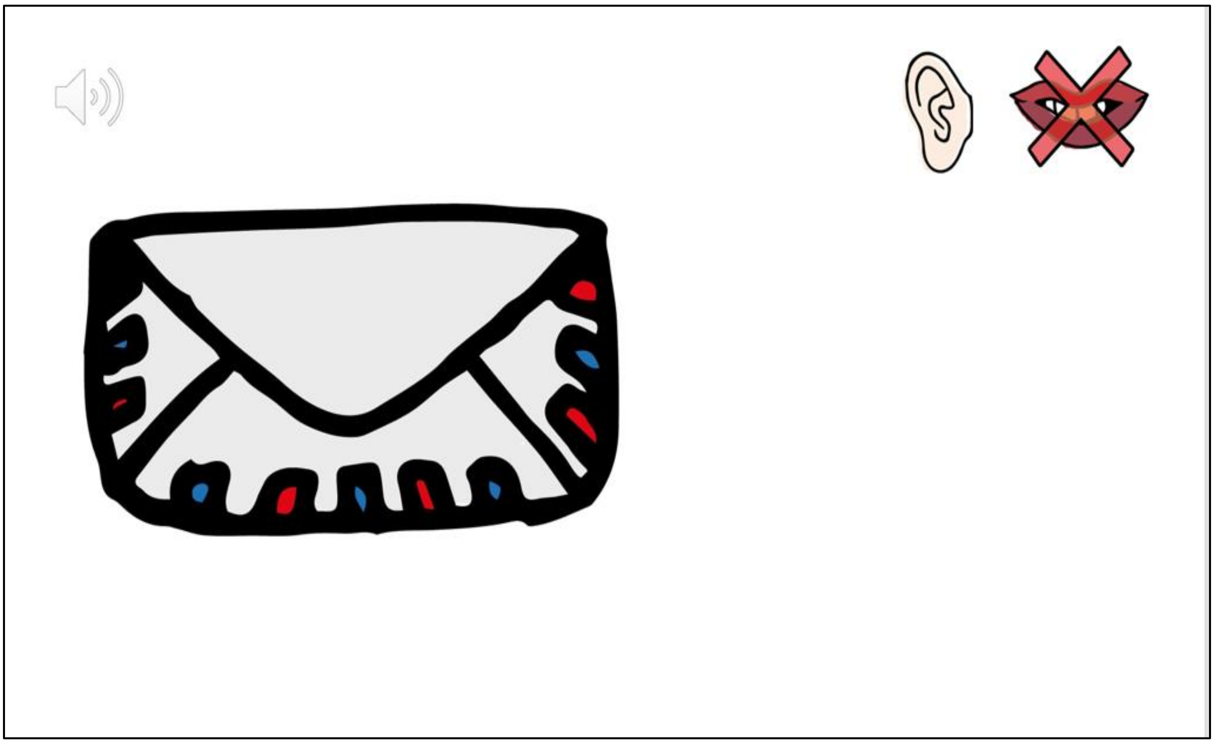


Figure 4.1: Example of visual instructional stimuli: repeat (production)

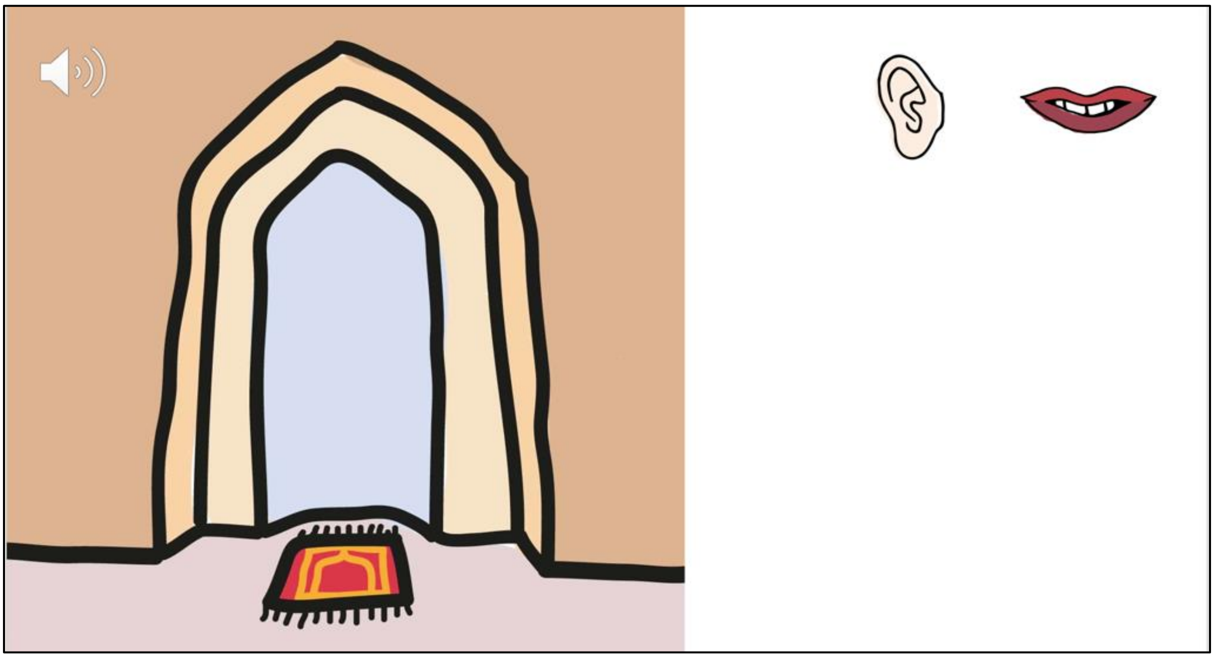


Figure 4.2: Example of visual instructional stimuli: listen and remain silent

4.5.2. The learning phase

After finishing the training phase, the children took a 5-minute break. Printed colouring sheets were used during the break between training and the recall task and the 5 minutes between the recall and the recognition task. The learning slides started with an encouraging emoji that appeared with the phrase "Let's learn!" appeared on the screen. The learning trials have the same structure as training, but they included 64 MSA words that were randomly divided into two sets according to the conditions of this study: listen and repeat. Every child learned the same words, but the words were assigned to different conditions according to the group.

4.5.3. The testing phase

The test phase was divided into three parts: recall, 5- minute break, and recognition task. During the recall phase, children were prompted to recall as many words as possible (e.g., 'Can you tell me the words you remember?'). A 5-minute break followed the recall phase to prepare for the recognition test, which included 64 questions (immediate recognition test). The same testing procedure was applied during the delayed test in week 1 and the delayed test in week 2.

Prior to starting the recognition test, the children were reminded to click on the image they thought was correct as fast as possible after hearing the question. At the beginning of test, the opening screen included a written instruction in English (which was also explained by the experimenter) to the children to listen to the question first then to choose the correct answer. The display included the "I'm ready!" button to click when participants fully understood the instruction. A blank screen appeared for 10 seconds followed by four images on the screen (Figure 4.3) with an auditory stimulus of the target word included in a question (i.e., /ʔajna al-sˤafiːħaʔ/ "where is the metal plate?"). The children were instructed to choose one of the four pictures by clicking on the image. The location of the images was randomized across the test.

The 64 testing trials presented four pictures on the screen to participants, organized into four distinct image combinations: (1) a target word, a semantic competitor, a phonological competitor, and an unrelated word (Figure 4.3); (2) a target word, a semantic competitor, and two unrelated distractors; (3) a target word, a phonological competitor, and two unrelated words; and (4) a target word and three unrelated distractors. This design was implemented to introduce variety in the visual and lexical contexts presented to participants, ensuring that the task engaged multiple aspects of word recognition and retrieval under different conditions. A blank screen appeared for 10 seconds to separate the test screens during the test. The testing session lasted up for around 30 minutes for each participant (i.e., excluding the 30 minutes to be trained and to learn the words).

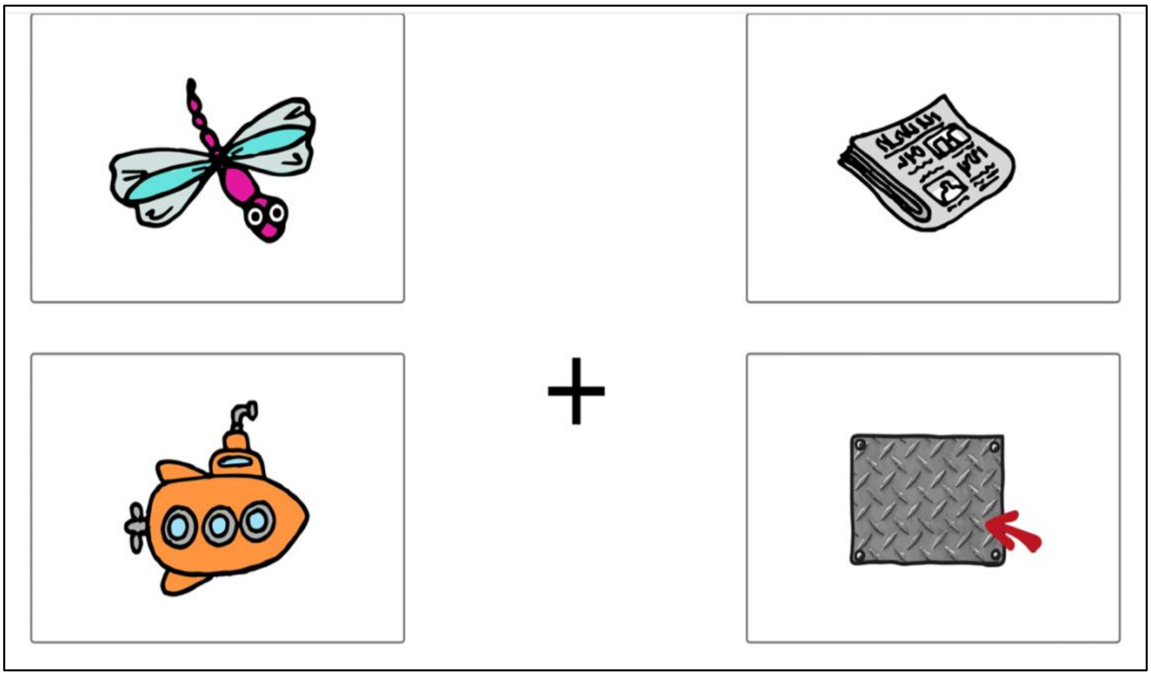


Figure 4.3: Example of a test trial demonstration as presented on Gorilla

Target: /sˤafiːħa/ "metal plate". Phonological competitor: / sˤaħi:fa/ "newspaper". Semantic competitor: / ɣawwaːsˤa/ "submarine". Unrelated word: / jaʕˈsuːb / "dragonfly".

4.6. Analysis:

Accuracy of answers (for recall) and accuracy and reaction time (for recognition tests) were the primary dependent variables in this study. The main effects of interest were whether items were repeated (produced) or observed silently (condition: listen vs. repeat), and the timing of the post-test (test: immediate, one-week delayed, and two-week delayed). Additionally, in the recognition test analysis, word type (target, phonological competitor, semantic competitor, unrelated) was considered to examine correct and incorrect responses (targets and competitors).

4.6.1. Coding

Children's responses to the recognition test were coded using Gorilla. Each test trial was evaluated for accuracy (1 = correct, 0 = incorrect). Recall tasks were manually coded in Excel, with items marked as recalled (= 1) or not recalled (= 0). Recall items were considered correct if they were retrieved from the MSA form (i.e., as learned). Since the children's L1 was English and MSA L2 was the main focus, any retrieval of an English word was coded as incorrect. For example, if the child recalled /ˈʃaːħinah/ (Shahina) correctly, they would score 1 (correct answer), but if the word was recalled as "truck" or "big car" in English, the score would be 0 (incorrect).

4.6.2. Statistical Analysis

Prior to analysis, data cleaning was done in R version 4.0.5. (R Core Team, 2021) after extracting the data from Gorilla. Data was organized to include subject, test time, reaction time, items and accuracy of answers. RTs were then log-transformed to reduce skewness. The analysis included 4492 data points across the three recognition tests.

To answer the research question, the analysis included two stages: (1) analysis of accuracy in recall; and (2) analysis of accuracy and reaction time in recognition. The analysis also included a sub-analysis of incorrect responses in the recognition test to examine the influence of word type (i.e., lexical competitors) to understand the mechanism underlying the production of newly learned words when target words are not chosen.

To analyse the accuracy data, a generalized linear mixed-effects model (GLMM) with a binomial distribution was used. Glmer, a type of regression model that can include random as well as fixed effects, was used to fit the model. The "bobyqa" optimizer was used for the maximum likelihood estimation. To analyse reaction time to correct answers, linear mixed-effects models were constructed and analyzed using the lme4 (version 1.1-26; Bates et al., 2015) and lmerTest (version 3.1-3; Kuznetsova et al., 2017) packages. The maximal random structure model included the interaction of all conditions with tests, and a fully random model was used whenever possible. If a model failed to converge, the random slopes were simplified. A multiple comparisons test was conducted using the package "emmeans" (Lenth, 2022) was implemented to identify significant differences among the groups.

4.6.3. Results

Overall recall and recognition results are shown in Table 4.3 which presents the mean and standard deviation (brackets) of accuracy (%) and reaction time (milliseconds) based on conditions (listen, repeat) in the three tests (immediate, one-week delayed, two-week delayed). Two-week delayed tests show higher accuracy rates for the "listen" condition in both recognition and recall and shorter RTs for both "listen" and "repeat" conditions compared to the immediate and one-week tests.

Table 4.3: Mean and SD (in parentheses) of correct answers (%) and RTs (in milliseconds) based on condition and test intervals

Recognition		immediate (n= 30)	one-week (n= 30)	two-week (n=30)
Accuracy Mean (SD)	Listen	64.0 (48.0)	73.3 (44.2)	81.8 (38.5)
	Repeat	65.1 (47.6)	70.1 (45.7)	72.8 (44.5)
RT Mean (SD)	Listen	3876.8 (1102.7)	3131.0 (1164.5)	2930.0 (968.2)
	Repeat	3909.8 (1132.0)	3164.1 (1173.6)	2916.3 (1001.4)
Recall		**immediate (n= 30)**	**one-week (n= 30)**	**two-week (n=30)**
Accuracy Mean (SD)	Listen	29.0 (45.4)	28.7 (45.2)	32.8 (47.0)
	Repeat	27.4 (44.6)	28.1.0 (44.9)	28.8 (45.3)

4.6.3.1. Accuracy in recall

A generalized linear mixed effect model (glmer) with binomial distribution was built to analyse accuracy across the three tests. The model included fixed effects of condition, treatment coded with listen as the baseline, and test time, treatment coded with immediate test as a baseline, and their interaction. This model included random intercepts for subject and item, and random slopes for test time. Model comparison was used to assess whether adding condition as a random slope would improve the fit. This showed that the addition of condition as a random slope did not improve the fit ($\chi2$ (15) = 15.3, p =.4). The best fitting model described in Table 4.4 shows no effect of condition, no effect of test, and no effect of interaction demonstrated in Figure 4.4 that shows the effect of condition and the results over test time.

Table 4.4: Mixed effects model output for the interaction between condition and test

Formula: glmer (correct ~ condition * test + (1 + condition*test | ID) + (1+ test | answer)

control = glmerControl(optimizer= "bobyqa"), family = binomial)

Fixed effects	**Estimate**	**SE**	**z-value**	**p-value**
(Intercept)	-1.06	0.18	-5.78	<.001***
Condition: repeat	-0.01	0.14	0.12	0.90
Test: one-week delayed	0.03	0.12	0.26	0.79
Test: two-week delayed	0.22	0.12	1.80	0.71
repeat*one-week delayed	0.03	0.15	0.22	0.81
repeat* two-week	-0.11	0.16	-0.69	0.48
Random effects variance (subject = 30, item = 64)				

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

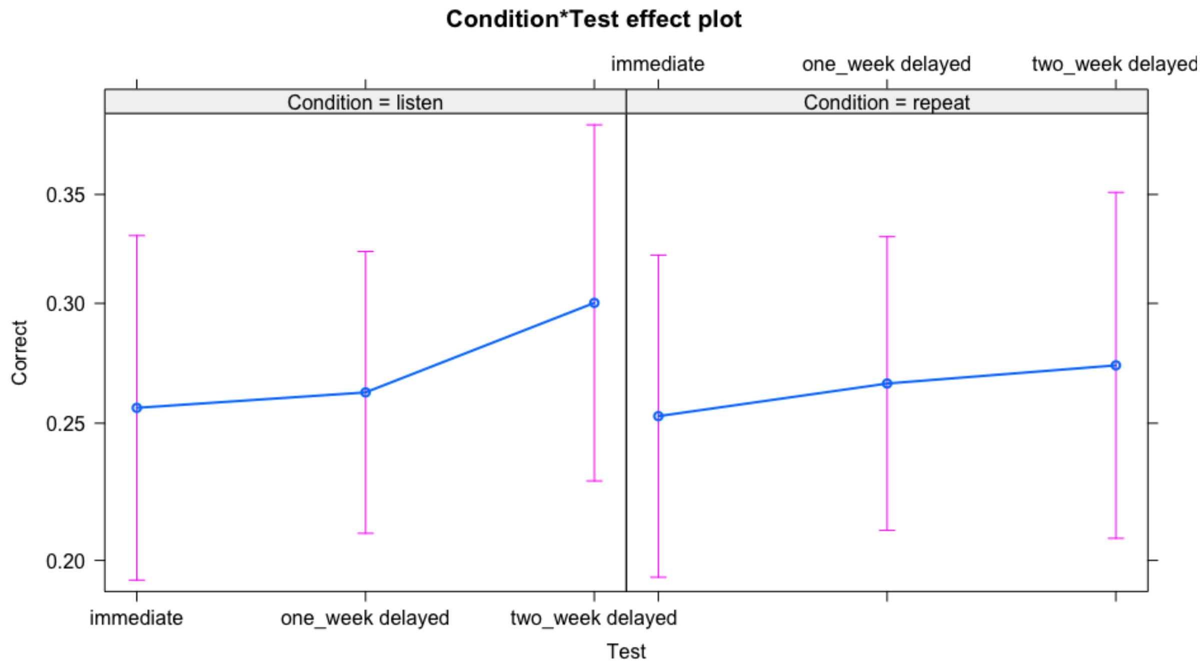


Figure 4.4: The interaction of condition and test time on accuracy of answers. y-axis is log-odds of answer accuracy

Despite the suggestion of a difference over time in Figure 4.4 (i.e., there seems to be an increase in performance during test time), the lack of an effect in Table 4.4 indicates that there are no significant differences across test times. Additionally, Table 4.5 shows that the condition had no effect on recall at any test point, as confirmed by a multiple comparisons test using the "emmeans" package (Lenth, 2022).

Table 4.5: Post-hoc tests with Bonferroni correction for accuracy based on the interactions between condition and test model in recall

Contrasts	**Estimate**	**SE**	**z-ratio**	**p-value**
Immediate test: listen - Immediate test: repeat	0.01	0.14	0.12	1.0000
One-week test: listen - One-week test: repeat	-0.01	0.13	-0.14	1.0000
Two-week test: listen - Two-week test: repeat	0.13	0.13	0.96	1.0000

* p < 0.05; **p < 0.01; ***p < .001

4.6.3.2. Accuracy in recognition

For recognition, the same recall analysis was used. Table 4.6 shows a significant effect of the one-week delayed test ($z = 0.29$, $p = 0.02$), a significant effect of the two-week delayed test ($z = 4.77$, $p < .001$), and a significant effect of the interaction between the repeat condition and the two-week delayed test ($z = -2.46$, $p = 0.01$). The analysis indicates that participants performed better in the delayed tests compared to the immediate test, as demonstrated in Figure 4.5. Overall, these results suggest that recognition performance improves over time, with participants showing better accuracy in the one-week and two-week delayed tests compared to the immediate test. Additionally, the interaction effect indicates that the repeat condition performs significantly worse in the two-week delayed test, highlighting a potential disadvantage of repetition over time.

Table 4.6: Mixed effects model output for the interaction between condition and test

Formula: glmer (correct ~ condition * test + (1 + condition*test | ID) + (1+ test | answer)

control = glmerControl(optimizer= "bobyqa"), family = binomial)

Fixed effects	**Estimate**	**SE**	**z-value**	**p-value**
(Intercept)	0.70	0.24	2.90	0.003**
Condition: repeat	0.06	0.14	0.44	0.66
Test: one-week delayed	0.70	0.30	0.29	0.02**
Test: two-week delayed	1.67	0.35	4.77	<.001***
repeat*one-week delayed	-0.45	0.28	-1.60	0.10
repeat* two-week	-1.00	0.40	-2.46	0.01 **
Random effects variance (subject = 30, item = 64)				

* $p < 0.05$; ** $p < 0.01$; *** $p < .001$

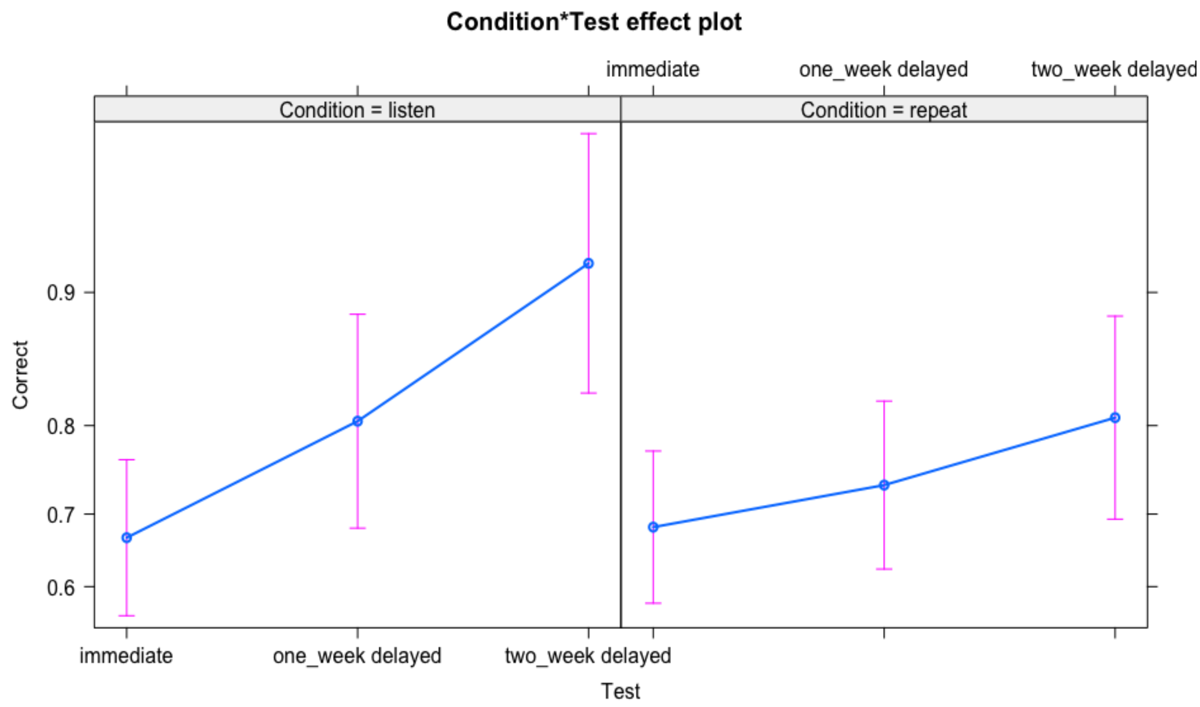


Figure 4.5: The interaction of condition and test time on accuracy of answers. y-axis is log-odds of answer accuracy

Figure 4.5 suggests that the retention of heard words (Condition: listen) may exhibit less decay over time, as indicated by the increasing trend. This pattern implies that the 'listen' condition potentially offers greater durability in memory retention compared to the 'repeat' condition at each time point. However, the p-values from the interactions between tests and conditions indicate that these trends are not statistically significant. Table 4.7 presents the results from a post-hoc analysis using the 'emmeans' package (Lenth, 2022), applied to the base model (condition*test) to compare the 'listen' and 'repeat' conditions across immediate, one-week, and two-week test intervals. Although the visual trend in Figure 4.5 appears to support a fundamentally different pattern, the post-hoc tests with Bonferroni correction do not confirm significant differences, except for a marginal trend at the two-week test point between 'listen' and 'repeat' ($z = 2.60$, $p = 0.13$).

Table 4.7: Post-hoc tests with Bonferroni correction for accuracy based on the interactions between condition and test model in recognition

Contrasts	**Estimate**	**SE**	**z-ratio**	**p-value**
Immediate test: listen - Immediate test: repeat	-0.06	0.14	-0.43	1.0000
One-week test: listen - One-week test: repeat	0.38	0.21	1.78	1.0000
Two-week test: listen - Two-week: repeat	0.93	0.36	2.60	0.13

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

4.6.3.3. Reaction time in recognition

A linear mixed-effects model (lmer) was fitted to evaluate the log-transformed reaction time to correct answers. The model included the interaction between condition and test time as fixed effects, random intercepts for subject and item, and by-subject random slopes for test. Adding condition as a random slope did not improve the model fit ($\chi^2(3) = 3.71$, $p = 0.29$).

Table 4.8 shows significant effects for the one-week and two-week delayed tests, with participants responding faster in these tests compared to the immediate test. Specifically, the one-week delayed test (t = -5.51, p < .001) and the two-week delayed test (t = -8.56, p < .001) both showed significantly shorter reaction times. Condition (repeat) and its interactions with test times were not significant, indicating no significant difference in reaction times between the repeat and listen conditions across test times. Figure 4.6 illustrates these findings, showing a clear trend of decreasing reaction times in the delayed tests. This suggests that participants became more efficient in responding to test items over time, possibly due to increased familiarity or improved retention of the material. Overall, the analysis indicates that children not only provided more accurate answers in the delayed tests (as shown in Table 4.6) but also responded more quickly. This enhancement in performance over time highlights the potential benefits of delayed testing in reinforcing memory.

Table 4.8: Mixed effects model output for reaction time (log-transformed)

Formula: lmer (RT_log ~ condition * test + (1 + Condition * Test | ID) + (1 + Test | Answer)

Fixed effects	Estimate	SE	t-value	p-value
(Intercept)	-121.7	0.39	- 311.8	<.001***
Condition: repeat	- 0.07	0.27	- 0.28	0.78
Test: one-week delayed test	- 3.74	0.67	- 5.51	<.001***
Test: two-week delayed test	- 4.34	0.50	- 8.56	<.001***
repeat* one-week delayed test	0.43	0.48	0.90	0.37
repeat* two-week delayed test	- 0.31	0.34	- 0.91	0.36
Random effects variance (subject = 30, item = 64)				

p-values are estimated using the *lmerTest* package in R (Kuznetsova et al. 2017). * $p < 0.05$; ** $p < 0.01$; *** $p < .001$

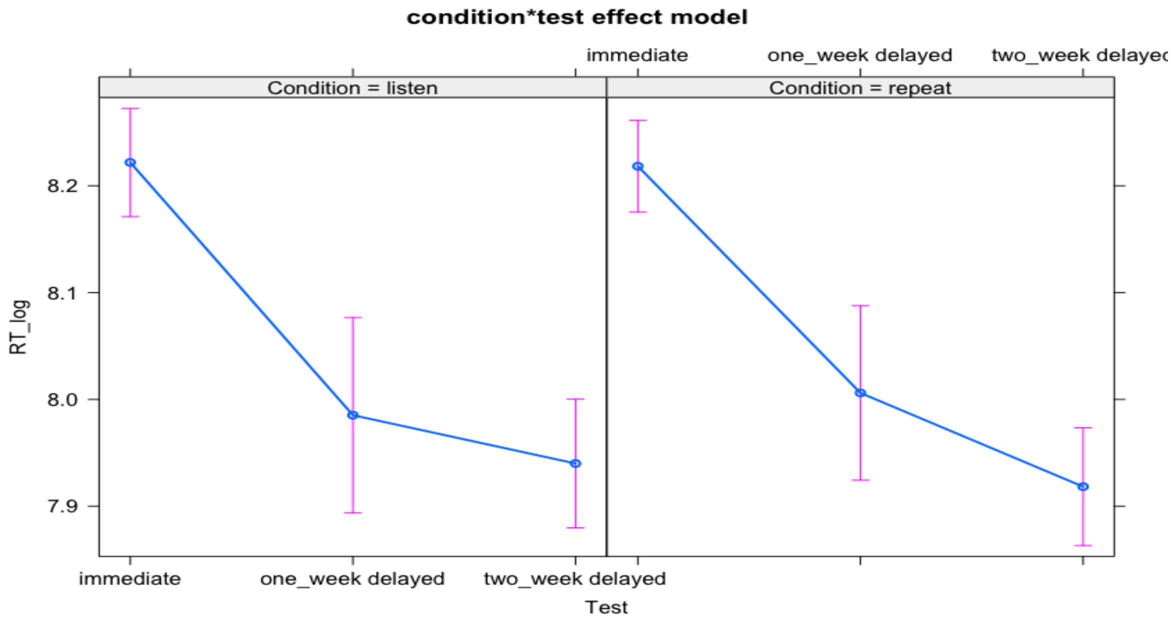


Figure 4.6: The interaction of condition and test time on RTs

A post-hoc test using the package “emmeans” (Lenth, 2022) was applied to the base model (condition*test) to compare the conditions (listen and repeat) across the test levels (immediate, one-week delayed, two-week delayed). As shown in Table 4.9, there are no significant differences between the listen and repeat conditions at any of the test times (immediate, one-week delayed, two-week delayed). This lack of difference suggests that both conditions performed similarly across all test intervals. The results indicate that the learning method—whether listening or repeating—did not significantly influence the reaction times across different time intervals.

Table 4.9: Post-hoc tests with Bonferroni correction for RTs based on the interactions between condition and test model

Contrasts	Estimate	SE	z-ratio	p-value
Immediate test: listen - Immediate test: repeat	0.00	0.01	0.19	1.0000
One-week test: listen - One-week test: repeat	-0.02	0.01	-1.04	1.0000
Two-week test: listen - Two-week: repeat	0.02	0.01	1.46	1.0000

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

4.7. Discussion

Most studies on the production effect have focused on adults, but there has been little research on the PE in children, particularly with child second language learners. Furthermore, childhood PE studies have systematically focused on short-term memory (i.e., immediate recall or recognition test) but have yet to examine the learning durability over different retention intervals. Therefore, the primary aim of experiment 2 was to investigate the PE in recall and recognition of L2 child learners of MSA using extended learning lists over three testing times. That is, experiment 2 aimed to examine the production effect in MSA word learning and memory retention among six-year-old bilingual children. It explored the extent to which producing words affects the recognition and recall of newly learned MSA vocabulary. The experiment also investigated the role of lexical competitors (words with similar semantic or phonological features to the target words) in recognition tests. Three recognition tests were used to examine the production effect hypothesized to impact word learning and memory (immediate, one-week delayed, and two-week delayed tests). Here, all children learned a list of 64 words divided into two conditions (i.e., listen and repeat).

The findings of this study indicate that, although there is no significant difference between learning under "listen" and "repeat" at each test point, there was a significant interaction effect, suggesting an overall difference in the pattern of learning. Specifically, words learned under the "listen" condition exhibit greater durability and an increase in accuracy over time, with a pronounced advantage becoming evident after a two-week delay. This pattern suggests either no production effect (PE) or some indication of a reversed PE.

4.7.1. How do different learning conditions, specifically listening versus repeating, influence memory retention and reaction times for newly learned words?

Results indicate that words learned under the "listen" condition may be more durable and exhibit less memory decay over time. This indicates that listening may provide a learning advantage that becomes more pronounced with time Comparisons between conditions at the two-week delay suggest that the listening condition gains a learning advantage over time. However, no significant differences were found between the "listen" and "repeat" conditions at other test times, indicating no difference in memory retention at immediate or one-week delays. This suggests that while the listening condition may show benefits specifically in the long term, both learning conditions are equally effective in the short term and at the one-week delay. There are a few possible explanations possibilities for the statistically significant PE deficit or null PE:

1. Incompatibility Between Perceptual Encoding and Activated Production Representation: The interplay between perceptual encoding and production representation is characterized by a competitive dynamic that can adversely affect cognitive performance. This notion is grounded in established research on cognitive load and working memory. For instance, Barrouillet et al. (2007) elaborate on how tasks that require sustained attentional resources can interfere with concurrent processes that also demand attention, reinforcing the concept of competition between encoding and production mechanisms.

2. Production-Induced Distraction During Critical Perceptual Processing: Empirical studies on working memory and cognitive distraction substantiate the phenomenon where production activities disrupt perceptual processing. Konstantinou et al. (2014) provide evidence that varying types of working memory demands can modulate distractor processing, either exacerbating or mitigating it, depending on the nature of the cognitive control mechanisms at play. This evidence supports the hypothesis that production activities may serve as a source of distraction, thereby disrupting the efficacy of perceptual encoding.

3. Cognitive Load Associated with Production Tasks: The detrimental effects of cognitive load on learning and memory processes are well-documented in the literature. Mierop et al. (2020) illustrates that cognitive load can impair information encoding into memory, even in cases where individual stimuli are initially encoded successfully. This observation aligns with the proposition that the cognitive demands inherent in verbal production could interfere with learning and encoding processes, potentially leading to diminished cognitive performance. The production effect (PE), which posits that producing words aloud during learning enhances memory, has been a topic of extensive research. However, recent findings suggest that this effect is not universally beneficial and can be influenced by various factors such as the type of sounds being learned, task complexity, cognitive load, and production timing. This discussion aims to provide a comprehensive discussion of these factors, integrating findings from recent studies to elucidate the nuanced nature of PE and its implications for memory and word learning.

The production effect (PE) is traditionally explained by the distinctiveness account, which argues that saying words aloud makes them more distinctive in memory, thereby enhancing recall (MacLeod et al., 2010; Jamieson et al., 2016). However, findings from the current study, along with those from Thorin et al. (2018), Zamuner et al. (2018), Baese-Berk and Michaud (2019), and López Assef et al. (2021), challenge this notion. These studies suggest that the PE may be weakened or even reversed under specific conditions, mainly when the produced words contain non-native or less frequent phonemes or task complexity increases. Experiment 2 in this thesis explores a unique linguistic context: children who have phonological awareness of Modern Standard Arabic (MSA) but are not native speakers. This situation parallels earlier research on the production and recall of non-native phonemes. Thorin et al. (2018) and Zamuner et al. (2018) found that producing words with non-native phonemes led to poorer recall than merely hearing them. Although the children in our study are familiar with MSA phonemes, the cognitive load of producing these sounds in a non-native language context may still be substantial. This observation supports the idea that the cognitive demands of processing less familiar sounds can reduce the distinctiveness benefit typically gained from production.

Baese-Berk and Michaud (2019) extend this observation to include phonologically complex or less common sounds within a language that individuals may be familiar with but do not use natively. This closely resembles the relationship our participants have with MSA. The cognitive demands of articulating MSA—familiar but not native—could diminish some of the mnemonic benefits typically associated with production. López Assef et al. (2021) further emphasized that the cognitive resources required to articulate such phonemes might offset the distinctiveness benefits. In our study, although the children are phonologically aware of MSA, producing these sounds in a controlled experimental setting requires significant cognitive resources, which may explain the observed attenuation of the production effect.

The distinctiveness account posits that produced words stand out and are therefore easier to remember. However, the current findings suggest that this advantage diminishes as task complexity increases. In this study, participants performed similarly on both heard and produced words, indicating that the cognitive load imposed by learning items in a mixed list format—and the resulting fatigue—can neutralize the distinctiveness advantage. Cognitive Load Theory (CLT) offers a framework to understand how working memory limitations affect learning and retention. CLT suggests that working memory has a limited capacity for processing information, which can be overwhelmed by high task complexity (Young et al., 2014). This is particularly relevant to the distinctiveness account, which argues that produced words are more memorable due to their uniqueness. However, when learners must process both heard and produced words within a mixed list, the cognitive load increases significantly due to the need to simultaneously manage and differentiate between the two tasks.

The intrinsic cognitive load of processing and remembering words and the extraneous load from switching between heard and produced words can exceed working memory capacity, leading to cognitive fatigue (Sweller, 2020; Jong, 2010). Cognitive fatigue, resulting from prolonged cognitive activity, impairs memory and learning. In this scenario, the effort required to manage the mixed list format may negate the distinctiveness benefits of produced words. As cognitive load increases, the distinctiveness advantage diminishes because the overburdened working memory cannot effectively process the unique features of produced words. Cognitive load is a critical factor in the effectiveness of the Production Effect. Producing words is a complex process that requires the integration of multiple cognitive functions, such as attention, phonological processing, and motor coordination. This complexity increases cognitive load, which can lead to cognitive fatigue and reduce the mnemonic benefits of production (Baddeley, 2012). In this study, participants had to learn items through both production and listening within a mixed list format. This likely exacerbated cognitive fatigue, thereby neutralizing the distinctiveness advantage of produced words. The cognitive load associated with producing words aloud, underscores the importance of considering the phonemic familiarity of learning materials. Reduced cognitive load with phonemically familiar material allows for more effective memory retention and word learning.

The researcher's observational notes tentatively noted that children waited between 2-3 seconds before producing the words after listening. This observation aligns with the findings of Kapnoula and Samuel (2022), who provided valuable insights into how brief delays between hearing and producing words can mitigate the negative impact of production. They found that introducing a two-second delay between hearing a new word and the requirement to produce it partially alleviated the detrimental effect. In contrast, a four-second delay eliminated this effect. This suggests that listeners need a brief moment to process and consolidate the new word immediately after hearing it. When production is required during this critical period, it disrupts this process.

Kapnoula and Samuel (2022) propose that this disruption could be due to an incompatibility between the developing perceptual encoding and the activated production representation, leading to active competition. Production might harm perceptual learning by distracting listeners from processing the perceptual input during this crucial moment. Their findings are supported by Baese-Berk and Samuel (2016; 2022), who demonstrated that requiring participants to produce something unrelated to the target sound (such as naming a printed letter instead of the heard syllable during the training trial) led to significantly poorer perceptual learning compared to having no production requirement, but better learning than producing the matching syllable. This supports both the distraction and incompatibility hypotheses, highlighting the complex interplay of factors affecting the production effect on memory and word learning.

It is crucial to highlight how recall and recognition results interacted under the two learning conditions. While recall performance remained broadly consistent across the different test times, showing no significant differences between the 'listen' and 'repeat' conditions, recognition performance demonstrated notable improvements over time, especially at the two-week delayed test. Specifically, the 'listen' condition exhibited a significant advantage in recognition accuracy in the two-week delay, although this benefit was not mirrored in the recall tasks. This distinction between recall and recognition could be attributed to the cognitive demands of each task. Recognition is typically considered less cognitively demanding than recall, allowing learners to rely on cues to identify the correct answer rather than actively retrieving information from memory (Eysenck & Eysenck, 1979). This reduced cognitive load may explain why participants performed better on recognition tasks over time, particularly for words learned through listening (Webb & Nation, 2017). In contrast, the higher cognitive demand for recall tasks may have contributed to the lack of significant differences between the learning conditions, as both conditions seemed to reach a cognitive threshold that limited improvements in recall performance (Karpicke et al., 2014). Thus, while listening may facilitate better recognition in the long term, its effect on recall appears minimal, suggesting that the type of memory task plays a crucial role in determining the effectiveness of each learning condition.

Together, these findings suggest that the interplay between cognitive load and task complexity significantly influences the effectiveness of different learning conditions. Balancing perceptual and production demands is crucial for optimizing memory retention and learning outcomes. Ultimately, the impact of learning conditions on memory retention and reaction times is multifaceted, indicating that both listening and repeating have unique advantages depending on the context.

4.7.2. To what extent can the production effect advantage in children be enhanced through delayed testing as opposed to immediate testing?

The results indicate that participants performed better in delayed tests, particularly in the two-week delayed test than in the immediate test, as evidenced by higher accuracy and shorter reaction times (RTs) for both the "listen" and "repeat" conditions. Moreover, the current study highlights the durability of learning, as demonstrated by better performance in the two-week delayed tests compared to the immediate and one-week delayed tests for both conditions. This suggests that despite a slight, not yet significant, increase in words learned under the "listen" condition, there is an overall trend towards improved learning over time. These findings suggest that consolidation over time enhances memory retention and retrieval efficiency, and the testing effect further reinforces this by improving long-term memory through repeated retrieval practice. Additionally, it implies that participants may have paid more attention to these words during the interim period, possibly due to increased familiarity, which improved learning and performance in the delayed tests. However, further evidence would be needed to confirm whether this increased attention was directly caused by the participants' familiarity with the words. According to Schmidt (1991), familiarity with a stimulus can increase the likelihood of attention being paid to it. Familiar items tend to be processed more efficiently and are more likely to be encoded into memory.

Yonelinas (2002) suggest that when participants are familiar with certain stimuli, such as words or images, they are more likely to perform better in subsequent tasks. This is because familiar stimuli require less cognitive effort, allowing for better focus and learning. Familiarity also plays a role in consolidation, which is critical for long-term memory. Familiar words are more likely to be retained and recalled during delayed tests because they have been encoded more strongly during the initial learning phase (Unsworth & Engle, 2007).These findings align with the theoretical framework posited by Nation (2017), who identifies two essential conditions for effective vocabulary acquisition: repetition and quality mental processing. Repetition ensures frequent encounters with vocabulary items, facilitating their transfer from short-term to long-term memory. Quality mental processing, which involves deep and meaningful word engagement, further enhances retention and recall. These principles are supported by the Levels of Processing hypothesis proposed by Craik and Lockhart (1972), which asserts that the depth of cognitive engagement significantly influences memory retention.

Nation (2017) categorises vocabulary learning into incidental and deliberate attention, which provides a valuable lens through which to interpret the current findings. Incidental attention involves encountering words during meaning-focused activities such as reading or listening, while deliberate attention involves intentionally focusing on vocabulary as distinct objects of study. Nation (2017) argues that deliberate attention, which typically involves deeper processing, is more efficient and effective for vocabulary learning. Nation (2001, 2017) also outlines four levels of processing for vocabulary learning: noticing, retrieval, varied meetings and use, and elaboration. These levels represent increasing cognitive engagement and are critical for effective vocabulary retention.

Further empirical evidence comes from a study by Icht and Mama (2022), which tested memory performance immediately, after one week, and after two weeks. The results indicated that while participants showed higher accuracy in immediate tests, long-term retention, measured by the production size, was significantly higher in the two-week delayed test. This underscores the importance of delayed testing in enhancing vocabulary retention, as it allows for better consolidation of learned material over time. The current study's use of generalized linear mixed models (GLMM) to analyze accuracy data mitigates the potential for type I errors associated with ANOVA in binomial data (Icht & Mama, 2022). The superior performance observed in delayed tests, particularly after two weeks, highlights the efficacy of delayed testing in vocabulary retention. The faster response times in children during delayed tests suggest a shift from conscious to automatic word retrieval, indicative of profound and practical learning. These findings resonate with the testing effect, extensively studied by Roediger and Butler (2011), which refers to improving memory retention when information is retrieved through testing rather than merely restudied. The testing effect operates through several mechanisms, including enhanced memory traces, active engagement, increased attention to retrieval cues, and feedback and correction (Roediger & Butler, 2011). Each retrieval act strengthens memory traces, making information more accessible for future recall. To further substantiate this argument, longitudinal studies have shown increased accuracy over time. For instance, Varela (2020) conducted a study on spaced repetition in a high school language course, revealing that students retained vocabulary significantly better over time when assessed at intervals of 30, 60, and 70 days post-treatment. These longitudinal findings collectively illustrate how structured repetition and delayed testing can effectively enhance vocabulary retention accuracy over time. Taken together, the findings from this study underscore the significant enhancement in vocabulary retention through delayed production and repeated testing, aligning with memory consolidation and testing effect theories. It emphasises the importance of delayed testing in promoting long-term retention and suggests incorporating both immediate and delayed retrieval strategies in educational practices to maximise vocabulary acquisition in children. Moreover, the study highlights delayed testing's critical role in understanding L2 Modern Standard Arabic (MSA) learning mechanisms. Focusing on immediate and delayed retrieval offers more profound insights into vocabulary acquisition and retention. These findings advance the understanding of effective language learning strategies, advocating for varied retrieval practices to enhance long-term retention in L2 MSA learners. Future research should explore these mechanisms to optimize vocabulary learning for diverse populations.

4.8. Conclusion and Limitations

The findings highlight the complexity of the Production Effect (PE), revealing that either no PE or a reverse PE is present, particularly in the context of listening tasks evaluated after one and two weeks. This study underscores the importance of delayed testing for enhanced vocabulary retention, aligning with theories on memory consolidation and the testing effect. Consistent results across different word types and learning conditions suggest that balancing perceptual and production demands is crucial for optimizing memory retention. Furthermore, the analysis of phonological competition indicates that phonological encoding plays a critical role in word recognition and recall, exerting a more pronounced influence than semantic competition. This insight is essential for understanding how PE influences Modern Standard Arabic (MSA) vocabulary acquisition in children, highlighting the potential variability in learning outcomes based on different learning conditions. However, several limitations must be acknowledged. The exclusive reliance on a free recall task restricts exploring alternative recall methods, such as pictorial recall, which could provide deeper insights into memory retrieval processes. Additionally, variability in the structure of testing trials may reduce comparability across trials and introduce unintended variability in participant responses; thus, a more standardized design is warranted in future studies. Moreover, the absence of two essential norming studies weakens the experimental design: one with native Arabic speakers to assess phonological and semantic associations related to target words and another with children to ensure familiarity with these words. Addressing these limitations could lead to a more comprehensive understanding of memory retrieval and learning strategies, ultimately enhancing vocabulary acquisition outcomes.

Chapter 5. Experiment 2

The pilot study (Chapter 3) assessed the effectiveness of the production effect (PE) on Modern Standard Arabic (MSA) learning, focusing on age, language background, and the learning conditions of "listen" or "repeat." The results showed that age influenced children's performance more than the learning method or language background. Experiment 1 (Chapter 4) built on the pilot study by addressing its limitations, such as the small sample size and simple word list. Using a list of 64 MSA words and controlling for age and language background, Experiment 1 aimed to examine PE and its effects on memory retention thoroughly. Participants were tested immediately, one week, and two weeks after learning under "listen" or "repeat" conditions. Results indicated better performance in delayed tests, with faster response times and higher recognition accuracy. Although there were no significant differences in recall accuracy across conditions or test times, a significant interaction suggested an attenuated reversed effect. Various factors, such as the word list, test delay, and distractor type, appeared to reduce the PE. This chapter also examines different production types (e.g., writing) and typical PE conditions (listening and repeating) in MSA learning.

5.1. Introduction: Writing as a form of production

As discussed in Chapter 2, the Production Effect (PE) refers to an enhanced memory for words produced versus those read silently. The literature presents mixed evidence: positive effects are observed in adults and children aged 5-10, while reversed effects occur in younger children (2-4 years) and adults learning non-native words. Additionally, some studies report no-effect cases where participants' performance is consistent across conditions. Studies investigating PE in adults have used various forms of production, such as mouthing, whispering, spelling, typing, and writing (MacLeod et al., 2010; Forrin et al., 2012; Quinlan & Taylor, 2013; Bodner et al., 2016; Icht & Mama, 2016; Zormpa et al., 2018). Child studies have used similar conditions: look only, look and listen, look and say, listen only, say aloud, read aloud, and read silently (Icht & Mama, 2015; Zamuner et al., 2017; Pritchard et al., 2019; López Assef et al., 2021). Despite extensive research into the effect overall, studies have yet to investigate writing as a type of production in children.

The writing process involves a complex interplay of encoding semantic, phonological, and orthographic information (Hayes et al., 2000). Writing a word requires lexical processing, which includes forming auditory and visual associations between a word and its representation. This process also involves developing a mental representation of the sound (phonological representation), enabling access to the word's orthographic representation during sub-lexical processing (Costa et al., 2011; Pacheo & Huertas, 2022). In this perspective, words consist of a series of letters which, in turn, have different representations calls "allographs". Allographs represent different visual manifestations of the same letter, aiding tasks like reading and writing (Schubert & McCloskey, 2013). These representations are influenced by typical fonts, which play a significant role in shaping them (Rothlein & Rapp, 2017). Each allograph is paired with a graphic motor pattern stored in the long-term memory (Van Galen, 1991; Schmidt, 2013).

When an individual writes, the movement pattern for each allograph is retrieved from long-term motor memory, initiating the motor program necessary for writing (Bonney, 1992; Guan et al., 2011; Lai & Leung, 2012). That is, when writing a letter, individuals first retrieve the specific letter representation, known as a grapheme, from memory. This process involves mapping phonemes (sound units) to graphemes (written symbols). Graphemes are the smallest units of writing that represent phonemes in alphabetic systems, forming the crucial link between spoken and written language (Vinci-Booher et al., 2019). Subsequently, they retrieve the corresponding shape of the letter, referred to as the allograph, to accurately produce the letter (Maurer, 2023). Allographs are crucial in word production as they specify the visual form of letters within words (Lambert & Quémart, 2018). These abstract letter identities represented by allographs are independent of visual variations, such as uppercase or lowercase forms, emphasizing their importance in reading and writing tasks (Kinoshita et al., 2019). The processes involved in written word production encompass cognitive functions like orthographic memory and motor actions required for writing. Allographic processes follow central stage processes and involve specifying how letter series will be produced, including the individual's unique way of writing, and executing the necessary motor sequences (Planton et al., 2013). The coordination of central cognitive processes with peripheral motor actions is crucial for generating written words effectively (Purcell et al., 2011).

Writing involves the intricate mapping of phonological (sound) to graphemic (written symbol) information in English or Arabic. This phonetic nature is essential for converting spoken language into written form. While both languages share phonetic elements, they have distinct characteristics that shape their writing systems and learning processes. The participants in this study were all exposed to English from birth but had phonological awareness of Modern Standard Arabic. This dual language exposure influences their cognitive processing during writing tasks. Understanding these specificities is crucial for analyzing how children map phonological to graphemic information across different languages and writing systems.

The English alphabet is a phonetic writing system designed to represent spoken language sounds rather than word meanings. It consists of 26 letters, which combine to form graphemes that map onto specific phonemes, allowing for a wide array of word representations through combinations of letters that mirror their pronunciation (Rayner et al., 2001). Despite irregularities in spelling, the primary function of the English alphabet is to capture speech sounds, as seen in the use of different letters such as 'c', 'k', and 'q' to represent the /k/ sound (Treiman, 1993). Learning to read English involves teaching the sounds associated with each letter and blending them to form words, highlighting their phonetic nature (Ehri, 2005).

In contrast, the Arabic writing system, while also phonetic, has unique characteristics. Arabic is built around consonantal root structures, typically comprising three consonants that convey core meanings, with vowels and affixes modifying these roots to express different grammatical forms (Ryding, 2005). The Arabic script includes 28 letters representing consonant sounds. While short vowels are not usually written in standard texts, they can be indicated with diacritical marks in educational or religious contexts (Versteegh, 2014). For example, the word "عِدْ"/ʕid/" promise!" can be read as "عِيد"/ʕiːd/"holiday" if the long vowel "i: "is used instead of the short vowel "i". This use of diacritical marks makes Arabic partially phonetic, especially when precise pronunciation is necessary, though, in everyday writing, readers infer vowels from context and linguistic knowledge (Brockett, 1985). Within Modern Standard Arabic (MSA), a letter's representation differs according to its spatial position in a word. For instance, the letter YaA "ي" can be presented as "يـ" at the beginning of a word or "ـي" at the end of a word (Boudelaa et al., 2019).

Writing Arabic presents unique challenges due to its cursive nature, where letters are connected, making segmentation complex (AbdAllah & Viriri, 2020). The visual complexity of Arabic letters can impede the processing of orthography, affecting reading and writing tasks (Ibrahim et al., 2013). Additionally, the informal, colloquial form of Arabic used in informal contexts can present challenges for spell-checking and analysis (Al-Jarf, 2023).

Learning an alphabetic writing system, such as English or Arabic, induces the representation of phonemic structure in the phonological lexicon (Ziegler & Goswami, 2005). The Double-Route process explains the encoding of words, involving phonological recoding (translating graphemes into phonemes) and logographic reading (recognizing whole words that do not follow regular grapheme-phoneme correspondences) (Kerek & Niemi, 2009), contributing to the selection of appropriate orthographic representations (Cardoso et al., 2012). Phonological information guides spoken production articulation, while orthographic information directs output in written production (Zhang & Wang, 2014). For instance, writing "cat" (Figure 5.1) involves activating phonological, semantic, and orthographic information, recalling the letters from long-term memory, and applying motor skills to write the word (Costa et al., 2011).

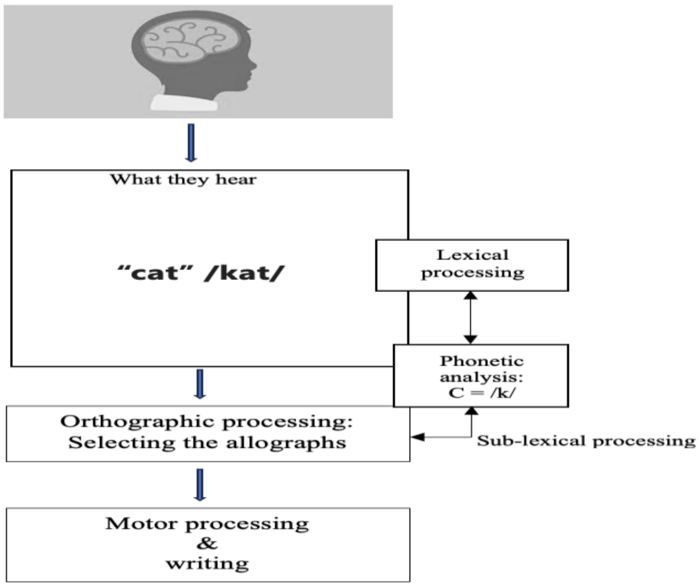


Figure 5.1: The lexical and sub-lexical processing of a word

The interaction between phonological and orthographic processing in writing and vocabulary learning can be framed through the lens of Dual Coding Theory (Paivio, 1971). This theory posits that human cognition involves two distinct but interconnected systems: one for verbal information, such as spoken or written words, and one for non-verbal or visual information, such as images or written text. In the context of writing and vocabulary acquisition, this theory suggests that engaging both the phonological (verbal) and orthographic (visual) systems simultaneously can create more robust mental representations of the target vocabulary (Clark & Paivio, 1991), thereby enhancing recall and recognition.

When learners encounter new vocabulary, particularly in written form, they rely on auditory input (phonology) and visual input (orthography) to reinforce their understanding and retention of the words. Writing activities, such as seeing the written form of a word and repeating it aloud, combine these dual systems by engaging the learner in phonological and orthographic processing. As Dual Coding Theory posits, this integrated engagement strengthens the encoding of new vocabulary, making it easier for learners to retrieve the words later during recall tasks (MacLeod et al., 2010; MacLeod & Bodner, 2017). The theory, therefore, highlights the value of incorporating both writing and verbal repetition into vocabulary learning strategies.

The relevance of Dual Coding Theory to the Production Effect (PE) becomes apparent when considering how repeating a word aloud after writing or hearing it further enhances phonological processing. This active verbal repetition reinforces memory by engaging both auditory and articulatory systems, which, combined with the visual system through writing, leads to stronger word retention (Bailey et al., 2021; Cyr et al., 2021). In this context, writing while vocalizing or repeating a word activates both cognitive systems, creating a more prosperous and more integrated memory trace.

Research supports the idea that engaging verbal and visual systems in writing and vocabulary learning improves retention and the application of newly learned words. Clark and Paivio (1991) argue that dual coding, processing both the spoken and written forms of words, allows for more robust encoding in memory. For example, Mayer and Sims (1994) found that learners who engage with verbal and visual information, such as written words alongside auditory repetition, perform better on tasks that measure retention and knowledge transfer. This suggests that writing activities involving both seeing and repeating words could enhance vocabulary retention.

In research on the Production Effect, where learners listen to and repeat words, the dual activation of phonological and orthographic systems helps explain the potential benefits of this learning condition. Though the PE emphasizes verbal repetition, the combined processing of phonological and orthographic information through writing may underlie its effectiveness in boosting word recall. Learners create stronger associative links between the sound of the word, its written form, and its meaning, allowing for more efficient retrieval in both immediate and delayed recall tests (Moreno & Mayer, 1999). This integration of writing, repetition, and verbal engagement provides a comprehensive strategy for enhancing vocabulary learning through PE.

Levie and Lentz (1982) conducted a comprehensive review of 55 experiments that compared learning outcomes from illustrated texts with those without illustrations. Their findings demonstrated that including text-relevant illustrations significantly improved comprehension and retention, with groups exposed to illustrated materials performing, on average, 36% better on comprehension tests. Illustrations were shown to guide learners' attention to key content, facilitating better understanding and memory retention. Furthermore, the presence of visuals enhanced students' enjoyment of the material and positively influenced their attitudes toward learning. Importantly, illustrations provided substantial support for learners with lower reading skills, helping them overcome challenges in understanding the text. The benefits were especially pronounced for comprehension questions directly related to the illustrated content, reinforcing the role of visuals in improving educational outcomes. These results align with the Dual Coding Theory (Paivio, 1971), which posits that learning is enhanced when verbal (text) and non-verbal (visual) systems are activated simultaneously. When both systems are engaged, learners can create richer mental representations, improving recall and deeper comprehension (Clark & Paivio, 1991). Integrating verbal and visual information supports the idea that incorporating relevant illustrations into educational materials can significantly enhance learning outcomes (Mayer, 2009).

Studies have found that writing enhances the distinctiveness of a word relative to other forms of production (e.g., singing, reading aloud, etc) (Levelt, 1999, 2001; Fairs et al., 2022). As discussed in section 2.3, Research on the production effect demonstrates that producing a word through speaking, writing, or typing generally leads to better memory retention than reading it silently, with writing showing some advantages over other forms of production (MacLeod et al., 2010; Ozubko & MacLeod, 2010). The additional effort and time required for handwriting contribute to making words more distinctive in memory. As explained earlier, writing helps create a stronger visual representation of the word's form, aiding in later recognition and spelling. Thus, writing offers unique advantages in making words more distinctive and memorable compared to other forms of production, likely due to its multi-sensory nature and the deeper processing it often requires. In this respect, written words involve visual, motor, and auditory processes (reading, articulating, hearing), whereas silent reading is purely visual. It is argued that improved memory is caused by additional information encoding and that the PE may be determined by the number of encoding processes that are involved in learning, emphasizing that at test locations, participants search for unique encoding information (Mama & Icht, 2016; Icht & Mama, 2021). Essentially, participants use the distinctive features of their encoding processes, such as producing words aloud, to aid in memory retrieval during these testing moments.

On the other hand, when comparing mouthing, writing, whispering, and silent reading, Forrin et al. (2012) found that speech production significantly improved memory, while writing did not show the same level of enhancement compared to speech. However, research on the Production Effect (PE) has yielded inconsistent findings, with some studies reporting strong effects while others show more elusive results. The distinctiveness of speech is often cited as a reason for these differences, as producing words aloud engages multiple sensory and motor systems, creating more distinctive memory traces than other production forms (MacLeod et al., 2010). Saying the word aloud provides auditory feedback and engages phonological knowledge, which can aid recall. While writing involves motor processes and creates a visual representation that can enhance memory compared to silent reading, it lacks the same level of distinctiveness and auditory feedback that speech offers.

Additionally, the distinctiveness account suggests that learning with visual aids, such as pictures, can also enhance recall, as visual stimuli offer unique and distinctive features. This is consistent with the dual coding theory, which posits that linking words to images strengthens memory by engaging verbal and image systems (Ensor et al., 2019; Roberts et al., 2023). Nevertheless, the inconsistent results across different modalities underscore the complexity of memory processes and suggest that the PE may not always have the same impact depending on the Using sentence writing as a strategy for enhancing vocabulary retention is based on the assumption that deeper semantic processing and contextual integration lead to better learning outcomes (Craik & Lockhart, 1972; Laufer & Hulstijn, 2001). However, Barcroft (2002) challenged this view, suggesting that sentence writing may hinder rather than facilitate L2 vocabulary acquisition under certain conditions.

Using sentence writing as a strategy for enhancing vocabulary retention is based on the assumption that deeper semantic processing and contextual integration lead to better learning outcomes (Craik & Lockhart, 1972; Laufer & Hulstijn, 2001). However, Barcroft (2002) challenged this view, suggesting that sentence writing may hinder rather than facilitate L2 vocabulary acquisition under certain conditions.

Barcroft's (2002) Type of Processing-Resource Allocation (TOPRA) model offers a framework for understanding how different types of processing during vocabulary learning compete for limited cognitive resources. According to this model, learners must distribute their cognitive resources across various processing types, such as semantic, structural, and form-based processing. The TOPRA model posits that focusing on one type of processing may reduce resources available for others, potentially creating trade-offs that affect learning outcomes.

Barcroft's (2004) seminal study marked a significant turning point in understanding the relationship between sentence writing and L2 vocabulary acquisition. In this research, Barcroft compared the effects of writing new words in sentences with word-picture repetition learning for L2 Spanish learners. The study employed a within-subjects design, with participants learning new vocabulary under different conditions across two experiments. The results were striking: participants consistently demonstrated poorer performance on measures of word form learning when they had engaged in sentence writing than when they had simply been exposed to word-picture pairs.

Specifically, in Experiment 1, participants learned 24 new Spanish words under three conditions: writing sentences, writing the words as many times as possible, or no writing. The results showed that both writing conditions led to significantly lower scores on immediate and delayed productive vocabulary tests compared to the no-writing condition. Experiment 2 replicated these findings with a different set of words and presentation intervals, further strengthening the evidence for the detrimental effects of sentence writing on initial vocabulary acquisition.

These findings were further corroborated and extended in Barcroft's (2006) study, which examined the effects of "forced output" during vocabulary learning. This research reinforced the negative impact of sentence writing and demonstrated that even less demanding forms of output, such as copying target words, could interfere with lexical acquisition. The study employed a between-subjects design with 114 L2 Spanish learners, comparing three learning conditions: word writing, sentence writing, and no writing. Both writing conditions resulted in significantly lower scores on productive vocabulary tests than the no-writing condition, with sentence writing showing the most pronounced negative effect.

Barcroft (2007) continued to explore this phenomenon, investigating the effects of word and fragment writing on L2 vocabulary learning. This study introduced a novel condition where participants wrote fragments of target words, providing insight into how different writing tasks might impact vocabulary acquisition. The results aligned with previous findings, showing that full-word and fragment writing led to poorer vocabulary learning outcomes than a no-writing control condition.

While Barcroft's Research primarily focused on L2 Spanish learners, subsequent studies have sought to validate these findings across different language pairs. For instance, Wong and Pyun (2012) examined the effects of sentence writing on L2 French and Korean lexical retention. Their study compared two word-learning conditions: writing target words in sentences and learning words in isolation. The results corroborated Barcroft's findings, demonstrating that sentence writing can have detrimental effects on L2 vocabulary retention, particularly for languages with writing systems that differ from learners' L1.

Several mechanisms have been proposed to account for the negative effects of sentence writing on vocabulary acquisition. One explanation is related to *cognitive resource allocation*, as described in Barcroft's TOPRA model. Sentence writing requires learners to distribute significant cognitive resources toward managing grammatical structures and understanding semantic relationships, which can detract attention from processing target word forms (Barcroft, 2004, 2006).

Another explanation is the concept of *output interference*. Producing written output may disrupt the initial encoding of new vocabulary, an effect referred to as the "output interference effect" (Barcroft, 2007; Rohrer et al., 2010). This interference can weaken the ability to internalize new word forms during the early stages of learning.

Semantic elaboration also plays a role. While elaboration on meaning can enhance the retention of already known words, it may inhibit the acquisition of new word forms when learners must focus on deeper meaning at the expense of form-based processing. This redirecting cognitive resources away from form processing can hinder early vocabulary learning efforts (Barcroft, 2004; Leow, 2015).

Finally, the task complexity involved in sentence writing, which demands lexical retrieval, syntactic processing, and semantic integration, may overwhelm learners' working memory capacities. This is especially true for learners with lower proficiency, where the cognitive load required to process multiple elements simultaneously can be detrimental to effective vocabulary acquisition (Robinson, 2001; Skehan, 2009).

According to Barcroft (2004; 2006), educators may need to reconsider the timing and implementation of sentence writing tasks in vocabulary learning curricula. Introducing sentence writing activities may be more effective after learners are exposed to new word forms through other methods, such as word-picture associations or contextualized input. Furthermore, these studies suggest that a more nuanced approach to vocabulary instruction is necessary, one that considers the cognitive demands of different learning tasks and their potential impact on lexical acquisition. For instance, instructors might consider employing a progression of tasks that gradually increase in complexity, beginning with form-focused activities and moving towards more semantically rich exercises as learners become more familiar with the target vocabulary (Nation, 2013; Schmitt, 2008).

These findings provide significant evidence regarding the potential drawbacks of sentence writing in second language (L2) vocabulary acquisition, yet several crucial areas remain underexplored, warranting further investigation. One such area is the long-term effects of sentence writing on vocabulary retention and usage. Existing studies predominantly focus on immediate or short-term retention outcomes, leaving a gap in our understanding of how sentence writing influences long-term vocabulary consolidation. Schmitt (2010) suggests that further research is needed to determine whether the benefits or limitations of sentence writing persist over extended periods and how they affect learners' ability to retain and use vocabulary effectively over time.

Another important aspect involves the role of learner proficiency in the effectiveness of sentence writing tasks. Studies have yet to fully consider how learners at varying proficiency levels respond to sentence writing in vocabulary acquisition. Investigating whether learners of lower versus higher proficiency benefit differently from such tasks could provide valuable insights, enabling more targeted and efficient instructional strategies. Laufer and Hulstijn (2001) emphasize the importance of accounting for proficiency levels when evaluating vocabulary learning interventions, highlighting a gap that future studies should address.

The exploration of task variations within sentence writing and other output-based activities is also underdeveloped. Different task types, such as word fragment completion, sentence completion, and guided writing, may affect vocabulary learning differently. Comparative studies examining these tasks can help identify which designs are most conducive to successful vocabulary retention and usage. Barcroft (2007) and Webb (2005) advocate for such investigations, as they could reveal optimal conditions for vocabulary acquisition.

Moreover, individual differences in learners play a pivotal role in approaching sentence writing tasks. Factors such as learning styles, working memory capacity, and other cognitive variables can mediate the effectiveness of these activities. Understanding how these individual differences influence vocabulary learning can inform more personalized and adaptive approaches to instruction. Skehan (2012) highlights the importance of accommodating cognitive diversity in language learning, suggesting that research should focus on how sentence writing tasks can be tailored to meet the diverse needs of learners.

Finally, the potential of multimodal learning in conjunction with sentence writing remains an area ripe for exploration. The interaction between written tasks and other modes of vocabulary acquisition, such as oral production or digital tools, could reveal more integrated and effective learning strategies. As Chun et al. (2016) discuss, Multimodal approaches may offer learners richer and more varied opportunities to engage with new vocabulary, enhancing retention and application in real-world contexts. Further investigation is needed to explore how combining sentence writing with other modalities can optimize vocabulary learning outcomes. Together, these areas represent critical directions for future research that could deepen our understanding of sentence writing's role in L2 vocabulary acquisition and lead to more effective and learner-centered pedagogical approaches.

In summary, research on the effects of sentence writing on L2 vocabulary acquisition, particularly in the work of Barcroft and subsequent studies, challenges the traditionally assumed benefits of this practice. Evidence suggests that sentence writing may not consistently enhance vocabulary learning, particularly during the early stages. This finding underscores the need for a deeper understanding of the cognitive demands imposed by different learning tasks, such as listening, repeating, and writing, and how these tasks impact vocabulary acquisition. Given the complexity of output and semantic elaboration, it is crucial to investigate how various production methods affect long-term retention, especially for children learning Modern Standard Arabic. This study aims to expand our understanding of writing as a production method in enhancing vocabulary retention, thereby refining vocabulary acquisition strategies and contributing to more effective L2 learning mechanisms.

The current study aimed to extend the findings of Study 1 by employing the same testing method with older children. Furthermore, this study introduced a novel component to children's Production Effect (PE) research: adding a writing task. Based on the outcomes of Study 1, it was hypothesized that the PE would be null due to the cognitive demands of the task, leading to comparable performance across all conditions. However, retention was expected to improve over time. Study 1 indicated no significant overall differences in recall accuracy between the "listen" and "repeat" conditions. However, participants exhibited enhanced performance over time in both conditions, particularly in the two-week delayed test. This pattern suggests that while the PE may not be immediately apparent, memory consolidation improves over time, as reflected by the interaction effect observed in the delayed tests. Also, the current study investigates the effect of writing as a form of production on memory retention. As explained earlier, writing enhances memory through additional encoding processes, such as orthographic and motoric encoding, making words more distinctive and memorable. Therefore, it is predicted that children will show a significant production effect when learning new words under the writing condition, with better recall and recognition performance than listening and repeating, especially in the two-week delayed test. This prediction is based on the depth of processing and multi-sensory engagement in writing.

This study employs the same procedure as in Study 1, presenting unknown Modern Standard Arabic (MSA) words in an intermixed design, tested with lexical competitors, to investigate the production effect (PE), focusing on writing as a production method. The study targets eight-year-old children who have developed the ability to write complex sentences in their first language (L1). They can apply these writing skills in their second language (L2) by analyzing phonemes and producing corresponding orthographic forms. In addition to re-examining the typical PE conditions of listening and repeating, this study introduces writing as an additional production type, expanding the scope of PE research in this age group.

Compared to the six-year-olds in Study 1, the predictions for this older group include several vital outcomes: (1) Enhanced overall performance, as the eight-year-olds are expected to demonstrate higher accuracy rates across all conditions due to their advanced cognitive development and greater language experience. (2) Superior performance in the writing condition, leveraging their more developed writing abilities, which may result in a more pronounced PE for writing compared to listening and repeating.

Based on these considerations, it is anticipated that eight-year-olds will exhibit a significant production effect in the writing condition, akin to the "listen + write" condition observed in adults by Icht and Mama (2016). However, if the PE shows a reversal or null effect, it may suggest that the cognitive load associated with writing, combined with the phonological demands of the task, increases phonological competition during testing, though this effect is expected to be less pronounced than in younger children. Finally, while prior studies that examined PE in childhood have focused on immediate tests (i.e., direct and short learning effects on memory), this study, building on the methodology used in study 2, examines the durability of learning of L2 children across three testing periods (immediate, one-week, and two-week). Study 2 has already demonstrated the benefits of testing at these intervals, showing that participants' performance improves over longer delays, particularly in the two-week delayed test. By including writing as an additional form of production, the current study aims to investigate further whether this method enhances long-term retention more effectively than the "listen" and "repeat" conditions used previously.

5.2. Methods and Design

This experiment examines the production effect (PE) in the acquisition of unfamiliar Modern Standard Arabic (MSA) vocabulary by 8-year-old children across three conditions: listening, repeating aloud, and writing. The study seeks to explore the differential impact of these modes of self-production on vocabulary learning outcomes, explicitly comparing active engagement through repetition and writing to passive listening. The primary research question addressed in this study is: Does adding a writing condition as a mode of production lead to greater vocabulary retention in children compared to listening and repeating? Using a within-subjects design, the experiment ensures that each word is presented twice across all conditions: twice via auditory input in the listening condition, once via auditory input, and once through self-production in the repeating and writing conditions. This approach provides equal exposure across conditions, allowing for a direct comparison of the effects of different production methods on vocabulary retention.

The study included three phases: training, learning, and testing. Children were first trained on nine unfamiliar words to ensure compliance with study conditions and testing procedure.

Words were presented under three conditions: Listen (children see the image and hear the associated word twice); Repeat (children see the image, hear the associated word once, and produce the word once); and Write (children see the image, hear the associated word once, and write it on the paper once). After being briefed on the conditions, the participants took a short break and then completed a free recall task, followed by a recognition test, as discussed in section 5.4. The participants were presented with the learning items via a PowerPoint presentation, then tested online using Gorilla Experiment Builder (https://app.gorilla.sc/admin/home), a web-based tool for creating and deploying experiments. However, the children participated in the experiment in person, while being supervised by the experimenter at Amanah Muath Trust, an Islamic and Arabic school located in Birmingham, UK.

5.3. Participants:

Participants were 30 students (10 males, 20 females, aged 8;0- 8;9 years old, M=8;5, SD= 2.83, range = 9 months) recruited from a mixed-age classroom at Amanah school. Age was controlled to address potential cognitive abilities across different age groups. Participants were grouped to A, B and C in order to counterbalance word condition as explained in Chapter 4 section 4.3. Ethics approval was obtained from the University of Birmingham Research Ethics Board, and consent was obtained from participants' parents. Participants received a "Thank you" sticker for their participation. According to parents' reports, all children were neurotypical with normal or corrected-to-normal vision and normal hearing and no background in any language impairment or delay (e.g., dyslexia). Participants were required to have been exposed to English from birth but to have phonological awareness of Modern Standard Arabic (i.e., English is their L1, but they know the sounds of Arabic letters) to control for any effect of language dominance and to minimize task difficulty for the children when learning (i.e., longer time in processing or recognizing speech sounds of learning materials). Additionally, participants needed to be able to write in Modern Standard Arabic (MSA) for this study. This was verified through a Language Background Questionnaire (Appendix 4) that the parents filled out, which was reviewed by the child's class teacher at Amanah School to confirm their level of MSA language ability. All 30 participating children completed the recall and recognition tests.

5.4. Materials

As in Experiment 1, 72 unfamiliar Modern Standard Arabic nouns were used in three conditions: listen, repeat, and write. There were 9 words on the training list, with 3 words assigned to each condition. The study list consisted of 63 words (the same list used in Study 2), with 21 words in each condition. Words were counterbalanced across the three lists corresponding to the study conditions: listen, repeat, and write, and checked for familiarity. Before the experiment, a picture-naming task was conducted to determine word familiarity. Despite the training list containing frequently used MSA words (such as "apple" and "banana"), most children could not name these words correctly in Arabic, although they could name the words in English, French, and Urdu. A few children could name some of the target words in Arabic - in these cases, those words were removed from any further analysis. Similar to experiment 2, words were presented randomly and in isolation to ensure independent acquisition during the learning phase. However, they were evaluated as competitors during the testing phase. To illustrate, lexically related words such as /raff/ “shelf,”/daff/ “tambourine,” and /xizaːna/ “closet” were displayed separately and randomly in an intermixed list during the learning process but were presented as competitors during the testing phase. The full list of materials is provided in section 4.3 of chapter 4, and visual stimuli are included in Appendix 5.

5.5. Procedure: Training, Learning, and Testing

The current experiment replicated the procedure used in experiment 2, which can be described in 3 phases:

5.5.1. The training phase:

Nine MSA words were randomly assigned to the conditions for training purposes. Each word was accompanied by an instructional emoji that represented its condition. Training slides began with the phrase “Let’s Practice!”. The experimental conditions were explained to the children by using instructional images (Figure 5.2). The listen condition was introduced with an image illustrating an ear and a cross over the mouth, and the participant was instructed to listen twice to the word and remain silent. The repeat condition was introduced with ear and mouth images, and they were directed to listen to the word once and repeat it once. The writing condition was presented with an ear and a hand with a pen, and they were directed to listen to the word once and write it down once. As this experiment investigates the benefit of writing as a form of production, children were given a blank piece of A4 paper and a pencil during the training phase. Participants were instructed to write down the word presented to them during the writing condition, as illustrated by the writing symbol in Figure 5.3. They were informed that the spelling did not matter and that the paper would be collected after the learning phase but would not be corrected to minimize the children's anxiety about writing the word in the correct form instead of focusing on learning the word.

Each item was presented in auditory and visual formats for 10 seconds, and there was a 10-second gap between each item. The training and learning phases were conducted using PowerPoint slides. The experimenter's role was to refer to the instructional image to remind the participants of the condition. After a 5-minute break, children were trained on the free recall task and the forced choice recognition test.

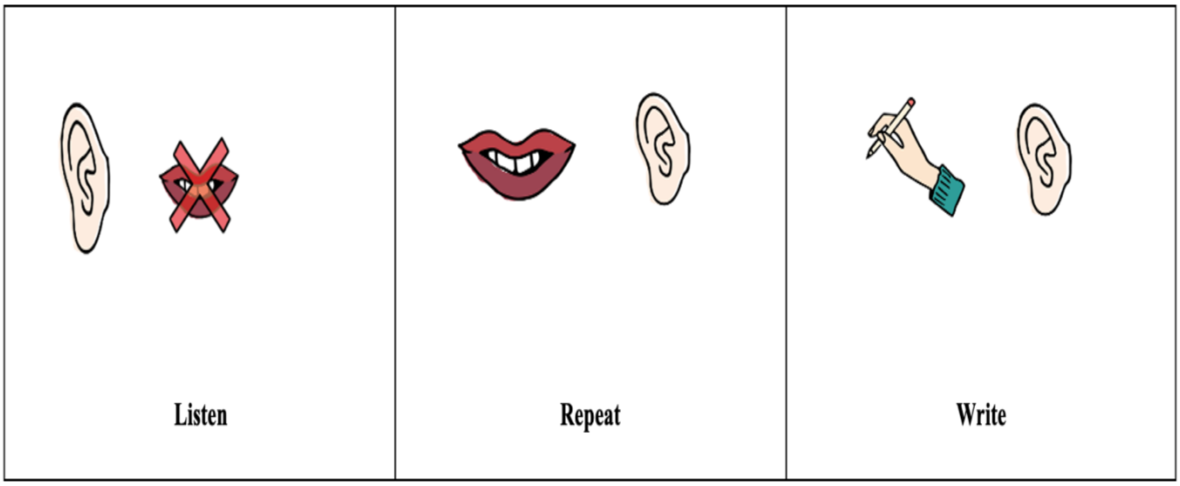


Figure 5.2: Visual instructional stimuli for the conditions

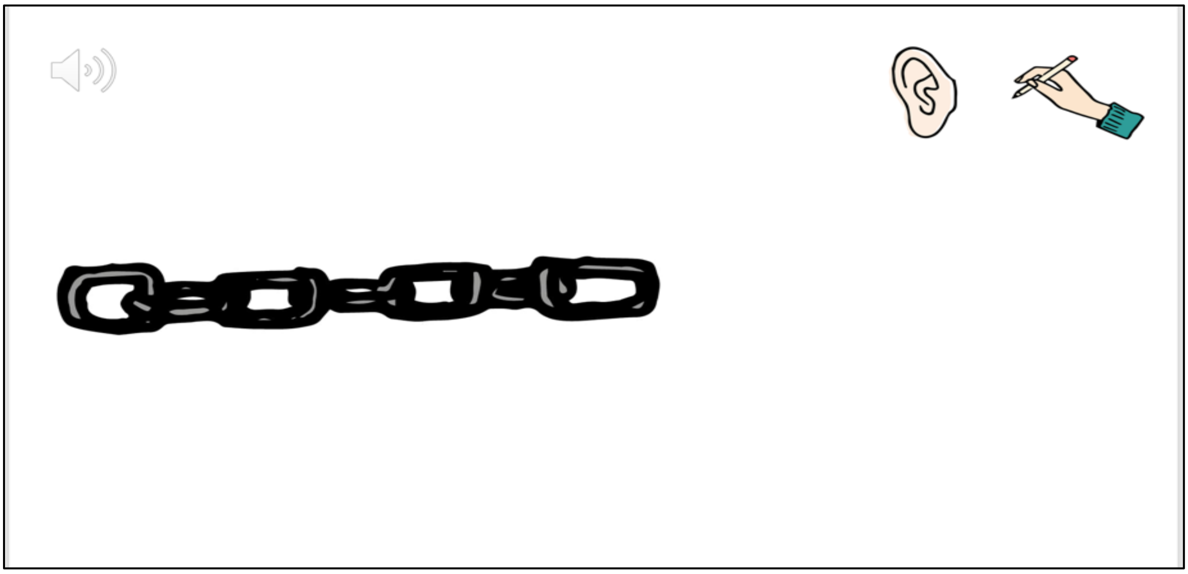


Figure 5.3: Example of a screen displaying the writing condition

5.5.2. The learning phase:

Learning trials had the same structure as training but included 63 Modern Standard Arabic (MSA) words. These words were randomly divided into three lists of 21 words each, corresponding to the study conditions: listen, repeat, or write. Each child participated in all three conditions, with items randomly intermixed across the conditions (i.e., presented in a randomized order). The order of the conditions was counterbalanced across participants to control for order effects. In the "listen" condition, children were instructed to listen attentively to each word they heard. In the "repeat" condition, children were asked to listen to the word then repeat it aloud immediately after hearing it. In the "write" condition, children were given the word orally and then instructed to write it down on a provided sheet of paper. All children learned the same set of 63 words, ensuring consistency across participants. By counterbalancing the conditions, we ensured that each child experienced a different sequence of the listening, repeating, and writing tasks. This means that the order in which the tasks were presented was varied across participants to prevent any order effects, where the position of a task might influence performance. For example, if the 'listen' task was always presented first, it might lead to a learning advantage simply due to the initial exposure. Counterbalancing eliminates this potential bias by distributing the sequence of tasks across participants. Additionally, within each condition, the words were presented randomly, meaning no participant encountered the same sequence of words. This randomization further reduces any potential learning advantages related to specific word order, ensuring that the results reflect the impact of the learning condition itself rather than the effects of encountering particular words earlier or later in the task.

5.5.3. The testing phase:

The test phase was divided into three parts: recall, 5- minute break, and recognition task. During the recall phase, children were prompted to recall as many words as possible (e.g., 'Can you tell me the words you remember?'). A 5-minute break followed the recall phase to prepare for the recognition test. During the recognition test, children were instructed to click on their preferred image as quickly as possible. The initial screen had written instructions: listen to the question, then select the answer. When participants understood, they clicked "I'm ready!" to start the test. There was a blank screen for 10 seconds, followed by four images and an auditory question.

In total, 63 trials were presented, each displaying four randomized images. The trials tested various combinations of target words alongside semantic and phonological competitors and unrelated words (Figure 5.4). Specifically, the image combinations for each trial were grouped into the following configurations: (1) a target word, a semantic competitor, a phonological competitor, and an unrelated word (Figure 5.4); (2) a target word, a semantic competitor, and two unrelated distractors; (3) a target word, a phonological competitor, and two unrelated distractors; and (4) a target word with three unrelated distractors. This design introduced diversity in the visual and lexical contexts presented to participants, aiming to engage various word recognition and retrieval aspects under distinct conditions. A 10-second blank screen was displayed between each trial to ensure separation of the test screens. The testing session lasted approximately 30 minutes per participant, excluding the additional 30 minutes for training and word learning.

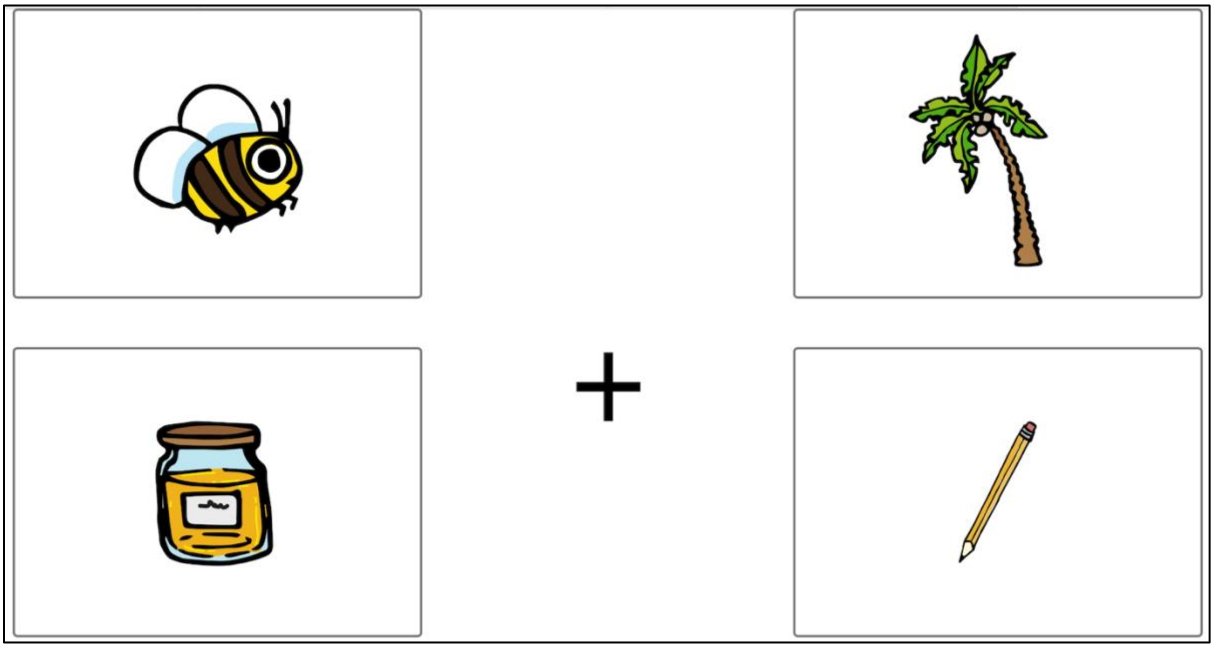


Figure 5.4: Example of a test trial as it appeared on Gorilla from Group (1)

Target: /naħla/ “bee”. Phonological competitor: /naxla/ “palm tree”. Semantic competitor: /ˈʕəsəl/ “honey”. Unrelated word: /mursa:m/ “pencil”.

5.6. Analysis

The same analysis was used as in the previous chapter. The primary dependent variables in this study were accuracy (for the recall test) and accuracy and reaction time (for the recognition test). Results were examined for two main effects: whether items were observed silently, repeated or written (conditions: listen, repeat, and write) and the time of the post-test (post-test: immediate, one-week delayed, or two-week delayed). The recognition test analysis included the word type (target, phonological competitor, semantic competitor, unrelated) as an additional factor to determine correct and incorrect responses.

5.6.1. Coding

Responses to the recognition test were coded in Gorilla. Testing trials were coded for accuracy (1 = correct, 0 = incorrect). Excel was used to code the recall tasks as recalled (= 1) or not recalled (= 0). The recalled item was considered correct if retrieved as learned in its MSA form. Any retrieval of English words was considered incorrect as the participants' L1 is English, and learning MSA is the primary focus. For example, if the child recalled /ˈʃa:ħinah/ (Shahina) as it is, they would score 1 (correct answer), but if the word were recalled as "truck" or "big car" in English, the score of the word marked to be 0 (incorrect). Two participants recalled some items in their spoken Arabic dialect which were marked as correctly recalled (e.g., /trylla:/ (Trilla) for the truck /ˈʃaːħinah/ and /du:la:b/ (Dulab) for closet /xizaːna/).

5.6.2. Statistical Analysis

Prior to analysis, data cleaning was conducted in R version 4.0.5 (R Core Team, 2021) after extracting the data from Gorilla. The data was organized to include subject, test, reaction time, items, and accuracy of answers. Reaction times (RTs) were log-transformed to reduce skewness. The analysis included 5759 data points across the three recall tests and 4651 data points across the three recognition tests.

The analysis included two stages: (1) analysis of accuracy in recall; and (2) analysis of accuracy and reaction time in recognition. The analysis also included a sub-analysis of incorrect responses in the recognition test to examine the influence of word type (i.e., lexical competitors) to understand the mechanism underlying the production of newly learned words when target words are not chosen.

To analyze the accuracy data, a generalized linear mixed-effects model (GLMM) with a binomial distribution was used. The model was fitted using glmer, a type of regression model that can include both random and fixed effects. The "bobyqa" optimizer was employed for the maximum likelihood estimation. To analyze reaction times for correct answers, linear mixed-effects models were constructed and analyzed using the lme4 (version 1.1-26; Bates et al., 2015) and lmerTest (version 3.1-3; Kuznetsova et al., 2017) packages. The maximal random structure model included the interaction of all conditions with tests, and a fully random model was used whenever possible. If a model failed to converge, the random slopes were simplified. A multiple comparisons test, using Bonferroni's correction, was conducted with the 'emmeans' package (Lenth, 2022) to identify significant differences among the groups.

5.6.3. Results

Overall recall and recognition results are shown in Table 5.1 which presents the mean and standard deviation (in brackets) of accuracy (%) and reaction time (milliseconds) based on the three conditions (listen, repeat, and write) in the three tests (immediate, one-week delayed, two-week delayed). The recall test showed a consistent pattern of superior recall accuracy for the "Write" condition, compared to the "Listen" and "Repeat" conditions, with the latter being the least effective. In the recognition test, both production conditions showed an advantage over the "Listen" condition. However, performance in the "Write" condition was better during the two-week delayed test. Reaction time shows an interesting pattern, with the "Repeat" condition showing faster response times in the immediate test, while the "Write" condition showed faster response times in the two-week delayed test.

Table 5.1: Mean and SD (in parentheses) of correct answers (%) and RTs (in milliseconds) based on condition and test intervals

Recall		**immediate (n= 30)**	**one-week (n= 30)**	**two-week (n=30)**
Accuracy Mean (SD)	Listen	38.7 (48.7)	42.9 (49.5)	40.0 (49.0)
	Repeat	28.8 (45.3)	29.5 (45.6)	33.2 (47.1)
	Write	53.1 (49.9)	57.2 (49.5)	65.5 (47.5)
Recognition		**immediate (n= 30)**	**one-week (n= 30)**	**two-week (n=30)**
Accuracy Mean (SD)	Listen	63.3 (48.2)	59.8 (49.0)	63.2 (48.2)
	Repeat	75.7 (42.9)	76.1 (42.6)	80.3 (39.7)
	Write	76.6 (42.3)	80.8 (39.4)	91.8 (27.4)
RT Mean (SD)	Listen	3324.1 (1204.8)	3261.0 (1288.8)	2795.5 (967.8)
	Repeat	3153.3 (1176.6)	3041.5 (1190.6)	2748.5 (1072.8)
	Write	3269.3 (1245.2)	3046.6 (1072.1)	2680.8 (988.9)

5.6.3.1. Accuracy in recall

A generalized linear mixed effect model (glmer) with a binomial distribution was used to evaluate recall accuracy, incorporating fixed effects of condition, test, and their interaction, along with random intercepts for subjects and items. Initially, the model included by-subject random slopes for the effects of condition, test, and their interaction, as well as by-item random slopes for the effect of Test. However, this led to unavoidable convergence errors. As a result, the random slopes were removed to simplify the model and ensure proper convergence. The model presented in Table 5.2 shows a significant effect of the repeat condition and the interaction between the write condition and the two-week delayed test. This is illustrated in Figure 5.4, which demonstrates the effect of condition and the results over test time. Figure 5.5 indicates no significant difference between the listen and write conditions overall, although the write condition shows significance in the two-week delayed test, as detailed in Table 5.3.

Table 5.2: Mixed effects model output for the interaction between condition and test

Formula: glmer (Correct ~ Condition*Test + (1 + | ID) + (1 | answer)

control = glmerControl(optimizer= "bobyqa"), family = binomial)

Fixed effects	Estimate	SE	z-value	p-value
(Intercept)	-0.15	0.16	-0.94	0.34
Condition: repeat	-0.87	0.14	-6.15	<.001***
Condition: write	0.06	0.14	0.48	0.62
Test: one-week delayed	0.19	0.13	1.39	0.16
Test: two-week delayed	0.05	0.13	0.42	0.67
repeat*one-week delayed	-0.15	0.18	-0.86	0.38
write*one-week delayed	-0.02	0.18	-0.11	0.91
repeat* two-week delayed	0.15	0.18	0.86	0.38
write*two-week delayed	0.55	0.18	3.01	0.002**

Random effects variance (subject = 30, item = 63)

* $p < 0.05$; ** $p < 0.01$; *** $p < .001$

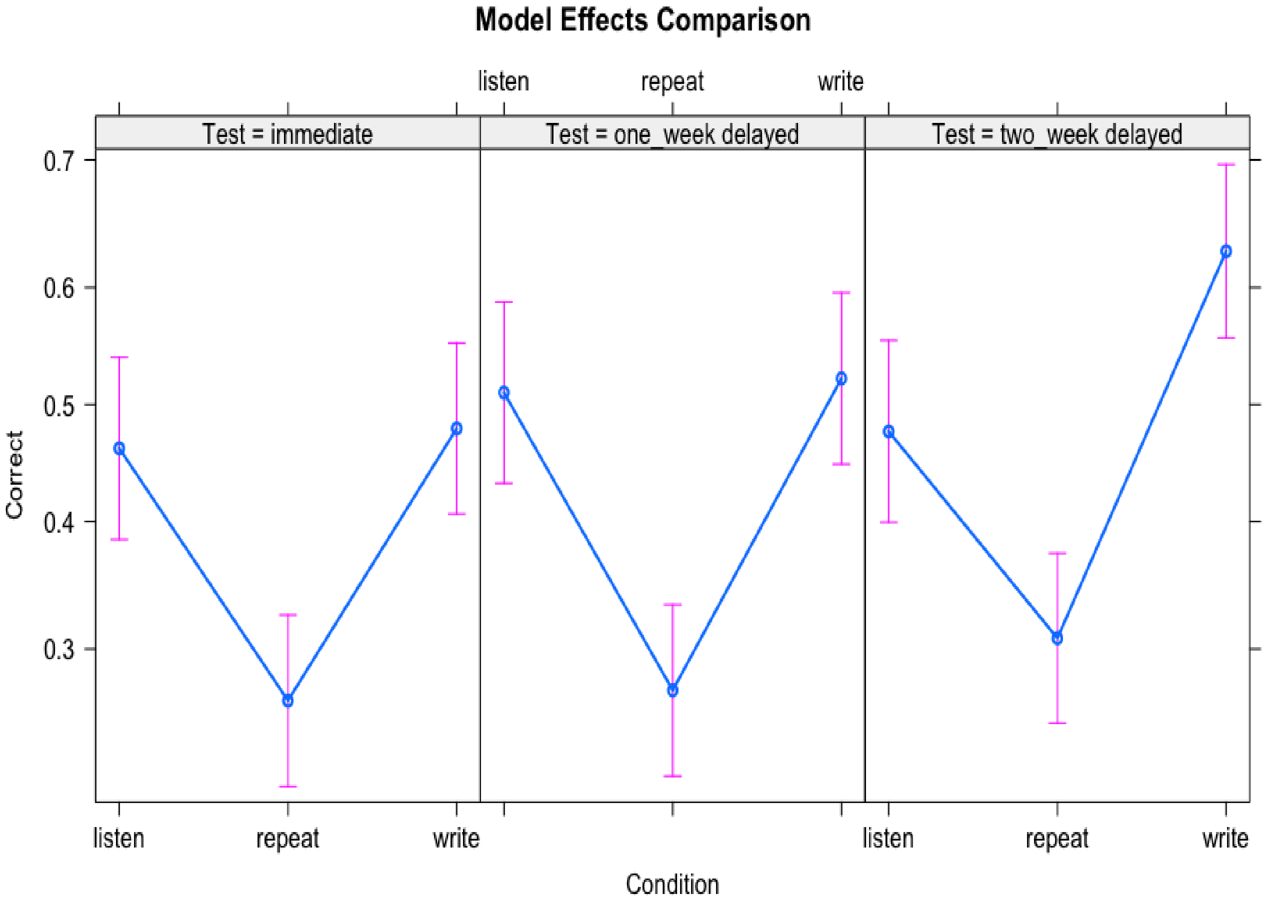


Figure 5.5: The interaction of condition and test time on accuracy of answers in recall. y-axis is log-odds of response accuracy

A multiple pairwise comparisons test using the package "emmeans" (Lenth, 2022) shows several significant differences between the conditions during the test intervals, as presented in Table 5.3. During the immediate test, there was a significant difference between the listen and repeat conditions (z-ratio = 6.15, $p < .001$), and between the repeat and write conditions (z-ratio = -7.12, $p < .001$), indicating that the repeat condition performed worse than both listen and write conditions. No significant difference was found between the listen and write conditions (z-ratio = -0.48, $p = 1.000$). In the one-week test, the pattern was similar. There was a significant difference between the listen and repeat conditions (z-ratio = 7.33, $p < .001$), and between the repeat and write conditions (z-ratio = -8.15, $p < .001$), again showing that the repeat condition was worse. There was no significant difference between the listen and write conditions (z-ratio = -0.34, $p = 1.000$).

In the two-week test, all conditions showed significant differences. The listen condition outperformed the repeat condition (z-ratio = 5.12, p < .001). However, the write condition showed a significant advantage over the listen condition (z-ratio = -4.41, p = .0004). The repeat condition performed significantly worse than the write condition (z-ratio = -10.0, p < .001). These results indicate that the repeat condition was consistently less effective across all test points. Notably, there was an emerging advantage for the write condition over the listen condition in the two-week delayed test, suggesting that writing may facilitate longer-term retention of the words.

Table 5.3: Post-hoc tests with Bonferroni correction for accuracy based on the interactions between condition and test model in recall

Contrasts	**Estimate**	**SE**	**z-ratio**	**p-value**
Immediate test: listen - Immediate test: repeat	0.87	0.14	6.15	<.001***
Immediate test: listen - Immediate test: write	-0.06	0.14	-0.48	1.0000
Immediate test: repeat - Immediate test: write	-0.94	0.13	-7.12	<.001***
One-week test: listen - One-week test: repeat	1.03	0.14	7.33	<.001***
One-week test: listen - One-week test: write	-0.04	0.14	-0.34	1.0000
One-week test: repeat - One-week test: write	-1.08	0.13	-8.15	<.001***
Two-week test: listen - Two-week test: repeat	0.71	0.14	5.12	<.001***
Two-week test: listen - Two-week test: write	-0.62	0.14	-4.41	0.0004***
Two-week test: repeat -Two-week test: write	-1.34	0.13	-10.0	<.001***

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

5.6.3.2. Accuracy in recognition

Table 5.4 shows that there were no significant differences between the listen and repeat conditions. However, a significant interaction was observed for the write condition in both the one-week delayed test (p = 0.04) and the two-week delayed test (p < 0.001), as demonstrated in Figure 5.5 and evident in the pairwise comparisons test in Table 5.5. This suggests that the written words are not only more durable compared to the listen and repeat conditions but also show an improvement in retention over time, as indicated by the increase in scores observed in Figure 5.6. This improvement could be attributed to the combined effects of the testing effect, where retrieval practice enhances memory, and the production effect, where actively writing down words leads to better long-term retention.

Table 5.4: Mixed effects model output for the interaction between condition and test

Formula: glmer (correct ~ condition * test + (1 | ID) + (1| Answer)

control = glmerControl(optimizer= "bobyqa"), family = binomial)

Fixed effects	Estimate	SE	z-value	p-value
(Intercept)	1.14	0.25	4.45	<.001***
Condition: repeat	-0.07	0.17	-0.43	0.66
Condition: write	0.21	0.17	1.20	0.22
Test: one-week delayed	-0.14	0.16	-0.87	0.38
Test: two-week delayed	0.13	0.17	0.77	0.43
repeat*one-week delayed	0.15	0.22	0.67	0.49
write*one-week delayed	0.46	0.23	1.99	0.04**
Repeat*Two-week delayed	0.22	0.23	0.95	0.34
Write*Two-week delayed	1.59	0.26	5.98	<.001**
Random effects variance (subject = 30, item = 63)				

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

The pairwise comparisons in Table 5.5 reveal significant differences in the delayed tests. Specifically, during the one-week delayed test, there was a significant difference between the listen and write conditions (z-ratio = -3.76, p < .001) and between the repeat and write conditions (z-ratio = -3.42, p = 0.02), indicating that the write condition led to better retention of the words compared to both listen and repeat conditions. In the two-week delayed test, the differences were even more pronounced. There was no significant difference between the listen and repeat conditions (z-ratio = -0.80, p = 1.000). However, there were significant differences between the listen and write conditions (z-ratio = -8.20, p < .001) and between the repeat and write conditions (z-ratio = -7.64, p < .001). This demonstrates that the write condition resulted in significantly better retention compared to both the listen and repeat conditions. These results suggest that the learning effect observed when participants are asked to write out the prompt words is not only more durable but also leads to an improvement in memory retention over time. This enhancement in retention is evident in the higher scores observed in the delayed tests, underscoring the effectiveness of writing as a powerful method for learning and retaining new words.

Table 5.5: Post-hoc tests with Bonferroni correction for accuracy based on the interactions between condition and test model in recognition

Contrasts	Estimate	SE	z-ratio	p-value
Immediate test: listen - Immediate test: repeat	0.07	0.17	0.43	1.0000
Immediate test: listen - Immediate test: write	-0.21	0.17	-1.20	1.0000
Immediate test: repeat - Immediate test: write	-0.29	0.17	-1.71	1.0000
One-week test: listen - One-week test: repeat	-0.07	0.17	-0.44	1.0000
One-week test: listen - One-week test: write	-0.67	0.18	-3.76	<.001***
One-week test: repeat - One-week test: write	-0.59	0.17	-3.42	0.02**
Two-week test: listen - Two-week test: repeat	-0.14	0.18	-0.80	1.0000
Two-week test: listen - Two-week test: write	-1.80	0.22	-8.20	<.001**
Two-week test: repeat -Two-week test: write	-1.66	0.21	-7.64	<.001**

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

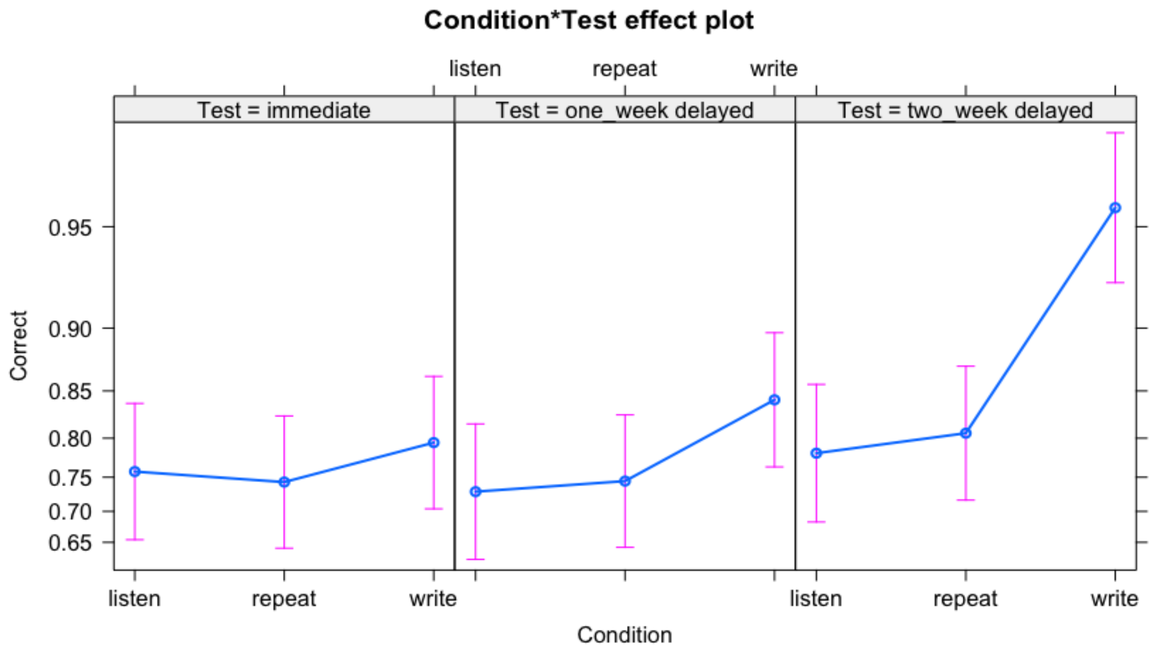


Figure 5.6: The interaction of condition and test time on accuracy of answers in recognition. y-axis is log-odds of answer accuracy

5.6.3.3. Reaction time in recognition

A linear mixed-effects model (lmer) was fitted to evaluate the log transformed reaction time to correct answers. The same base model as for accuracy was fitted which included the interaction between condition and test time as fixed effects, and by-subject random slopes for test. Adding random intercepts caused unavoidable convergence issue, hence the simplified model is the best fitted model.

Table 5.6 highlights a significant effect observed in the two-week delayed test ($p < .001$). This is further illustrated in Figure 5.7, where participants demonstrated faster reaction times in the two-week delayed test compared to both the immediate and one-week delayed tests. Specifically, reaction time (RT) was significantly faster in the two-week delayed test ($z = -5.86$, $p < .001$), indicating an improvement in retrieval speed over time. However, no significant differences in reaction times were found between the conditions (listen, repeat, write), as shown by the pairwise comparisons in Table 5.7, only improvement over time across all conditions. For example, the comparisons between the listen and repeat conditions ($z = -0.80$, $p = 1.000$) and between the listen and write conditions ($z = -1.20$, $p = 1.000$) at the two-week test point did not show significant differences.

Overall, the reaction time results suggest improved performance in the two-week delayed test, implying that participants became more efficient in responding to test items over time, possibly due to increased familiarity or better retention of the material. However, the pairwise analysis (Table 5.7) indicates that the interaction between conditions and test intervals was not significant, meaning that the type of condition (listen, repeat, write) did not significantly influence reaction times within each test interval.

Table 5.6: Mixed effects model output for reaction time (log-transformed)

Formula: lmer (RT_log ~ Condition*Test + (1 | ID) + (1 | Answer)

Fixed effects	Estimate	SE	z-value	p-value
(Intercept)	8.01	2.91	275.08	<.001***
Condition: repeat	-8.71	2.38	-0.36	0.71
Condition: write	1.82	2.39	0.76	0.44
Test: one-week delayed	-1.44	2.45	-0.59	0.55
Test: two-week delayed	-1.47	2.51	-5.86	<.001***
repeat*one-week delayed	-2.21	3.13	-0.70	0.48
write*one-week delayed	-4.77	3.15	-1.51	0.13
Repeat*Two-week delayed	2.89	3.17	0.91	0.36
Write*Two-week delayed	-4.05	3.18	-1.27	0.20
Random effects variance (subject = 30, item = 63)				

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

Table 5.7: Post-hoc tests with Bonferroni correction for reaction time based on the interactions between condition and test model

Contrasts	Estimate	SE	z-ratio	p-value
Immediate test: listen - Immediate test: repeat	0.00	0.02	0.36	1.0000
Immediate test: listen - Immediate test: write	-0.01	0.02	-076	1.0000
Immediate test: repeat - Immediate test: write	-0.02	0.02	-1.23	1.0000
One-week test: listen - One-week test: repeat	0.03	0.02	1.30	1.0000
One-week test: listen - One-week test: write	-0.00	0.02	-0.06	1.0000
One-week test: repeat - One-week test: write	0.02	0.02	1.23	1.0000
Two-week test: listen - Two-week test: repeat	-0.02	0.02	-0.83	1.0000
Two-week test: listen - Two-week test: write	0.02	0.02	0.92	1.0000
Two-week test: repeat -Two-week test: write	0.04	0.02	1.96	1.0000

* $p < 0.05$; **$p < 0.01$; ***$p < .001$

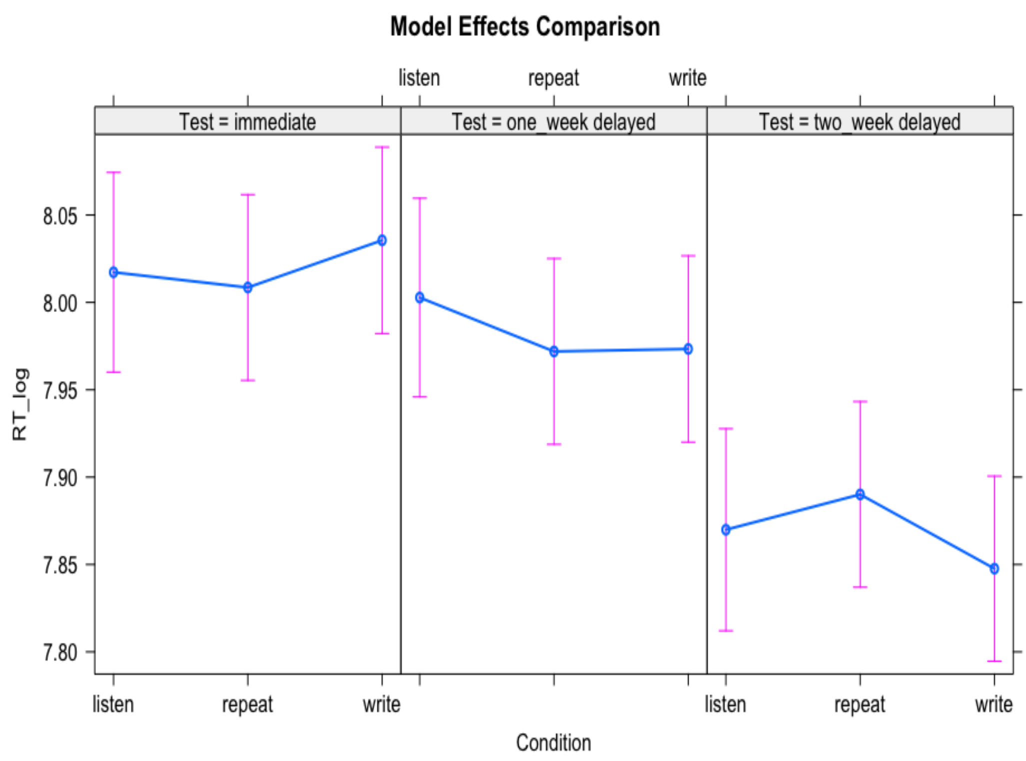


Figure 5.7: The interaction of condition and test on response time in the recognition test

5.7. Discussion

Many studies on the production effect (PE) in children have utilized different conditions such as listening, reading, saying, and looking to assess the impact of producing words on memory, as discussed in section 5.1. The current study aims to investigate writing as a production condition within the framework of PE theory in eight-year-old children. This experiment employed a within-subject design where children learned 63 unfamiliar Modern Standard Arabic (MSA) words in a mixed list under three conditions: listen, repeat, and write.

The results indicate that the "Write" condition consistently outperforms the "Listen" and "Repeat" conditions in the recall test, with "Repeat" being the least effective (Figure 5.5). In the recognition test (Figure 5.6), the performance of the conditions varies across the three testing intervals. Initially, "Listen" and "Write" perform similarly, with "Repeat" being lower. At the one-week delayed test, all conditions show improvement, with "Write" achieving the highest performance. By the two-week delayed test, the "Write" condition exhibits a significant increase in correct responses, indicating superior long-term retention, while "Listen" and "Repeat" remain stable. These findings suggest that writing enhances retention and retrieval speed more effectively over time than listening or repeating. Reaction time (Figure 5.7) shows distinct patterns, with the "Repeat" condition leading to quicker responses immediately after learning, while the "Write" condition results in faster responses after a two-week delay. This shift implies that writing may enhance information retention and retrieval speed over time. The findings demonstrate that production activities, mainly writing, are more effective in facilitating learning than listening and repeating. While the effect of delayed testing was extensively discussed in the previous chapter, this section discusses the implications of writing on vocabulary learning and memory retrieval. The findings are analyzed within the broader context of the Production Effect (PE) theory, focusing on the underlying learning mechanisms that enhance memory retention and retrieval in young learners.

How Does Writing Enhance Vocabulary Acquisition and Memory Retention in L2 Learning Compared to Listening and Repeating Conditions?

Writing offers significant advantages over passive listening and repetition by involving more active, demanding, and multimodal cognitive activities. According to Chen et al. (2018), the additional effort required to write words out instead of simply repeating them engages learners in deeper cognitive processing, which enhances their ability to remember the vocabulary over time. The current findings indicate that writing facilitates better memory retention in L2 vocabulary learning through these mechanisms, making it more effective than listening and repeating condition.

Writing, by its nature, demands a higher level of cognitive engagement than listening or repeating. It integrates visual, motor, and cognitive processes, forming more durable memory traces. This deeper cognitive processing is essential for effectively acquiring and retaining vocabulary. Learners are compelled to consider each word's spelling, structure, and meaning when writing words, facilitating a more comprehensive understanding and robust memorization of vocabulary items (Reed, 2006). Engaging in the orthographic representation of words requires the learner to undertake more intricate cognitive activities than merely listening or repeating, leading to deeper processing and, consequently, enhanced long-term retention.

A key distinction in this study is the performance differences between recall and recognition across the learning conditions. Recall requires retrieving information without external cues, whereas recognition involves identifying learned information from provided options, typically requiring less cognitive effort (Craik & Lockhart, 1972). The "Write" condition significantly improved recall accuracy, especially in delayed tests. This result is consistent with the generation effect (Nairne & Widner, 1987), which asserts that generating information (in this case, writing the word) enhances memory retention by engaging deeper cognitive processes. Writing involves a complex interaction of visual, motor, and linguistic systems, leading to more robust encoding and durable memory traces, which explains the superior recall performance. The multimodal engagement through dual coding theory (Paivio, 1986), where both visual and auditory systems are activated, further supports the enhanced recall and recognition observed in the writing condition. In contrast, the "Listen" and "Repeat" conditions showed less pronounced improvements in recall, particularly in the two-week delayed test. Listening, a passive form of learning, does not sufficiently engage learners in active retrieval or deep cognitive processing, resulting in weaker memory traces and lower recall performance. While the "Repeat" condition was more active than listening, it still failed to improve recall over time significantly. This suggests that without the generative writing component, rote repetition does not effectively enhance memory retention (Barcroft, 2007).

For recognition tasks requiring lower cognitive demand, all conditions showed improvement, though the "Write" condition consistently outperformed the others. This superior performance in recognition suggests that writing, by engaging multiple sensory modalities, creates more retrieval pathways, facilitating easier word recognition even after longer intervals. A significant finding of this study is the role of time in highlighting the advantages of the "Write" condition.

One of the key findings of this study was the significant interaction between learning condition and test interval, with the writing condition showing progressively better performance over time. While performance was similar across conditions in immediate tests, the two-week delayed tests revealed a clear advantage for writing in both recall and recognition tasks. The writing condition showed higher accuracy and faster retrieval times in delayed retention tests, suggesting that writing strengthens long-term memory consolidation and improves retrieval efficiency. This time-dependent improvement aligns with the testing effect (Roediger & Butler, 2011), where the process of active retrieval during learning strengthens memory traces, leading to better long-term retention. In contrast, listening and repeating did not yield the same long-term benefits, as the absence of active retrieval in these conditions resulted in weaker memory traces, particularly in delayed recall tasks.

Interestingly, no significant differences between the listening and repeating conditions in delayed recall and recognition tests were found. While repetition showed a modest improvement in recognition immediately after learning, the lack of sustained improvement over time suggests that repetition alone does not provide sufficient cognitive engagement for long-term retention. Listening, a purely passive activity, resulted in the weakest performance in recall and recognition. These findings align with research indicating that rote repetition and passive listening do not engage the deeper cognitive processes necessary for effective vocabulary acquisition and retention (Laufer & Hulstijn, 2001). Both conditions rely on shallow processing, which is less effective for creating durable memory traces, especially in delayed testing.

Several cognitive mechanisms can explain the superior performance of the "Write" condition. First, writing engages learners in active retrieval, making words more distinctive and easier to recall (MacLeod et al., 2010). The distinctiveness account suggests that actively producing information, such as writing, makes the learned material stand out in memory, enhancing recall and recognition. This distinctiveness, combined with the dual coding of visual and auditory information, explains why the writing condition led to significantly better retention in delayed tests. Additionally, writing involves deeper levels of processing than either listening or repeating, consistent with Craik and Lockhart's (1972) levels of processing theory. The motor, visual, and cognitive processes required to generate written words create stronger and more durable memory representations than passive listening or repeating. This depth of processing is crucial for transferring information from short-term to long-term memory, which explains the superior performance of the writing condition in delayed tests.

This principle aligns with cognitive load theory (Sweller, 1988), differentiating intrinsic, extraneous, and germane cognitive load. Writing imposes germane cognitive load as a complex cognitive activity—the cognitive effort that directly facilitates learning and schema construction (Sweller et al., 1998; Paas et al., 2003). By introducing desirable difficulties, writing challenges the learner and promotes deeper learning and better vocabulary retention (Bjork & Bjork, 2011; McNamara & Healy, 2000). Though effortful in the short term, these desirable difficulties create conditions that lead to more robust long-term learning outcomes (Bjork, 1994; Roediger & Karpicke, 2006). In L2 vocabulary acquisition, writing tasks serve as a form of productive retrieval practice, and they have been shown to enhance retention more effectively than passive review (Barcroft, 2007; Webb, 2005).

Moreover, the multimodal nature of writing engages visual, motor, and linguistic processes, creating more elaborate and distinctive memory representations of new vocabulary items (Perfetti & McCutchen, 2000; Longchamp et al., 2008). By combining the generation effect with dual coding, writing emerges as a powerful tool for vocabulary acquisition. Learners actively produce language while creating visual representations of words, strengthening memory traces through multiple cognitive channels. This multifaceted engagement with vocabulary significantly increases the likelihood of long-term retention and retrieval.

Therefore, writing facilitates deeper cognitive processing and enhanced memory retention in L2 vocabulary learning by engaging learners in complex cognitive processes, introducing desirable difficulties, leveraging the generation effect, utilizing dual coding, and enhancing distinctiveness. These mechanisms collectively make writing a highly effective vocabulary acquisition and retention strategy, outperforming listening and repetition by demanding higher cognitive effort and engaging multiple sensory modalities. Consequently, writing leads to more robust and durable learning outcomes.

5.8. Conclusion and Limitations

The findings of this study demonstrate that writing is a more effective strategy for both recall and recognition of L2 vocabulary than listening and repeating. Writing engages learners in deeper cognitive processing, promotes active retrieval, and creates more distinctive memory representations, resulting in superior long-term retention. The significant improvements observed in delayed recall and recognition tasks emphasize the importance of incorporating productive, multimodal activities like writing into language learning practices to optimize vocabulary acquisition and memory retention. In contrast, while listening and repeating offer some short-term benefits, they do not facilitate the same depth of processing or long-term retention as writing. Despite these robust findings, several limitations must be considered. First, the sample size in Experiment 2, which involved three conditions, could have been more significant in increasing statistical power and improving the generalizability of the results. Second, the increased cognitive load in the writing condition may have introduced variability in performance, particularly among participants with lower proficiency in Modern Standard Arabic (MSA), highlighting the need for future studies to account for individual differences. Another limitation lies in the lack of focus on individual cognitive differences, such as working memory or attention, which could have influenced the effectiveness of the different learning conditions. Lastly, repeated exposure to the same items across multiple testing intervals might have introduced learning effects independent of the conditions. Future research should address these limitations by increasing the participant pool, incorporating assessments of cognitive differences, and exploring the potential of combining writing with other active learning strategies to enhance vocabulary learning outcomes further.

Chapter 6. General Discussion

This thesis conducted three experiments to assess the production effect (the PE), a memory benefit associated with producing information during encoding (by writing or repeating words aloud), in facilitating children's learning of unfamiliar vocabulary. This study examined several contextual factors, including the type of production method (listening, repeating, or writing), assessment tasks (recall vs recognition), linguistic background, age, and the delay before testing. The aim of this thesis was to investigate the role of production in the vocabulary acquisition process during second language learning. It focused on how learners associate unfamiliar Modern Standard Arabic (MSA) vocabulary with known words in English or other Arabic dialects (i.e., Yamani, Egyptian, Cairene, etc). This association is based on contextual learning strategies employed during the encoding phase. This thesis is not a comprehensive investigation into learning Arabic more broadly; rather, it focuses on the precise mapping of new lexical items within the existing linguistic knowledge of young MSA learners. This chapter will focus on the findings from Experiments 1 and 2, discussing the various factors that influenced MSA vocabulary acquisition, with particular emphasis on the role of production in enhancing learning and retention.

To briefly summarise the results of the two studies: Experiment 1 (Chapter 4) evaluated the effect of production on MSA word learning, with participants learning a list of 64 MSA words under two conditions: listening and repeating. Words were chosen to better control for factors like word difficulty and familiarity, ensuring a more focused investigation of the Production Effect (PE). This experiment also controlled the participants' age, minimizing any potential influence on memory performance. Free recall and recognition tests occurred immediately, one week, and two weeks after learning. The findings show no significant difference between the learning conditions.

Experiment 2 (Chapter 5) extended the investigation to children aged eight, examining the PE under three conditions: listen, repeat, and write. These children were chosen because they possess the advanced motor and cognitive skills required for writing words in Modern Standard Arabic (MSA) and have reliably developed writing abilities, which younger children may not have achieved to the same extent. This approach aimed to facilitate a more controlled investigation of writing as a production method. Therefore, this experiment offered a new approach by including a "write" condition alongside the existing "listen" and "repeat" conditions. This inclusion addresses a critical gap in the literature on PE in children. While previous studies have explored various forms of vocal production, the influence of writing as a distinct production modality on vocabulary learning still needs to be examined in children's PE studies. The findings show that the "write" condition enhanced the distinctiveness and memorability of learned words. These findings highlight the significant role of writing in enhancing memory retention and underscore its relevance in learning vocabulary.

In conclusion, the experiments provide valuable insights into the Production Effect and its role in second language vocabulary acquisition among children. While the results did not consistently show a significant advantage for production methods across all conditions, they shed light on the specific role of writing in vocabulary learning. These findings suggest that writing, as a production method, may offer distinct benefits for memory retention in specific contexts, highlighting the need for further investigation into its potential to enhance vocabulary acquisition.

6.1. The Production Effect on Vocabulary Learning

The findings show varied impacts of PE on vocabulary recall and recognition across immediate, one-week, and two-week delayed tests. In Experiment 1, there was no significant difference between the listening and repeating conditions regarding recall and recognition accuracy across immediate and delayed tests, indicating that neither condition provided a distinct advantage for retention over time. In Experiment 2, the writing condition outperformed both listening and repeating conditions, particularly in the two-week delayed tests, highlighting the durability of memory retention when words are written as opposed to only listened to or repeated aloud. Moreover, participants’ reaction times were significantly faster in delayed tests across all conditions, suggesting enhanced retrieval efficiency over time.

6.1.1 The Null Effect of Production

Experiment 1 investigated the production effect (PE) in learning Modern Standard Arabic (MSA) vocabulary among 6-year-old children. This age group was chosen strategically to control for cognitive abilities across participants and ensure a consistent baseline. However, the absence of a significant production effect in this study challenges the generalizability of the PE. It suggests that deeper processing through active production may not universally enhance memory retention, particularly in young learners or with certain types of linguistic material, as it was assumed in the children's PE literature. The observed interaction between learning conditions and memory outcomes indicates that the learning task design may have placed excessive cognitive demands on participants during the production phase, potentially complicating memory consolidation. This aligns with the cognitive load theory (Sweller, 1988), which posits that learning is impaired when working memory capacity is exceeded. In the context of young children learning unfamiliar Arabic vocabulary, the production task may have consumed the cognitive resources needed for effective encoding and consolidation.

The null effect observed in this experiment, where no significant difference in memory was found between produced and non-produced items, underscores the variability and complexity inherent in the production effect. This finding resonates with other studies that have reported null effects in specific contexts. For instance, Hourihan and Smith (2016) found that in face-name association tasks, producing only the name (and not the face) did not enhance recall, indicating that both elements needed to be produced for the effect to manifest.

These results suggest that various factors, including participant characteristics, task design, and material properties, moderate the production effect. The complexity of Modern Standard Arabic for young learners may have played a crucial role in the observed null effect. The unfamiliarity of Arabic script and phonology for young learners can increase cognitive load, potentially overshadowing production benefits. This increased cognitive demand aligns with the findings of Yue et al. (2013), who demonstrated that the production effect can be attenuated when concurrent tasks tax cognitive resources. Furthermore, the developmental stage of the 6-year-old participants may have influenced the results. The executive functions and working memory capacities of young children are still developing (Diamond, 2013), which may have limited their ability to benefit from the distinctiveness typically afforded by production, a key mechanism proposed by Dodson and Schacter (2001) in their distinctiveness account of the production effect.

The null effect observed in this study also highlights the importance of considering the specific characteristics of the linguistic material being learned. Modern Standard Arabic, with its complex orthography and phonology, may present unique challenges for young learners. As Abu-Rabia (1997, 1999) discussed, many partially vowelized Arabic words are homographs, which can be read as different lexical items. This complexity may have interacted with the production task, potentially negating any benefits typically associated with the production effect. Additionally, the interaction between the production effect and serial position, as observed by Saint-Aubin et al. (2021) and Cyr et al. (2021), may have played a role in the null effect. These researchers found that produced items are less well recalled on the first serial positions than silently read items, while the reverse pattern is observed for the recency portion of the curve. In the context of learning Arabic vocabulary, this interaction may have been further complicated by the unfamiliarity of the material.

In conclusion, the null effect observed in this experiment with young children learning Modern Standard Arabic vocabulary underscores the need for a nuanced understanding of the production effect. It suggests that the effectiveness of production as a learning strategy may be contingent on factors such as age, linguistic complexity, and cognitive load.

6.1.2 The Effect of Writing on Vocabulary Learning

The observed pattern of results supports the notion that production methods that require a high degree of active engagement, such as writing, are advantageous for long-term vocabulary retention. Interestingly, the repeating condition's relatively lower effectiveness than listening and writing underscores the importance of task complexity in enhancing PE. The findings suggest that merely repeating words aloud may lack the cognitive depth to yield durable memory traces, especially compared to the more integrative process involved in writing.

MacLeod et al. (2010) suggest that stronger memory retention is achieved through encoding processes that engage multiple sensory modalities and involve deeper cognitive processing. In other words, memories are more likely to be retained when the information is processed in a prosperous, multisensory manner, requiring deeper thought and analysis. This perspective aligns with Paivio's Dual Coding Theory (1986), which posits that cognition involves two distinct but interconnected systems: one for processing verbal information and another for non-verbal (imagery) information. The results of Experiment 2 corroborate this perspective, demonstrating that tasks demanding higher levels of cognitive engagement, explicitly writing, consistently yielded superior retention outcomes compared to more passive activities such as listening or simple repetition. This pattern of results supports the notion that deeper, more active processing leads to more robust memory formation and retention. The writing condition, which necessitates more significant cognitive effort in motor coordination, visual processing, and semantic elaboration, appears to create stronger memory traces than the comparatively less demanding tasks of listening or repeating. These findings can be interpreted through the lens of Dual Coding Theory. Writing engages both the verbal system (through linguistic processing) and the imagery system (through the visual and motor aspects of forming letters), potentially creating dual memory codes. This dual encoding may provide multiple retrieval pathways, enhancing retention and recall. Furthermore, these results support the Levels of Processing

Theory (Craik & Lockhart, 1972), which posits that deeper engagement with material, such as the cognitive and motor processes involved in writing, enhances retention more than shallower, surface-level processing like listening or essential repetition. The writing task likely promotes deeper semantic processing and elaboration, leading to more durable memory traces.

While the literature on the Production Effect primarily addresses memory and the cognitive aspects, theories on vocabulary acquisition provide a rich context for understanding the Production Effect observed in this study. One theoretical framework particularly relevant to these findings is Transfer-Appropriate Processing (TAP) (Morris et al., 1977). TAP asserts that memory retrieval is optimized when the cognitive processes used during learning are aligned with those required during recall. In this study, the enhanced retention seen in the writing condition may be partly due to the alignment between the physical act of writing and the retrieval tasks used in the delayed tests, engaging visual and motor memory pathways. The relatively lower retention in the repeat condition, by contrast, might reflect a limited match between the task demands and the retrieval processes needed for recognition and recall, a finding echoed by Barcroft’s (2007) work on production tasks and cognitive load in vocabulary acquisition. According to Barcroft (2007), the relationship between production tasks and vocabulary acquisition is complex and influenced by cognitive load. Their study examined the effects of word and fragment writing during L2 vocabulary learning, finding that these tasks can sometimes interfere with lexical acquisition. Barcroft (2007) proposed that when learners are required to engage in output-based tasks (such as writing) during initial vocabulary learning, the cognitive resources devoted to the production process may detract from those available for form-meaning mapping. This aligns with the concept of limited processing resources in second language acquisition. Barcroft's findings suggest that the effectiveness of production tasks in vocabulary learning may depend on the learner's proficiency level, the complexity of the target words, and the specific stage of vocabulary acquisition.

The differential impacts of the various production methods on delayed retention also support the TOPRA Model (Barcroft, 2015), which emphasizes the necessity of balancing semantic, structural, and mapping processes in vocabulary acquisition. The model proposes that extensive focus on one processing type, such as listening or repetition, may limit cognitive resources available for other aspects of learning. The writing condition in Experiment 2 required semantic and structural processing and added a motor element that deepened encoding in line with the TOPRA model’s prediction. This integrative task appears to optimize lexical input processing by engaging multiple cognitive resources, thereby enhancing retention compared to less engaging tasks like repetition.

Furthermore, the significant differences in reaction times and accuracy between immediate and delayed tests in the writing condition reflect Barcroft’s insights into Lexical Input Processing (2015), which argues that vocabulary acquisition benefits from varied and context-rich tasks. As the findings indicate, the increased retention observed in the writing condition likely reflects the added cognitive and motor processing layers, reinforcing memory consolidation over time.

Finally, an important consideration is the bilingual background of participants, as this may modulate the effects of various learning methods. The findings resonate with Multicompetence Theory (Cook, 2002), which suggests that bilinguals may activate multiple linguistic systems, influencing vocabulary learning strategies in an additional language. This theory posits that the knowledge of multiple languages creates a unique cognitive state that differs from that of monolinguals, potentially affecting how new linguistic information is processed and stored. The delayed retention improvement observed in the writing condition could be linked to participants' heightened awareness of cross-linguistic differences when encountering new vocabulary in a non-native language, such as Modern Standard Arabic. Bilinguals often demonstrate enhanced metalinguistic awareness, which may lead to more analytical approaches to language learning. When engaging in writing tasks, these learners might be more attuned to the subtleties of the new language, comparing it with their existing linguistic knowledge. This aligns with studies on bilingual processing, where cross-linguistic cognitive engagement, mainly through writing, enhances retention by fostering multiple connections to lexical items (Wu & Thierry, 2010). Writing may activate the target language and the learner's other known languages, creating a rich network of associations. This multi-layered processing could contribute to stronger memory traces and more effective retrieval paths. Moreover, the bilingual advantage in executive functions, such as cognitive flexibility and inhibitory control (Bialystok et al., 2012), may affect how these learners approach different vocabulary learning tasks. The writing condition, being more cognitively demanding, might better leverage these enhanced executive functions, leading to more effective encoding and subsequent retention of new vocabulary. It is also worth considering that bilinguals' experience with different writing systems and orthographies could influence their performance in the writing condition. Depending on the participants' language background, participants might find certain aspects of the MSA script more or less challenging, potentially affecting the cognitive resources allocated to the task and, consequently, the strength of the formed memories.

In essence, the findings underscore that tasks involving higher cognitive engagement, such as writing, promote stronger memory retention in vocabulary learning. This aligns with theories highlighting the role of multisensory and deeply processed encoding, such as Dual Coding Theory and Levels of Processing. The Production Effect appears particularly potent when tasks integrate visual, motor, and semantic elements, as seen in the writing condition, which likely created dual memory codes and facilitated multiple retrieval pathways. Furthermore, the bilingual background of participants may enhance these effects, as their heightened metalinguistic awareness and cognitive flexibility support more analytical and durable learning outcomes.

Taken together, these findings suggest that the Production Effect's efficacy in vocabulary learning may depend on the cognitive demands of the task and the learner's linguistic background. Active engagement through writing, which incorporates both multisensory and cognitive processes, appears beneficial for retention, as demonstrated in the increased memory durability in delayed tests. These results highlight that task complexity and the depth of processing play central roles in memory formation. At the same time, bilingual advantages, such as enhanced metalinguistic awareness and cognitive flexibility, may further amplify retention. Thus, the Production Effect in vocabulary learning seems most effective when tasks are sufficiently engaging to facilitate robust encoding, especially in learners with complex linguistic backgrounds.

6.2. Assessing Testing and Retention Intervals

6.2.1. Comparison of Recall and Recognition Tests in the Production Effect

The differentiation between recall and recognition in studying the production effect is vital because it can illuminate the specific aspects of memory that production enhances. Recall tasks require participants to retrieve information without the aid of cues, often tapping deeper into memory storage and relying heavily on the ability to access and reconstruct past experiences or learned information. In the context of the production effect, recall tasks can be particularly expressive, demonstrating how effectively a word or concept has been encoded through production (e.g., speaking or writing). Enhanced recall in production conditions suggests that the act of producing information makes it more distinct in memory, potentially through the creation of unique neural or cognitive traces (Krishnan et al., 2017; Candry et al., 2020; Endres et al., 2020; López Assef et al., 2023).

Recognition, conversely, presents participants with cues, requiring them to determine whether they have seen these cues before. This task relies more on familiarity and less on the ability to access detailed memory traces (McDaniel & Einstein, 1993; Cox & Shiffrin, 2017). In studies of the production effect, recognition tasks can reveal whether producing information increases the general feeling of knowing or familiarity without necessarily enhancing the ability to recall specifics without cues (Bodner et al., 2016; Zrompa et al., 2018).

This pattern is particularly evident in studies employing mixed-list designs, where the production effect is significantly more pronounced for recall than recognition within the same list (Landauer & Bjork, 1978). This suggests that production influences memory beyond a simple familiarity boost, which primarily benefits recognition. While recall typically demands deeper cognitive processing and the generation of more detailed memory traces, recognition relies more on surface-level familiarity. Therefore, the unique neural or cognitive traces created by production may facilitate the retrieval processes required for recall more than the familiarity-based processes used in recognition. This differentiation explains why a boost for recall due to production does not automatically translate to a boost for recognition, as observed in some studies. In Experiment 2, however, the reverse pattern was evident, with more significant differences emerging under recognition than recall. This suggests that, in certain contexts, production may impact familiarity-based processes more strongly, enhancing recognition rather than recall.

This distinctiveness effect, significant for recall tasks, could explain why the production effect is generally more significant for recall than recognition (MacLeod & Bodner, 2017; Rumbaugh & Landau, 2018). The distinctiveness provided by production makes the encoded information more accessible during recall tasks, which require detailed memory traces, rather than recognition tasks, which rely on familiarity.

In the current thesis, these assumptions were tested in experiments 1 and 2 to explore whether the production effect is more pronounced for recall due to its influence on distinctiveness. Utilizing both types of tasks in research on the production effect allows for a comprehensive understanding of how production influences different layers of memory. This dual approach can help delineate whether the benefits of production are more about creating distinct, durable memories or enhancing a general sense of familiarity, each of which might be more beneficial in different contexts.

The results of Experiment 1 showed that the accuracy rates in recall tasks were similar across all testing times, indicating no significant advantage of the "repeat" condition over the "listen" condition for recall tasks. This outcome suggests that producing the word aloud did not significantly enhance the ability to recall the word from memory more than merely listening to it. In contrast, for recognition tasks, while the overall accuracy improved over time for both conditions, the "listen" condition showed increasing accuracy rates at the delayed test points relative to the repeat condition. This implies that the words heard have been encoded more deeply or retained better over time, possibly due to lower cognitive load than the production condition, as discussed in the previous section.

In Experiment 2, the results show significant differences in performance based on the learning condition. For recall tasks, the 'write' condition consistently outperformed the 'listen' and 'repeat' conditions across immediate, one-week, and two-week delayed tests. Notably, children in the 'write' condition demonstrated significantly higher recall accuracy, particularly in the two-week delayed test. In contrast, the recognition tasks show that while all conditions improve performance over time, the 'repeat' condition was particularly effective immediately after learning, with children responding more quickly and accurately than in the other conditions. However, as time progressed, the 'write' condition demonstrated the highest accuracy in the two-week delayed recognition test, suggesting that writing may support retention and precise word recognition under testing conditions. This progression aligns with the notion that enhanced recall typically facilitates improved recognition, given that recognition is generally considered a less demanding cognitive task. As previously discussed, these findings imply that writing as a production activity can significantly strengthen recall and recognition of newly learned words, likely due to the additional cognitive processes engaged during writing, such as deeper semantic and phonological encoding.

Taken together, these findings highlight the distinct effects of production on recall and recognition. Writing consistently enhanced recall across all intervals, particularly over longer delays, suggesting that it strengthens memory by making information more distinctive and durable. Repeating provided an initial advantage for recognition, but writing ultimately led to the highest accuracy at delayed intervals, indicating that writing supports familiarity and distinctiveness. Overall, these results show that recall benefits most from deep, distinct memory traces, while recognition can be improved by familiarity-based processes.

6.2.2. Repeated Testing and Evaluating Vocabulary Learning:

The overall results from both experiments demonstrate improved performance in delayed tests (one-week and two-week intervals) compared to immediate tests, underscoring the efficacy of repeated testing in enhancing vocabulary learning and retention. This section emphasizes the role of repeated testing across conditions and its impact on memory consolidation in vocabulary acquisition contexts.

Research has consistently shown that repeated testing significantly enhances learning outcomes and retention, making it an essential tool for vocabulary acquisition. For instance, Karpicke and Roediger (2008) found that repeated retrieval yields stronger long-term retention than repeated studying, highlighting how retrieval, rather than mere exposure, consolidates information more effectively in memory. This effect has direct implications for vocabulary studies, where repeated retrieval of words deepens word knowledge, and fosters recall and recognition over time (e.g., Barcroft, 2004).

In Experiment 1, children's recognition accuracy showed improvement over time. In the 'listen' condition, accuracy increased from 64.0% (SD = 48.0) in the immediate test to 81.8% (SD = 38.5) in the two-week delayed test, while the 'repeat' condition improved from 65.1% (SD = 47.6) to 72.8% (SD = 44.5) over the same period. This improvement suggests that incorporating retrieval practice with spaced intervals can enhance retention by promoting deeper encoding and facilitating retrieval over extended periods. Notably, the listening condition demonstrated a more pronounced improvement, indicating that listening may contribute to more robust long-term retention in this context. Although these changes were observed, the statistical analysis indicated that they did not reach a level of significance, thus limiting the strength of conclusions about the testing effect.

In Experiment 2, the 'write' condition yielded even greater gains in accuracy over time. Recall accuracy in this condition improved from 53.1% (SD = 49.9) in the immediate test to 65.5% (SD = 47.5) in the two-week delayed test. Recognition accuracy showed a significant increase, rising from 76.6% (SD = 42.3) in the immediate test to 91.8% (SD = 27.4) in the delayed test ($p < 0.001$). These findings underscore the role of the 'write' condition in enhancing both immediate and long-term retention, supporting the testing effect principle, which posits that retrieval strengthens memory when practiced over intervals.

Writing seems to benefit from both the production effect and spaced retrieval practice, illustrating that varied engagement with vocabulary—through semantic, phonological, and orthographic encoding—enhances retention more effectively than listening or repeating alone. This pattern aligns with Butler's (2010) findings that repeated testing outperforms repeated studying, promoting not only retention but also the transfer of knowledge to new contexts, a critical aspect of vocabulary acquisition.

However, the optimal delay length for testing is a crucial aspect to consider. While some delay between learning and retrieval reinforces memory, excessive delay can lead to forgetting, reducing the benefits of repeated testing. Research suggests that one to two-week delays enhance retention and recall, but extending the interval too long without intermediate retrieval can lead to substantial memory decay (Karpicke & Roediger, 2008). In the current studies, the one-week and two-week delayed tests effectively supported memory retention, as recognition and recall scores improved over time. This suggests that appropriately spaced intervals—such as those applied in Experiment 1 and Experiment 2—are essential for consolidating vocabulary into long-term memory.

Therefore, incorporating repeated testing at appropriate intervals provides continuous memory reinforcement, making it more durable and facilitating better retention than other learning strategies. As Butler (2010) emphasized, this interval-based approach strengthens memory traces and enhances knowledge transfer, demonstrating that strategically spaced retrieval intervals are crucial for optimizing vocabulary acquisition.

In conclusion, repeated retrieval and delayed testing emerge as powerful facilitators of vocabulary learning, promoting retention and incremental acquisition. These findings have significant implications for educational practice, where implementing delayed retrieval assessments could optimize vocabulary retention and support students in achieving more robust, long-term comprehension. By strategically using spaced retrieval intervals, educators can encourage deeper processing and better knowledge retention, aligning with the testing effect principle to ensure that learned information remains accessible and applicable over time.

6.3. Practical Implications and Applications

This thesis has examined the production effect (PE) within Modern Standard Arabic (MSA) learning, shedding light on the importance of writing on L2 language acquisition among bilingual children. The findings contribute to our understanding of the production effect in language learning. This knowledge can be valuable for future research on language learning mechanisms and may ultimately inform the development of practical language instruction.

The superiority of the "write" condition observed in Experiment 3 underscores the benefits of integrating writing into language learning activities. Writing engages multiple cognitive processes, including semantic, phonological, and orthographic integration, leading to deeper cognitive processing and more durable learning outcomes. This finding is consistent with Dual Coding Theory posited by Paivio (1986), which suggests that information processed through both verbal and non-verbal channels enhances memory retention. Language instructors can incorporate writing exercises into their curriculum to enhance vocabulary retention. For instance, students can be encouraged to write sentences or short paragraphs using new vocabulary words, which will help reinforce their learning through dual coding processes. However, it is essential to note that these strategies' effectiveness depends on the student's age and developmental stage, as younger children may still need to fully develop the cognitive abilities required to benefit from these writing exercises.

Additionally, research should examine the effectiveness of combining different learning modalities in language instruction, such as auditory, visual, and kinesthetic. Understanding how these modalities interact and support each other can help educators design more comprehensive and effective language learning programs.

In conclusion, the findings of this thesis provide valuable insights into the production effect in language learning, highlighting the importance of age-appropriate instruction, task complexity, contextual cues, and the integration of writing activities. However, further research is needed to complement these findings and explore the optimal use of writing in language teaching methods. This ongoing research will ultimately contribute to developing more effective and evidence-based language instruction strategies.

6.4. Limitations and Future Directions

Besides specific limitations mentioned in experimental chapters, several general limitations of this research should be addressed in future studies. This section will provide an overview of the limitations encountered in the current thesis and propose future research directions. The focus areas will include data collection methods, participant diversity and language background, training materials, and experimental design. Recommendations for refining these aspects will be discussed to enhance the understanding of language learning processes and improve educational strategies for L2 learners.

6.4.1. Methodological Limitations

A notable limitation across all three studies is the lack of a control group to provide a baseline for direct comparison. The initial research objective was to employ eye-tracking to investigate how production influences memory and learning in a second language (L2), focusing on children's real-time processing. However, the COVID-19 pandemic led to a necessary shift in methodology from an in-lab eye-tracking setup to a laptop-based forced-choice task. This adaptation, while focusing on recognition accuracy and response time, significantly diverges from the original design.

Without a control group, the studies lack a definitive point of reference to isolate the effect of production on memory and learning outcomes, which limits the ability to conclude with certainty the extent of production’s impact. Additionally, while the current design effectively integrates recognition accuracy and response time, a more controlled baseline comparison would allow a more explicit interpretation of how production influences response times in recognition tasks. Consequently, while these findings contribute valuable insights into the relationship between production and L2 learning, the absence of a control group limits the precision with which we can attribute observed effects solely to the production component. Future studies incorporating a control group are essential for validating these findings and offering more granular insights into the unique role of production in enhancing L2 recognition and response efficiency.

6.4.2. Challenges in Data Collection with Children Populations

Several challenges were encountered in recruiting and collecting data from child participants. To engage the children and mitigate potential reluctance, the study was framed as a fun game where they could "defeat the computer." Despite this effort, convincing the children and their parents of the study's appeal proved difficult. Many children expressed boredom or hesitation, exacerbated by the frustrations of their parents, who were tasked with explaining the game and assisting without direct interference. These difficulties were compounded by practical challenges, such as accommodating various scheduling preferences, including early morning sessions. Additionally, some families expressed discomfort with instructions provided via Zoom, necessitating alternative methods of communication and support. These experiences underscore the complexities of conducting research with young participants, particularly in a remote setting during a pandemic (Appendix 6).

In Experiments 1 and 2, the procedure required approximately one hour to cover all phases, including training, learning, and testing, on the first day. This extended duration posed significant challenges in maintaining the children's attention and engagement throughout the session. The length of the session may have led to increased anxiety or stress for some participants, further complicating the data collection process. Future studies aiming to collect data with children should consider breaking the session into shorter, more manageable segments spread over multiple days to reduce fatigue and maintain engagement. Incorporating more interactive and varied tasks could help sustain children's interest. Flexibility in scheduling and using engaging, child-friendly communication tools are crucial for accommodating the needs of young participants and their families. Addressing these limitations can enhance the reliability and validity of data collected in research settings with children.

6.4.3. Language Background

The participants in the present experiments represented diverse linguistic backgrounds, a variable extensively explored in second language (L2) acquisition research. No significant differences in Modern Standard Arabic (MSA) learning emerged in the pilot study based on participants' linguistic backgrounds. However, Experiments 1 and 2 concentrated on multilingual learners whose dominant language was English, and within these experiments, knowledge of Arabic dialects did not significantly influence MSA acquisition. This outcome contrasts with previous research that emphasizes the role of related dialectal or language knowledge in shaping L2 learning processes (Jarvis & Pavlenko, 2008). In this study, while dialectal Arabic did not seem to affect MSA acquisition in this specific group, further investigation is necessary, particularly in learners for whom Arabic is the dominant language.

Future research should investigate the influence of Arabic dialects on MSA acquisition across a more comprehensive array of linguistic backgrounds. Learners who have grown up speaking an Arabic dialect may exhibit distinct acquisition patterns compared to those without such a background. The structural overlap between MSA and particular Arabic dialects is a plausible explanation for this variability. For instance, some dialects, such as Gulf Arabic and specific variants of Levantine Arabic, display substantial phonological and syntactical similarities with MSA (Benmamoun, 2000). This overlap may facilitate knowledge transfer, accelerating MSA acquisition. Conversely, dialects that differ significantly from MSA, such as North African varieties (e.g., Moroccan or Algerian Arabic), may introduce interference, complicating the acquisition of MSA due to conflicting grammatical and phonological patterns (Boussofara-Omar, 2006).

Recognizing the role of dialectal Arabic in MSA learning is crucial for designing targeted pedagogical strategies. Learners whose dialects closely align with MSA may benefit from instructional methods that build on their existing knowledge, focusing on points of convergence and reducing the need for explicit instruction of overlapping structures. Conversely, learners from dialects more divergent from MSA may require concentrated instruction on the areas of most significant disparity, particularly complex grammatical features such as verb morphology, sentence structure, and the use of definite and indefinite articles, which vary considerably across dialects.

Additionally, learners from dialectal Arabic backgrounds may differ in their exposure to formal MSA, especially in educational systems where MSA is introduced at varying stages. For example, learners from regions where MSA serves as the medium of instruction from an early age may exhibit different learning trajectories than those who primarily encounter MSA through media or in formal contexts. This factor may account for some of the variability observed in MSA learning outcomes.

Although this study focuses on MSA acquisition among specific learner groups in the UK, the generalizability of these findings could be enhanced by examining other multilingual contexts. Including L2 learners from a broader spectrum of linguistic and cultural backgrounds would provide valuable insights into how prior knowledge of structurally related languages impacts the production effect and overall language learning. For example, future research could explore how learners of European languages, such as French or Spanish, languages that share numerous cognates and phonological features, perform in production-based tasks. These similarities might facilitate vocabulary acquisition and pronunciation, thereby reducing cognitive load and enhancing memory retention (Odlin, 1989).

Moreover, studying learners immersed in Arabic-speaking countries could yield important insights into how immersion in a target language environment interacts with L2 learning processes, particularly regarding memory consolidation and the production effect. Immersion offers continuous, rich input, which has been shown to enhance phonological and syntactic accuracy in L2 learners (Doughty & Long, 2003). Through regular exposure to and use of MSA in authentic contexts, learners may internalize phonological patterns more effectively, lessening their dependence on L1 phonology and improving both pronunciation and recall of newly acquired words.

In conclusion, while the current findings did not demonstrate a significant impact of dialectal Arabic on MSA learning, this does not negate the possibility of dialectal influence under different learning conditions or among learners with varied linguistic backgrounds. Understanding how dialectal variation interacts with L2 acquisition, particularly in immersion contexts or among a more diverse range of learners, will be critical for future research. These findings also have significant implications for language pedagogy, suggesting that more individualized approaches considering learners' linguistic backgrounds may be necessary to optimize L2 learning outcomes.

6.4.4. Age of Acquisition

One limitation of the present study is the lack of consideration for the age of acquisition (AoA) factor. While age was included as a factor, the specific age at which participants began learning Modern Standard Arabic (MSA) was not considered. Research indicates that AoA can significantly impact language acquisition and production, influencing vocabulary retention and phonological development (Ellis & Morrison, 1998; Gierut & Morrisette, 2011). Including AoA as a variable could provide deeper insights into the relationship between vocabulary acquisition and learning outcomes in MSA. Future studies should incorporate a detailed language background questionnaire to gather information on when participants began learning Arabic. This addition would allow for a more nuanced analysis of how early exposure to the language affects subsequent learning and proficiency, thereby enriching our understanding of the critical factors contributing to successful language acquisition.

6.4.5. Writing and Spelling Challenges in MSA Training Materials

In experiment 2, writing was used as a production technique without focusing on spelling difficulties faced by L2 learners of Modern Standard Arabic (MSA). The assumption was that participants could easily pronounce the words based on their proficiency levels. However, it became clear that learners encountered specific challenges, such as accurately writing and spelling MSA words. Writing in MSA is challenging due to its complex script and the differences between spoken dialects and the written standard. To assess accuracy in Experiment 2, we defined correct responses as those in which participants demonstrated orthographic accuracy measured by how closely the participants' written responses matched the target words in MSA. This ensured that written components of MSA proficiency were considered when evaluating participant performance. Learners may struggle with MSA's orthographic rules, which are often more complex than those of their native languages or dialects. This gap in the training materials likely contributed to difficulties in accurately recalling and writing MSA words. Future studies should focus on spelling and orthographic rules in writing tasks for L2 learners of MSA. Training materials could include explicit instruction on common spelling patterns, practice exercises, and feedback to help learners correct errors. Phonetic training alongside writing tasks could also help learners understand the relationship between sounds and their written forms in MSA. Additionally, increasing exposure to written Arabic through reading and writing activities can improve proficiency. These activities should gradually increase complexity to build learners' confidence and competence in recognizing and producing written MSA. Addressing these challenges in future research will enhance the learning experience and outcomes for L2 learners of MSA. By refining study designs to include a comprehensive focus on spelling, orthographic rules, and phonetic training, educators can better support learners in overcoming writing difficulties, leading to improved language acquisition and retention.

Chapter 7. Conclusion

This research explored the production effect (PE) as an approach aimed at enhancing vocabulary learning. It focused on the verbal or written production of words to facilitate and deepen the learning process. The study evaluated the distinct impact of production versus non-production methods on children's second language acquisition, underscoring the necessity of adapting instructional strategies to account for differences in age and cognitive ability to optimize learning outcomes.

Learning new words in children is influenced by various factors, including the production effect on word recognition (Zamuner et al., 2017). Research on this effect presents mixed findings. Some studies have demonstrated a positive impact, notably improved memory for spoken words in children aged 5-10 (Pritchard et al., 2019; Icht & Mama, 2015), while others have reported mixed or null results (Zamuner et al., 2018; López Assef et al., 2021). These inconsistencies suggest that the effectiveness of the production effect may be moderated by factors such as age, the type of stimuli (words vs. objects), and the experimental design (e.g., mixed vs. blocked presentation of items). Although evidence for the production effect in children exists, its robustness and influence on the durability of learning remain unclear.

The mechanisms underlying word learning evolve as children grow. Early word production may precede the formation of phonological categories, suggesting a progression from holistic word learning to more structured phonological acquisition (Vihman, 2016). Furthermore, children can learn novel words through environmental cues, such as overhearing conversations (Akhtar, 2005). This demonstrates the significance of external factors in word acquisition, which further complicates the role of the production effect in isolated learning tasks.

Experiment 1 investigated the Production Effect (PE) with an expanded set of 64 Modern Standard Arabic (MSA) words, accounting for age and language background. Six-year-old bilingual children learned the words in either a listen or repeat condition. Tests were conducted immediately, one week later, and two weeks later. Although delayed tests showed overall improvement across both groups, no significant production effect was observed.

Experiment 2 extended the investigation by incorporating a writing condition, comparing it to listening and repeating. The results revealed that the "write" condition yielded superior performance, particularly in delayed recall tests. Writing appears to engage more comprehensive cognitive processes, including semantic, phonological, and orthographic integration, aligning with Paivio's Dual Coding Theory (1986). According to this theory, cognition involves two distinct channels for processing information—verbal and non-verbal—and writing activates both channels simultaneously. This multimodal engagement strengthens memory retention by providing multiple retrieval pathways, which enhance the durability of learning over time, particularly in delayed tests.

In the context of language learning, the integration of visual (orthographic) and motoric (writing) tasks observed in the "write" condition supports deeper cognitive processing, leading to more durable learning outcomes. The multimodal nature of writing aligns with the theoretical framework of Dual Coding Theory, which posits that engaging verbal and non-verbal systems creates a more prosperous and interconnected cognitive representation. This finding suggests that writing not only aids in immediate learning but also significantly contributes to long-term retention, further supporting Paivio's theory that multiple cognitive pathways enhance memory resilience.

Taken together, these findings demonstrate that the production effect (PE) is influenced by a confluence of factors, including cognitive ability, task complexity, and the contextual environment in which learning occurs. Age and cognitive development play pivotal roles in determining how effectively production enhances learning, with older children demonstrating better memory retention and a greater capacity to handle more complex tasks. Additionally, the type of stimuli (words vs. objects) and experimental design (mixed vs. blocked presentation of items) moderate the efficacy of PE. Contextual factors, such as the clarity of definitions, the visual context, and the relevance of words to prior knowledge, also significantly impact word acquisition. Writing, in particular, strengthens memory retrieval, leading to more comprehensive vocabulary retention through a multifaceted approach.

In light of these findings, several recommendations can be made. First, educators should incorporate active learning strategies, including interactive activities, multimedia integration, and experiential learning, into vocabulary instruction. These strategies cater to various learning styles and enhance the retention and retrieval of new vocabulary. Introducing desirable difficulties, such as writing exercises, can also improve retention, particularly for older children. Managing cognitive load is critical to ensure that students are not overwhelmed by task demands, and the use of writing as a learning tool should be encouraged, as it engages multiple cognitive processes that support long-term retention. Finally, instructional approaches should be tailored to children's developmental stages, ensuring that learning strategies are aligned with their cognitive and linguistic capacities.

To conclude, this research highlights the complexity of the production effect in vocabulary acquisition, emphasizing the role of multiple cognitive and contextual factors. Future research should explore how these factors interact in diverse learning contexts to refine educational practices further and advance our understanding of cognitive processes in language learning.

References

Abdulhameed, A. (2023). رسمياً.. 'ترند' و"تروبقة" تدخلان قاموس اللغة العربية, *Alarabiya*. https://www.alarabiya.net/culture-and-art/2023/

Abdulrahaman, M. D., Faruk, N., Oloyede, A. A., Surajudeen-Bakinde, N. T., Olawoyin, L. A., Mejabi, O. V., Imam-Fulani, Y. O., Fahm, A. O., & Azeez, A. L. (2020). Multimedia tools in the teaching and learning processes: A systematic review. *Heliyon*, *6*(11). https://doi.org/10.1016/j.heliyon.2020.e05312

Abello-Contesse, C. (2009). Age and the critical period hypothesis. *ELT journal*, *63*(2), 170-172.

Abu-Rabia, S. (2000). Effects of exposure to literary Arabic on reading comprehension in a diglossic situation. *Reading and Writing*, *13*(1-2), 147-157.

Abuhakema, G. (2012). Heritage and non-heritage language learners in Arabic classrooms: Inter and intra-group beliefs, attitudes and perceptions. *Journal of the National Council of Less Commonly Taught Languages*, *12*, 73–106. https://doaj.org/article/f241a67fde724e5094f5e43e52b5d8d8

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50–68. https://doi.org/10.1037/a0014411

Adams, G. F. (2014). *Supporting language access: Teaching talk without words*. University of California.

Adwani, P., & Shrivastava, S. (2017). Analysis of factors affecting second language acquisition. *International Journal of Social Sciences and Management*, *4*(3), 158–164. https://doi.org/10.3126/ijssm.v4i3.17247

Ahrabi, M. F., Borzabadi, D. F., & Khomeijani, A. A. F. (2020). Incidental vocabulary learning and retention from audiovisual input and factors affecting them. *English Teaching & Learning*, *45*(2), 167–188. https://doi.org/10.1007/s42321-020-00066-y

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*(2), 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Al-Ani, S. H. (2014). *Arabic phonology: An acoustical and physiological investigation* (Vol. 61). Walter de Gruyter.

Al-Khatib, M. A., & Sabbah, E. H. (2008). Language choice in mobile text messages among Jordanian university students. *SKY Journal of Linguistics*, *21*, 37-65.

Al-Sahafi, M. (2015). The role of Arab fathers in heritage language maintenance in New Zealand. *International Journal of English Linguistics*, *5*(1), 73–83. https://doi.org/10.5539/ijel.v5n1p73

Al-Sughayer, K. I. (1990). Aspects of comparative Jordanian and modern standard Arabic phonology [Doctoral dissertation, Michigan State University].Holes, C. (2004). Modern Arabic: Structures, functions, and varieties (Revised ed.). Georgetown University Press.Laks, L., & Berman, R. A. (2014). A new look at diglossia: Modality-driven distinctions between spoken and written narratives in Jordanian Arabic. Language Learning, 64(4), 819-849.Al-Omari, A. (2019). The phonology of Jordanian Arabic: Stratal or parallel OT? *Journal of Universal Language*, *20*(1), 1-43.

Alabd, A. (2016). *Heritage language learners in L2 Arabic classes: Challenges and instructional strategies* [Master's thesis, The American University in Cairo]. AUC Knowledge Fountain. https://fount.aucegypt.edu/etds/304

Aladdin, A. (2010). Non-Muslim Malaysian learners of Arabic (NMMLAs): An investigation of their attitudes and motivation towards learning Arabic as a foreign language in multiethnic and multicultural Malaysia. *Procedia-Social and Behavioral Sciences*, *9*, 1805-1811.

Alavi, M. and Leidner, D. (2001). Review: knowledge management and knowledge management systems: conceptual foundations and research issues. *Mis Quarterly*, *25*(1), 107. https://doi.org/10.2307/3250961

Albirini, A. (2014). The role of the colloquial varieties in the acquisition of the standard variety: The case of Arabic heritage speakers. *Foreign Language Annals*, *47*(3), 447–463. https://doi.org/10.1111/flan.12087

Albirini, A. (2016). *Modern Arabic sociolinguistics: Diglossia, variation, codeswitching, attitudes and identity*. Routledge.

Albirini, A. (2017). The acquisition of Arabic as a first language. In E. Benmamoun & R. Bassiouney (Eds.), *The routledge handbook of arabic linguistics* (pp. 227–248). Routledge. https://doi.org/10.4324/9781315147062

Alcock, K. J., Ngorosho, D., Deus, C., & Jukes, M. C. (2010). We don't have language at our house: disentangling the relationship between phonological awareness, schooling, and literacy. *British Journal of Educational Psychology*, *80*(1), 55-76.

Aldawood, Z., Hand, L., & Ballard, E. (2023). Language learning environments for Arabic-speaking children in New Zealand: Family demographics and children's Arabic language exposure. *Speech, Language and Hearing*, *26*(4), 1–12. https://doi.org/10.1080/2050571x.2023.2212537

Alemi, M., & Tayebi, A. (2011). The influence of incidental and intentional vocabulary acquisition and vocabulary strategy use on learning L2 vocabularies. *Journal of Language Teaching and Research*, *2*(1), 81.

Alhamdan, A. A., Murphy, M. J., & Crewther, S. G. (2022). Age-related decrease in motor contribution to multisensory reaction times in primary school children. *Frontiers in Human Neuroscience*, *16*, 967081. https://doi.org/10.3389/fnhum.2022.967081

Aljadani, A. S. (2020). The Influence of Frequency on the Acquisition and Textbooks Design of Second Language Vocabulary. *International Journal of Linguistics.* https://doi. org/10.5296/ijl. v12i3,16831

Allen, J. (1999). *Words, words, words: Teaching vocabulary in grades 4-12*. Stenhouse Publishers.

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, *135*(2), 298–313. https://doi.org/10.1037/0096-3445.135.2.298

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419-439.

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of experimental child psychology*, *87*(2), 85-106.

Alresaini, S. (2017). Acquisition of Modern Standard Arabic by Speakers of Different Arabic Colloquial Varieties: Resumption in Object Relative Clauses. *Arab World English Journal (AWEJ) Volume*, *7*.

Alvarez, E., & Morgenstern, A. (2024). Third-Generation heritage Spanish acquisition and socialization: Word learning and overheard input in an LA-based Mexican Family. *Languages*, *9*(3), 108.

Alzu'bi, A. et al. (2023). Acquisition of modern Standard Arabic by speakers of different Arabic dialects: A linguistic study. *Journal of Arabic Linguistics*, *35*(2), 112–135.

Amin, H. U., & Malik, A. S. (2014). Memory retention and recall process. In N. Kamel & K. N. Malik (Eds.), *EEG/ERP analysis* (pp. 219–237). CRC Press. https://doi.org/10.1201/b17605

Andari, I. A. M. Y. (2023). Factors influence The acquisition of vocabulary by young learners. *Kumarottama: Jurnal Pendidikan Anak Usia Dini*, 153-166.

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & Kriegstein, K. (2020). Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. *Educational Psychology Review*, *32*(3), 815–850. https://doi.org/10.1007/s10648-020-09527-z

Andreou, G., & Karapetsas, A. (2004). Verbal abilities in low and highly proficient bilinguals. *Journal of Psycholinguistic Research*, *33*(5), 357–364. https://doi.org/10.1023/b:jopr.0000039545.16783.61

Anthony, J. L., Williams, J. M., Aghara, R. G., Dunkelberger, M., Novak, B., & Mukherjee, A. D. (2010). Assessment of individual differences in phonological representation. *Reading and Writing*, *23*(8), 969–994. https://doi.org/10.1007/s11145-009-9185-7

Antony JW, Ferreira CS, Norman KA, Wimber M. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends Cogn Sci*, *21*(8):573-576. https://doi.org/10.1016/j.tics.2017.05.001

Arnedt, C. S., & Gentile, J. R. (1986). A test of dual coding theory for bilingual memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *40*(3), 290–299. https://doi.org/10.1037/h0080100

Ashby FG, Maddox WT. (2011). Human category learning 2.0. *Ann N Y Acad Sci*. 1224:147-161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 1–36). Elsevier Academic Press. https://doi.org/10.1016/S0079-7421(06)46001-1

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442–481.

Ashby, J. (2003). *Phonological structure in spoken and written language: Evidence for a shared representation*. Scholar works.

Atasheneh, N., and Naeimi, M. (2015). Vocabulary learning through using mechanical techniques vocabulary learning strategy. *Theory and Practice in Language Studies*, *5*(3), 541–553. https://doi.org/10.17507/tpls.0503.13

Athari, P. (2021). *Infant speech development in the context of mother-infant interaction in transition from the precanonical to the canonical babbling stage*. McGill University .

Atkinson, R. C. (1986). Human memory and learning: Progress and prospects. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 3-19). North-Holland.

Atkinson, R. C. (2011). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In W. K. Estes, S. S. Stevens, & A. E. Ratliff (Eds.), *Advances in psychological theory* (Vol, 2, pp. 289–356). Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1977). Human memory: A proposed system and its control processes. *Human Memory*, 7–113. https://doi.org/10.1016/b978-0-12-121050-2.50006-5

Azizian, A., & Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *Journal of Cognitive Neuroscience*, *19*(12), 2071-2081.

Baddeley, A. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829-839.

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford University Press.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. (2020). Working memory. In A. Baddeley, M. W. Eysenck, & M. C. Anderson (Eds.), *Memory*. Routledge. https://doi.org/10.4324/9780429449642

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362-365. https://doi.org/10.1080/14640746608400055

Baddeley, A. D. (1985). *Working memory*. Oxford University Press.

Baddeley, A. D. (1997). Human memory: Theory and practice. Psychology Press.

Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77–92). Oxford University Press.

Baddeley, A. D. (2002). Working memory: An overview. In H. L. Pashler (Ed.), *Stevens' handbook of experimental psychology, Vol. 3: Learning, motivation, and emotion* (pp. 475–499). John Wiley & Sons.

Baddeley, A. D., & Hitch, J. G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). Academic Press.

Baddeley, A. D., & Lieberman, K. (2017). Spatial working memory. In A. Baddeley (Ed.), *Exploring working memory* (pp. 206–223). Routledge.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and* *executive control* (pp. 28–61). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005

Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, *40*(10), 1737–1743. https://doi.org/10.1016/S0028-3932(01)00146-4

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*(4), 981–1005. https://doi.org/10.3758/s13414-019-01725-4

Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36. https://doi.org/10.1016/j.jml.2015.10.008

Bahhari, A. (2020). Arabic language maintenance amongst sojourning families in Australia. *Journal of Multilingual and Multicultural Development*, *44*(5), 1–13. https://doi.org/10.1080/01434632.2020.1829631

Bai, Z. (2018). An analysis of English vocabulary learning strategies. *Journal of Language Teaching and Research*, *9*(4), 849–855. https://doi.org/10.17507/jltr.0904.24

Bailey, H. R., MacLeod, C. M., & Kregiel, B. (2021). The production effect in memory: A prominent mnemonic in children. Journal of Child Language, 42(5), 1102-1124.

Bakker, I., Takashima, A., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*, 116-130.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340-357.

Baluška, F., & Levin, M. (2016). On having no head: Cognition throughout biological systems. *Frontiers in Psychology*, *7*, 902. https://doi.org/10.3389/fpsyg.2016.00902

Bandura, A. (1969). Social-learning theory of identificatory processes. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 213-262). Rand McNally & Company.

Bandura, A., & Evans, R. I. (2006). *Albert Bandura*. Insight Media.

Barachetti, C., Majorano, M., Rossi, G., Antolini, E., Zerbato, R., & Lavelli, M. (2021). Vocabulary production in toddlers from low-income immigrant families: Evidence from children exposed to romanian-italian and nigerian english-italian. *Journal of Child Language*, *49*(2), 408–421. https://doi.org/10.1017/s0305000921000222

Barcroft, J. (2004). Second language vocabulary acquisition: A framework for effective teaching. The Modern Language Journal, 88(2), 222-243.

Barlow, M. C. (1928). The r?le of articulation in memorizing. *Journal of Experimental Psychology*, *11*(4), 306–312. https://doi.org/10.1037/h0071449

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570-585. https://doi.org/10.1037/0278-7393.33.3.570

Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, *36*(6), 1129–1147. https://doi.org/10.1111/j.1551-6709.2012.01243.x

Barton, B., & Brewer, A. A. (2020). Attention and working memory in human auditory cortex. In S. Hatzopoulos, A. Ciorba, & P. H. Skarzynski (Eds.), *The human auditory system-basic features and updates on audiological diagnosis and therapy* (pp. 3–35). IntechOpen. https://doi.org/10.5772/intechopen.85537

Barua, S. (2023). The interplay between implicit instruction and second language acquisition. *International Journal of Academic Pedagogical Research (IJAPR)*, *7*(7), 9-21. http://ijeais.org/wp-content/uploads/2023/7/abs/IJAPR230703.html

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bayley, P. J., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/no recognition, forced-choice recognition, and the human hippocampus. *Journal of Cognitive Neuroscience*, *20*(3), 505-512.

Bédécarrats, A., Chen, S., Pearce, K., Cai, D., & Glanzman, D. L. (2018). RNA from trained aplysia can induce an epigenetic engram for long-term sensitization in untrained aplysia. *Eneuro*, *5*(3), 1–11. https://doi.org/10.1523/eneuro.0038-18.2018

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 553.

Benitez, V. L., Yurovsky, D., & Smith, L. B. (2016). Competition between multiple words for a referent in cross-situational word learning. *Journal of Memory and Language*, *90*, 31–48. https://doi.org/10.1016/j.jml.2016.03.004

Bennett, H., Denston, A., & Arrow, A. (2023). The effectiveness of a parent-implemented, phonological awareness programme on the phonological awareness skills of preschool children. *The Australian Journal of Language and Literacy*, *46*, 125–143. https://doi.org/10.1007/s44020-023-00034-6

Berghe, R., Oudgenoeg-Paz, O., Verhagen, J., Brouwer, S., de Haas, M., de Wit, J., Willemsen, B., Vogt, P., Krahmer, E., & Leseman, P. (2021). Individual Differences in Children's (Language) Learning Skills Moderate Effects of Robot-Assisted Second Language Learning. *Frontiers in Robotics and AI*, *8*. https://doi.org/10.3389/frobt.2021.676248

Bernardi, M., Leonard, H. C., Hill, E. L., & Henry, L. A. (2016). Brief report: Response inhibition and processing speed in children with motor difficulties and developmental coordination disorder. *Child Neuropsychology*, *22*(5), 627–634. https://doi.org/10.1080/09297049.2015.1014898

Bertels, J., Kolinsky, R., Coucke, D., & Morais, J. (2015). When the spatial location of a novel word matters: Evidence for the use of orthographic cues in word learning. Language, *Cognition and Neuroscience*, *30*(1-2), 127-139.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & cognition*, *35*, 201-210.

Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. (2016). Articulating what infants attune to in native speech. *Ecological Psychology*, *28*(4), 216–261. https://doi.org/10.1080/10407413.2016.1230372

Beyon, J., & Cox-Boyd, C. (2020). The influence of font type and color on online information recall. *North American Journal of Psychology*, *22*(1). https://openurl.ebsco.com/openurl?sid=ebsco:plink:scholar&id=ebsco:gcd:141816959&crl=c

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, *19*(1), 19–23. https://doi.org/10.1177/0963721409358571

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in cognitive sciences*, *16*(4), 240-250.

Bisson, M. J., van Heuven, W. J., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology*, *68*(7), 1306-1326.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher et al. (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). Worth Publishers.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). MIT Press.

Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, 56-64.

Blankenship, T. L., Keith, K., Calkins, S. D., & Bell, M. A. (2018). Behavioral performance and neural areas associated with memory processes contribute to math and reading achievement in 6-year-old children. *Cognitive Development*, *45*, 141–151. https://doi.org/10.1016/j.cogdev.2017.07.002

Blaxton, T. A. (1989). An investigation of the multiple forms of memory used in conceptually driven and data-driven reading tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 657–668.

Block, R. A. (2003). Psychological timing without a timer: The roles of attention and memory. In H. Helfrich (Ed.), *Time and mind II: Information processing perspectives* (pp. 41-59). Hogrefe & Huber Publishers.

Blunt, J. R., & VanArsdall, J. E. (2021). Animacy and animate imagery improve retention in the method of loci among novice users. *Memory & Cognition*, *49*(7), 1360–1369. https://doi.org/10.3758/s13421-021-01175-0

Bock, K., Levelt, W., & Gernsbacher, M. A. (2002). Language production: Grammatical encoding. In G. M. Altman (Ed.), *Psycholinguistics: Critical concepts in psychology* (pp. 405–452). Academic Press.

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1711–1719. https://doi.org/10.1037/a0028466

Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, D.-L., & Bernstein, D. M. (2016). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology*, *70*(2), 93–98. https://doi.org/10.1037/cep0000082

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review*, *21*, 149-154.

Boers, F. (2013). Cognitive linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching*, *46*(2), 208–224. https://doi.org/10.1017/s0261444811000450

Boers, F., & Lindstromberg, S. (2008). How cognitive linguistics can foster effective vocabulary teaching. *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*, *1*(1), 1–64. https://doi.org/10.1515/9783110199161.1.1

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529. https://doi.org/10.1162/jocn.2009.21306

Boll-Avetisyan, N. (2012). Probabilistic phonotactics in lexical acquisition: The role of syllable complexity. In P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, & B. Kühnert (Eds.), *Consonant clusters and structural complexity* (pp. 257–272). De Gruyter. https://doi.org/10.1515/9781614510772

Bosch, L., and Segers, E. (2020). First and second language vocabulary affect early second language reading comprehension development. *Journal of Research in Reading*, *43*(3), 290–308. https://doi.org/10.1111/1467-9817.12304

Bosma, E., Heeringa, W., Hoekstra, E., Versloot, A., & Blom, E. (2017). Verbal working memory is related to the acquisition of cross-linguistic phonological regularities. *Frontiers in Psychology*, *8*, 1487. https://doi.org/10.3389/fpsyg.2017.01487

Botvinick, M. M., Wang, J., Cowan, E., Roy, S., Bastianen, C., Patrick Mayo, J., & Houk, J. C. (2009). An analysis of immediate serial recall performance in a macaque. *Animal Cognition*, *12*, 671-678.

Boufares, L. (2019). *Investigating learners' retention of lexical items through the use of visual aids* [Doctoral dissertation, Université Ibn Khaldoun-Tiaret-]. ProQuest Dissertation Publishing. http://dspace.univ-tiaret.dz/handle/123456789/936

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*(3), 219–231. https://doi.org/10.1037/0097-7403.20.3.219

Bozon, B., Davis, S., & Laroche, S. (2003). A requirement for the immediate early gene zif268 in reconsolidation of recognition memory after retrieval. *Neuron*, *40*(4), 695–701. https://doi.org/10.1016/s0896-6273(03)00674-3

Bradley, C., & Pearson, J. (2012). The sensory components of high-capacity iconic memory and visual working memory. *Frontiers in Psychology*, *3*, 355. https://doi.org/10.3389/fpsyg.2012.00355

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Breen, E., Pomper, R., & Saffran, J. (2019). Phonological learning influences label–object mapping in toddlers. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1923-1932.

Buchsbaum, B. R. (2016). Working memory and language. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 863-875). Academic Press.

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*(3), 433–461. https://doi.org/10.1017/s0272263111000040

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174-205). Cambridege University Press.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133. https://doi.org/10.1037/a0019902

Butler, Y. G. (2012). Bilingualism/multilingualism and second-language acquisition. In T. K. Bhatia & W. C. Ritchai (Eds.), *The handbook of bilingualism and multilingualism* (pp. 109-136). Routledge.

Byrne, J. H. (2017). *Learning and memory: a comprehensive reference*. Academic Press.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31-42.

Camos, V., & Portrat, S. (2014). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review*, *22*(4), 1029–1034. https://doi.org/10.3758/s13423-014-0772-5

Candry, S., Decloedt, J., & Eyckmans, J. (2020). Comparing the merits of word writing and retrieval practice for L2 vocabulary learning. *System*, *89*, 102206. https://doi.org/10.1016/j.system.2020.102206

Carpenter, J. (2011) Impacts of GM Crops on Biodiversity. *GM Crops, 2*, 7-23. http://dx.doi.org/10.4161/gmcr.2.1.15086

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268-276.

Carpenter, S. K., & Geller, J. (2019). Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. *Quarterly Journal of Experimental Psychology*, *73*(2), 211–224. https://doi.org/10.1177/1747021819879416

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 92–101. https://doi.org/10.1037/a0024828

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, *13*(5), 826–830. https://doi.org/10.3758/bf03194004

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*(6), 633–642. https://doi.org/10.3758/bf03202713

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354-380.

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology*, *9*(1), 57. https://doi.org/10.3389/fpsyg.2018.00401

Chang, Y., Delaney, P. F., & Verkoeijen, P. P. (2019). The testing effect in immediate recognition: Tests of the episodic context account. *Journal of Cognitive Psychology*, *31*(8), 825-838.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22. https://doi.org/10.1016/s1364-6613(02)00005-0

Chen Q, Mirman D. (2015). Interaction between phonological and semantic representations: time matters. *Cogn Sci*; 39(3):538-58. https://doi.org/10.1111/cogs.12156

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*(5), 693-713.

Chen, H. C., Vaid, J., Bortfeld, H., & Boas, D. A. (2008). Optical imaging of phonological processing in two distinct orthographies. *Experimental Brain Research*, *184*(3), 427-433.

Chen, J., Zhang, L., & Qian, W. (2022). Cognitive differences between readers attentive and inattentive to task-related information: An eye-tracking study. *Aslib Journal of Information Management*, *75*(5), 917–939. https://doi.org/10.1108/ajim-01-2022-0007

Chen, O., Castro-Alonso, J., Paas, F., & Sweller, J. (2018). Undesirable difficulty effects in the learning of high-element interactivity materials. *Frontiers in Psychology*, *9*, 1483. https://doi.org/10.3389/fpsyg.2018.01483

Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research*, *1665*, 50–64. https://doi.org/10.1016/j.brainres.2017.03.023

Chen, Q., & Mirman, D. (2015). Interaction between phonological and semantic information in spoken word recognition: Insights from eye-tracking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1297-1312.

Cheung, S., Kan, P., Winicour, E., & Yang, J. (2018). Effects of home language input on the vocabulary knowledge of sequential bilingual children. *Bilingualism Language and Cognition*, *22*(5), 986–1004. https://doi.org/10.1017/s1366728918000810

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Cho, K. W., & Feldman, L. B. (2013). Production and accent affect memory. *Phonological and Phonetic Considerations of Lexical Processing*, *8*(3), 295–319. https://doi.org/10.1075/ml.8.3.02cho

Cho, K. W., & Feldman, L. B. (2016). When repeating aloud enhances episodic memory for spoken words: Interactions between production- and perception-derived variability. *Journal of Cognitive Psychology*, *28*(6), 673–683. https://doi.org/10.1080/20445911.2016.1182173

Chomsky, N. (2011). Language and other cognitive systems. what is special about language? *Language Learning and Development*, *7*(4), 263–278. https://doi.org/10.1080/15475441.2011.584041

Church, B. A., Jackson, B. N., & Smith, J. D. (2021). Exploring explicit learning strategies: A dissociative framework for research. *New ideas in psychology*, *60*, 100817

Church, K., & Liberman, M. (2021). The future of computational linguistics: On beyond alchemy. *Frontiers in Artificial Intelligence*, *4*, 625341.

Cleeremans, A., & Jiménez, L. (2013). Implicit learning and consciousness: A graded, dynamic perspective. In L. Jimenez (Eds.), *Implicit learning and consciousness* (pp. 1-40). Psychology Press.

Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology*, *1*(1), 65–92. https://doi.org/10.1515/labphon.2010.005

Cockcroft, K. (2015). The role of working memory in childhood education: Five questions and answers. *South African Journal of Childhood Education*, *5*(1), 01-20. https://www.scielo.org.za/scielo.php?pid=S2223-76822015000100004&script=sci_arttext

Cohen, A. D. (2014). *Strategies in learning and using a second language*. Routledge.

Collins, A. G. E., & Frank, M. J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences*, *115*(10), 2502–2507. https://doi.org/10.1073/pnas.1720963115

Collins, A. G. E., Ciullo, B., Frank, M. J., & Badre, D. (2017). Working memory load strengthens reward prediction errors. *The Journal of Neuroscience*, *37*(16), 4332–4342. https://doi.org/10.1523/jneurosci.2700-16.2017

Conway, A., Jarrold, C., Kane, M., Miyake, A., & Towse, J. (Eds.). (2007). *Variation in working memory*. Oxford University Press.

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*(3), 341–361. https://doi.org/10.1016/0749-596x(87)90118-5

Cook, J. L., & Cook, J. L. (2010). *The world of children*. Allyn & Bacon.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, *33*(2), 185-209.

Cook, V. (Ed.). (2002). Portraits of the L2 user. Multilingual Matters.

Cook, V. (Ed.). (2003). Effects of the second language on the first. Multilingual Matters.

Cook, V. J. (1991). The poverty-of-the-stimulus argument and multicompetence. Second Language Research, 7(2), 103-117.

Cook, V. J., & Bassetti, B. (2010). *Language and bilingual cognition*. Psychology Press.

Cook, V., & Wei, L. (Eds.). (2016). The Cambridge handbook of linguistic multi-competence. Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. https://doi.org/10.1017/s0140525x01003922

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/s0079-6123(07)00020-9

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57.

Cowan, N. (2012). *Working memory capacity*. Routledge, Taylor & Francis Group.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223. https://doi.org/10.1007/s10648-013-9246-y

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science*, *14*(5), 1066–1074. https://doi.org/10.1111/j.1467-7687.2011.01060.x

Cowan, N., Lichty, W., & Grove, T. R. (1990). Properties of memory for unattended spoken syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 258.

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. https://doi.org/10.1037/rev0000076

Coyne, M. D., McCoach, D. B., Ware, S., Austin, C. R., Loftus-Rattan, S. M., & Baker, D. L. (2019). Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children*, *85*(2), 163–179. https://doi.org/10.1177/0014402918789162

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/s0022-5371(72)80001-x

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159–180. https://doi.org/10.1037/0096-3445.125.2.159

Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences*, *103*(24), 9315-9320.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*(6), 365-373.

Crystal, J. D., & Wilson, A. G. (2015). Prospective memory: A comparative perspective. *Behavioural Processes*, *112*, 88–99. https://doi.org/10.1016/j.beproc.2014.07.016

Csizér, K. (2009). Learning experiences, selves and motivated learning behaviour: A comparative analysis of structural models for Hungarian secondary and university learners of English. *Motivation, language identity and the L2 self/Multilingual Matters*.

Cummings, A. G., Piper, R. E., & Pittman, R. T. (2018). The importance of teaching vocabulary: The whys and hows. *An Online Journal for Literacy Educators*, *4*(6). https://digitalcommons.tamusa.edu/read_faculty/4/

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, *14*(2), 175–187. https://doi.org/10.2307/3586312

Cummins, J. (2009). Bilingual and immersion programs. *The handbook of language teaching*, 159-181.

Cummins, J. (2017). Bilingual education and language competence: Cognitive considerations. *Journal of Educational Psychology*, *109*(8), 987–996.

Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, *137*, 57–75. https://doi.org/10.1016/j.jecp.2015.03.005

Cyr, A., MacLeod, C. M., & Saint-Aubin, J. (2021). The Production Effect interacts with serial positions. Psychological Bulletin, 147(5), 467-486.

Czamolewski, M. (1996). An empirical validation for the natural log transformation of reaction time. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting/Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting*, *40*(24), 1287–1287. https://doi.org/10.1177/154193129604002471

Daana, H. A. (2017). Phonological development of an Arabic-English bilingual child during the one-word stage. *Linguistics and Literature Studies*, *5*(5), 354–364. https://doi.org/10.13189/lls.2017.050504

Dagenbach, D., & Carr, T. H. (Eds.). (1994). *Inhibitory processes in attention, memory, and language* (pp. 189-239). Academic Press.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507–534. https://doi.org/10.1080/01690960143000074

Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal*, *97*(4), 902–916. https://doi.org/10.1111/j.1540-4781.2013.12045.x

Dajani, B. A. S., Mubaideen, S., & Omari, F. M. A. (2014). Difficulties of Learning Arabic for non-native speakers. *Procedia-Social and Behavioral Sciences, 114*, 919-926. https://doi.org/10.1016/j.sbspro.2013.12.808

Dang, J., King, K. M., & Inzlicht, M. (2021). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, *25*(2), 137-143.

Daniels, M. (2000). *Dancing with words: Signing for hearing children's literacy*. Bloomsbury Publishing.

Danilina, S., & Shabunina, V. (2020). Intentional vs incidental ESP vocabulary acquisition by political science students. *Advanced Education*, *7*(16), 18–27. https://doi.org/10.20535/2410-8286.200729

Dasí, C., Soler, M. J., Cervera, T., & Ruiz, J. C. (2008). Influence of articulation rate on two memory tasks in young and older adults. *Perceptual and motor skills*, *106*(2), 579-589.

Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech, Language, and Hearing Research*, *38*(6), 1199–1211. https://doi.org/10.1044/jshr.3806.1199

Davis, M. H. (2003). Connectionist modelling of lexical segmentation and vocabulary acquisition. In P. Quinlan (Ed.), Connectionist models of development: Developmental processes in real and artificial neural networks (pp. 151-187). Psychology Press.

de Boysson-Bardies, B. (2001). *How language comes to children: From birth to two years*. Mit Press.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*, 453-482.

DeCaro, M. S., & Maricle, D. E. (2011). Working memory. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and development* (pp. 1579–1581). Springer. https://doi.org/10.1007/978-0-387-79061-9_3100

Deeb, A. (2017). *Learning vocabulary through memorization* [Doctoral dissertation, Lebanese American University]. https://doi.org/10.26756/th.2017.26

Dehn, M. J. (2011). *Working memory and academic learning: Assessment and intervention*. John Wiley & Sons.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), Focus on Form in Classroom Second Language Acquisition. Cambridge: Cambridge University Press.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. *Focus on form in classroom second language acquisition*, *28*, 42-63.

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. *The handbook of language teaching*, 119-138.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893–1906. https://doi.org/10.1162/jocn.2008.21142

Dewaele, J. M. (2015). Bilingualism and multilingualism. In K. Trach, C. ilie, T. Sandel (Eds.), *The international encyclopedia of language and social interaction* (pp. 1-11). Jhon Wiley & Sons. https://doi.org/10.1002/9781118611463/wbielsi108

Dewhurst, S. A. (1999). Cognitive effort and recollective experience in recognition memory. *Memory*, *7*(2), 129–146. https://doi.org/10.1080/741944067

Dey, A., & Sommers, M. S. (2015). Age-related differences in inhibitory control predict audiovisual speech perception. *Psychology and Aging*, *30*(3), 634–646. https://doi.org/10.1037/pag0000033

Dhahir, O. (2015). Studying Arabic as an additional language together with Arab heritage Language learners: The intellectual aspects of sociocultural interactive strategies. *Al-'Arabiyya*, *48*, 43–59. https://www.jstor.org/stable/44654038

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/s1366728902003012

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Frontiers in Human Neuroscience*, *11*, 233. https://doi.org/10.3389/fnhum.2017.00233

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

Dixon, L. Q., Zhao, J., Shin, J.-Y., Wu, S., Su, J.-H., Burgess-Brigham, R., Gezer, M. U., & Snow, C. (2012). What we know about second language acquisition: A synthesis from four perspectives. *Review of Educational Research*, *82*(1), 5–60. https://doi.org/10.3102/0034654311433587

Dochy, F. J. R. C., & Alexander, P. A. (1995). Mapping prior knowledge: A framework for discussion among researchers. European *Journal of Psychology of Education*, *10*(3), 225-242.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*(1), 155–161. https://doi.org/10.3758/bf03196152

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473. https://doi.org/10.1111/1467-9280.00290

Dudukovic, N. M., DuBrow, S., & Wagner, A. D. (2009). Attention during memory retrieval enhances future remembering. *Memory & Cognition*, *37*(7), 953–961. https://doi.org/10.3758/mc.37.7.953

Dufour, S., & Nguyen, N. (2016). Does talker-specific information influence lexical competition? Evidence from phonological priming. *Cognitive Science*, *41*(8), 2221–2233. https://doi.org/10.1111/cogs.12472

Dufour, S., & Peereman, R. (2003). Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition*, *31*(8), 1271–1283. https://doi.org/10.3758/bf03195810

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Dunn, K., & Iwaniec, J. (2021). Exploring the relationship between second language learning motivation and proficiency. *Studies in Second Language Acquisition*, *44*(4), 1–31. https://doi.org/10.1017/s0272263121000759

Durlik, J., Szewczyk, J., Muszyński, M., & Wodniecka, Z. (2016). Interference and inhibition in bilingual language comprehension: Evidence from polish-english interlingual homographs. *PLOS ONE*, *11*(3), e0151430. https://doi.org/10.1371/journal.pone.0151430

Durrant, P., Siyanova-Chanturia, A., Kremmel, B., & Sonbul, S. (2022). *Research methods in vocabulary studies*. John Benjamins.

Dyson, A. T., & Amayreh, M. M. (2000). Phonological errors and sound changes in Arabic-speaking children. *Clinical Linguistics & Phonetics*, *14*(2), 79-109.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie* [Memory: A Contribution to Experimental Psychology]. Leipzig, Germany: Duncker & Humblot.

Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of Neurosciences*, *20*(4), 155–156. https://doi.org/10.5214/ans.0972.7531.200408

Eghbaria-Ghanamah, H., Ghanamah, R., Shalhoub-Awwad, Y., Adi-Japha, E., & Karni, A. (2021). Long-term benefits after a rhyme-repetition based intervention program for kindergarteners: Better reading and spelling in the first grade. *Developmental Psychology*, *58*(2), 252–269. https://doi.org/10.1037/dev0001284

Eisenbeiß, S. (2005). Documenting child language. *Language documentation and description*, 3.

Eisenbeiss, S. (2009). Generative approaches to language learning.

Eisenbeiss, S. (2010). Talking to children: The role of child-directed speech in language development.

Ekstrand, B. R. (1966). Backward associations. *Psychological Bulletin*, *65*(1), 50.

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. (1966). A frequency theory of verbal-discrimination learning. *Psychological Review*, *73*(6), 566.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*(2), 515–523. https://doi.org/10.1037//0278-7393.24.2.515

Ellis, N. C. (2001). *Memory for language* (pp. 33-68). na.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*, (pp. 63-103).

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(02), 305–352. https://doi.org/10.1017/s027226310505014x

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The modern language journal*, *92*(2), 232-249.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology A*, *49*(1), 234–250. https://doi.org/10.1080/027249896392883

Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language learning*, *51*, 1-46.

Elsherif, M. M., Preece, E., & Catling, J. C. (2023). Age-of-acquisition effects: A literature review. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(5), 812–847. https://doi.org/10.1037/xlm0001215

Elsherif, M. M., Wheeldon, L., & Frisson, S. (2022). Phonological precision for word recognition in skilled readers. *Quarterly Journal of Experimental Psychology*, *75*(6), 1021–1040. https://doi.org/10.31234/osf.io/vftxd

Endres, T., Leber, J., Böttger, C., Rovers, S., & Renkl, A. (2020). Improving lifelong learning by fostering students' learning strategies at university. Psychology Learning & Teaching, 19(3), 336-354.

Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research*, *51*(6), 1580–1587. https://doi.org/10.1044/1092-4388(2008/07-0210)

Engelkamp, J., & Zimmer, H. D. (1997). Sensory and motor aspects of memory for actions. *Psychological Research*, *60*(1-2), 87–99.

Enk, A. and Cate, O. (2020). "languaging" tacit judgment in formal postgraduate assessment: the documentation of ad hoc and summative entrustment decisions. *Perspectives on Medical Education*, *9*(6), 373-378. https://doi.org/10.1007/s40037-020-00616-x

Epstein, J. N., Langberg, J. M., Rosen, P. J., Graham, A., Narad, M. E., Antonini, T. N., Brinkman, W. B., Froehlich, T., Simon, J. O., & Altaye, M. (2011). Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations. *Neuropsychology*, *25*(4), 427–441. https://doi.org/10.1037/a0022155

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, *102*(2), 211. https://psycnet.apa.org/fulltext/1995-24067-001.html

Escobar, G. P., Tuninetti, A., Antoniou, M., & Escudero, P. (2023). Understanding preschoolers' word learning success in different scenarios: Disambiguation meets statistical learning and eBook reading. *Frontiers in Psychology*, *14*, 1118142. https://doi.org/10.3389/fpsyg.2023.1118142

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*(5), 369–377. https://doi.org/10.1037/h0046888

Eysenck, M. W., & Eysenck, M. C. (1979). *Personality and individual differences*. Plenum Press.

Farnia, F., and Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics*, *32*(4), 711–738. https://doi.org/10.1017/s0142716411000038

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*(1), 1-5.

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory*, *20*(7), 655-666.

Fecher, N., & Johnson, E. K. (2018). Effects of language experience and task demands on talker recognition by children and adults. *The Journal of the Acoustical Society of America*, *143*(4), 2409-2418.

Fehlmann, B., Coynel, D., Schicktanz, N., Milnik, A., Gschwind, L., Hofmann, P., ... & de Quervain, D. J. (2020). Visual exploration at higher fixation frequency increases subsequent memory recall. *Cerebral cortex communications*, *1*(1), tgaa032.

Feldman, J. (2016). The simplicity principle in perception and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(5), 330–340. https://doi.org/10.1002/wcs.1406

Feng, X., & Webb, S. (2020). Perceived usefulness of elaborative and memorization strategies for L2 vocabulary learning. *Language Teaching Research*, *24*(4), 481-503.

Ferguson, C. A. (1959). Diglossia. Word, 15(2), 325-340.

Ferguson, G. R. (2013). Language practices and language management in a UK Yemeni community. *Journal of Multilingual and Multicultural Development*, *34*(2), 121–135. https://doi.org/10.1080/01434632.2012.724071

Fiorella, L., & Mayer, R. E. (2015). Eight ways to promote generative learning. *Educational Psychology Review*, *27*, 709-715.

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2014). Metacognition: Computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1280-1286.

Forrin, N. D., & MacLeod, C. M. (2016). Order information is used to guide recall of long lists: Further evidence for the item-order account. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 125.

Forrin, N. D., & MacLeod, C. M. (2017). Relative speed of processing determines color–word contingency learning. *Memory & Cognition*, *45*, 1206-1222.

Forrin, N. D., Groot, B., & MacLeod, C. M. (2016). The d-Prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(7), 1090.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*(7), 1046–1055. https://doi.org/10.3758/s13421-012-0210-8

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). North-Holland.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental science*, *21*(5), e12659.

Fougnie, D. (2008). The relationship between attention and working memory. In N. B. Johansen (Ed.), *New Research on short-term Memory* (pp. 1–45). Nova Science Publishers, Inc. http://www.psy.vanderbilt.edu/students/fougnidl/Fougnie-chap1.pdf

Friedman, G. N., Johnson, L., & Williams, Z. M. (2018). Long-Term visual memory and its role in learning suppression. *Frontiers in Psychology*, *9*, 1896. https://doi.org/10.3389/fpsyg.2018.01896

Friedmann, N., & Rusou, D. (2015). Critical period for first language: The crucial role of language input during the first year of life. *Current Opinion in Neurobiology*, *35*(1), 27–34. https://doi.org/10.1016/j.conb.2015.06.003

Friesen, D. C., & Jared, D. (2012). Cross-language phonological activation of meaning: Evidence from category verification. *Bilingualism: Language and Cognition*, *15*(1), 145–156. https://doi.org/10.1017/s1366728910000489

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241. https://doi.org/10.1111/j.1467-9280.1996.tb00366.x

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *58*(3), 934-945.

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, *49*, 87–115. https://doi.org/10.1146/annurev.psych.49.1.87

Galdo, M., Weichart, E. R., Sloutsky, V. M., & Turner, B. E. (2022). The quest for simplicity in human learning: Identifying the constraints on attention. *Cognitive Psychology*, *138*, 101508. https://doi.org/10.1016/j.cogpsych.2022.101508

Galli, G. (2014). What makes deeply encoded items memorable? Insights into the levels of processing framework from neuroimaging and neuromodulation. *Frontiers in Psychiatry*, *5*, 61.

Gamble, C., Wilkins, M., Aliponga, J., Koshiyama, Y., Yoshida, K., & Ando, S. (2018). Learner autonomy dimensions: What motivated and unmotivated EFL students think. *Lingua Posnaniensis*, *60*(1), 33–47. https://doi.org/10.2478/linpo-2018-0003

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, *18*, 579-583.

Gardner, R. C. (2005). Integrative motivation and second language acquisition. In Z. Dörnyei & R. Schmidt (Eds.), Motivation and second language acquisition (pp. 1-19). University of Hawaii Press.

Gardner, R. C., & Lambert, W. E. (1972). Attitudes and motivation in second-language learning. Newbury House Publishers.

Gardner, R. C., & MacIntyre, P. D. (1992). A student's contributions to second language learning. Part I: Cognitive variables. *Language Teaching*, *25*(04), 211–220. https://doi.org/10.1017/s026144480000700x

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105-132.

Gaskell, M. G., & Ellis, A. W. (2009). Word learning and lexical development across the lifespan. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3607–3615. https://doi.org/10.1098/rstb.2009.0213

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266. https://doi.org/10.1016/s0010-0285(02)00003-8

Gasparini, S. (2004). Implicit versus explicit learning: Some implications for L2 teaching. *European Journal of Psychology of Education*, *19*(2), 203–219. https://doi.org/10.1007/bf03173232

Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*(2), 319–333. https://doi.org/10.1017/s0272263199002090

Gass, S. M., & Selinker, L. (2001). Second Language Acquisition, an Introductory Course (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S. M., & Selinker, L. (2008). Second language acquisition: An introductory course. Routledge.

Gates, A. I. (1917). Experiments as the relative efficiency of men and women in memory & reasoning. *Psychological Review*, *24*(2), 139.

Gathercole, S. E. (1998). The development of memory. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(1), 3-27.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017/s0142716406060383

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Psychology Press.

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*(2), 110–119. https://doi.org/10.3758/bf03213478

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. M. (2001). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 481–488. https://doi.org/10.1037/0278-7393.27.2.481

Gathercole, S. E., Woolgar, F., Kievit, R. A., Astle, D., Manly, T., & Holmes, J. (2016). How common are WM deficits in children with difficulties in reading and mathematics? *Journal of Applied Research in Memory and Cognition*, *5*(4), 384–394. https://doi.org/10.1016/j.jarmac.2016.07.013

Ghorbani, M. and Riabi, N. (2011). The impact of memory strategy instruction on learners' efl vocabulary retention. *Theory and Practice in Language Studies*, *1*(9). https://doi.org/10.4304/tpls.1.9.1222-1226

Gierut, J. A., & Morrisette, M. L. (2011). Effect size in clinical phonology. Clinical Linguistics & Phonetics, 25(11-12), 975-980.

Gillon, G. T. (2005). Phonological awareness.

Gionet, S., Guitard, D., & Saint-Aubin, J. (2022). The production effect interacts with serial positions. *Experimental Psychology*, *69*(1), 12–22. https://doi.org/10.1027/1618-3169/a000540

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 5-16.

Glanzer, M., & Cofer, C. N. (1971). Are we aware of structural features of stimuli? *Journal of Experimental Psychology*, *84*(1), 27-34.

Glavaš, G., & Štajner, S. (2015). Simplifying lexical simplification: Do we need simplified corpora? *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Confere*, *2*, 63–68. https://doi.org/10.3115/v1/p15-2011

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Godfroid, A., Ahn, J., Choi, I. N. A., Ballard, L., Cui, Y., Johnston, S., ... & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563-584.

Godwin-Jones, R. (2018). Contextualized vocabulary learning. *Language Learning & Technology*, *22*(3), 1–19. https://doi.org/10125/44651

Gogate, L. (2022). *An interactive approach to enhancing early word learning: From research to best practice*. Cambridge Scholars Publishing.

Goh, W. D., & Pisoni, D. B. (2003). Effects of lexical competition on immediate memory span for spoken words. *The Quarterly Journal of Experimental Psychology Section A*, *56*(6), 929–954. https://doi.org/10.1080/02724980244000710

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

González-Fernández, B., & Schmitt, N. (2017). Vocabulary acquisition. In S. Loewen & M. Sato (Eds.), *The routledge handbook of instructed second language acquisition* (pp. 280–298). Routledge. https://doi.org/10.4324/9781315676968

González-Valenzuela, M.-J., & Martín-Ruiz, I. (2016). Effects on reading of an early intervention program for children at risk of learning difficulties. *Remedial and Special Education*, *38*(2), 67–75. https://doi.org/10.1177/0741932516657652

Goodman, J., Marsh, R., Peterson, B. S., & Packard, M. G. (2013). Annual research review: The neurobehavioral development of multiple memory systems - implications for childhood and adolescent psychiatric disorders. *Journal of Child Psychology and Psychiatry*, *55*(6), 582–610. https://doi.org/10.1111/jcpp.12169

Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., & Stern, J. (1969). Alcohol and recall: State-dependent effects in man. *Science*, *163*(3873), 1358-1360.

Goswami, U. (1990). A special link between rhyming skill and the use of orthographic analogies by beginning readers. *Journal of Child Psychology and Psychiatry*, *31*(2), 301-311.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 981–998. https://doi.org/10.1080/02724980443000386

Granena, G. (2018). Cognitive aptitudes and l2 speaking proficiency. *Studies in Second Language Acquisition*, *41*(2), 313-336. https://doi.org/10.1017/s0272263118000256

Greenwood, S. C. (2010). *The power of words: Learning vocabulary in grades 4-9*. R&L Education.

Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.

Grünbaum, T. (2022). Responsibility for forgetting to do. *Erkenntnis*, *89*(2), 755–776. https://doi.org/10.1007/s10670-022-00554-6

Grünbaum, T., Oren, F., & Kyllingsbæk, S. (2021). A new cognitive model of long-term memory for intentions. *Cognition*, *215*, 104817. https://doi.org/10.1016/j.cognition.2021.104817

Gruneberg, M. M., & Morris, P. E. (1994). *Theoretical aspects of memory* (Vol. 12). Taylor & Francis.

Gu, P., and Johnson, R. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, *46*(4), 643–679. https://doi.org/10.1111/j.1467-1770.1996.tb01355.x

Guerrout, R., Azzaoui, C., & Omari, M. (2019). *Caregivers' role in child's language acquisition* [Doctoral dissertation, Ahmed Draia University-Adrar]. Ahmed Draia University Library. https://dspace.univ-adrar.edu.dz/jspui/handle/123456789/2476

Guo, Y. (2021). Exploring the dynamic interplay between foreign language enjoyment and learner engagement with regard to EFL achievement and absenteeism: A sequential mixed methods study. *Frontiers in Psychology*, *12*, 766058. https://doi.org/10.3389/fpsyg.2021.766058

Habash, N. Y. (2010). *Introduction to Arabic natural language processing*. Morgan & Claypool Publishers.

Hackl, H. (2018). Aptitude for vocabulary acquisition. *Exploring Language Aptitude: Views from Psychology, the Language Sciences, and Cognitive Neuroscience*, 43–56. https://doi.org/10.1007/978-3-319-91917-1_3

Haddad, Y. A. (2006). Dialect and standard in second language phonology: The case of Arabic. *Sky Journal of Linguistics*, *24*(3), 147–171. https://www.linguistics.fi/julkaisut/SKY2006/Haddad.pdf

Hallin, A. E., & Van Lancker Sidtis, D. (2017). A closer look at formulaic language: Prosodic characteristics of Swedish proverbs. *Applied Linguistics*, *38*(1), 68–89.

Halwani, N. (2017). Visual aids and multimedia in second language acquisition. *English Language Teaching*, *10*(6), 53–59. https://eric.ed.gov/?id=EJ1143525

Hamydy, E. and Hind, B. (2023). Investigating the effect of reciprocal teaching on summarizing among moroccan college students. *Journal of English Language Teaching and Applied Linguistics*, *5*(1), 45-52. https://doi.org/10.32996/jeltal.2023.5.1.6

Harley, T. A. (2014). *The psychology of language: From data to theory* (4th ed.). Psychology Press.

Harris, J. A., Miniussi, C., Harris, I. M., & Diamond, M. E. (2002). Transient storage of a tactile memory trace in primary somatosensory cortex. *Journal of Neuroscience*, *22*(19), 8720–8725. https://doi.org/10.1523/jneurosci.22-19-08720.2002

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child development*, *65*(1), 41-57.

Hattie, J. A., & Yates, G. C. (2014). Using feedback to promote learning. *Applying science of learning in education: Infusing psychological science into the curriculum*, 45-58.

Hattie, J., & Yates, G. C. (2014). *Visible learning and the science of how we learn*. Routledge.

Hazem, A. H., & Meteab, W. Y. (2019). Nominal Constructions in Modern Standard Arabic With Reference to English. *International Journal of Applied Linguistics & English Literature*.

He, A. W. (2010). The heart of heritage: Sociocultural dimensions of heritage language learning. *Annual Review of Applied Linguistics*, *30*, 66-82.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Blackwell Publishing.

Heikkilä J., Alho K., Hyvönen H., Tiippana K. (2015). Audiovisual semantic congruency during encoding enhances memory performance, *Exp. Psychol.* 62, 123–130.

Henry, L., & Winfield, J. (2010). Working memory and educational achievement in children with intellectual disabilities. *Journal of Intellectual Disability Research*, *54*(4), 354-365.

Hernandez, A., Li, P., & Macwhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, *9*(5), 220–225. https://doi.org/10.1016/j.tics.2005.03.003

Hindman, A. H., Farrow, J. M., Anderson, K., Wasik, B. A., & Snyder, P. A. (2021). Understanding child-directed speech around book reading in toddler classrooms: Evidence from early head start programs. *Frontiers in Psychology*, *12*, 717983. https://doi.org/10.3389/fpsyg.2021.719783

Hoff, E., Core, C., Place, S., Rumiche, R., Senor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(01), 1–27. https://doi.org/10.1017/s0305000910000759

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.

Hollingworth, H. L. (1913). Characteristic differences between recall and recognition. *The American Journal of Psychology*, *24*(4), 532–544. https://doi.org/10.2307/1413450

Holzen, K., Fennell, C., & Mani, N. (2018). The impact of cross-language phonological overlap on bilingual and monolingual toddlers' word recognition. *Bilingualism Language and Cognition*, *22*(3), 476–499. https://doi.org/10.1017/s1366728918000597

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*(4), 534–537. https://doi.org/10.1016/s0022-5371(72)80036-7

Hopman, E. W., & MacDonald, M. C. (2018). Production practice during language learning improves comprehension. *Psychological science*, *29*(6), 961-971.

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *62*(4), 242.

Howard-Jones, P., Ioannou, K., Bailey, R., Prior, J., Yau, S. H., & Jay, T. (2018). Applying the science of learning in the classroom. *Impact*, *2*, 9-12.

Hu, M. (2013). A study of the relationship between the linguistic distance of Chinese dialects and lexical acquisition of Chinese as a second language. Language Teaching Research, 17(4), 443-462.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition*, *21*(2), 181-193.

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258–286). Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. Language Learning, 51(3), 539-558.

Hummel, K. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics*, *30*(2), 225-249. https://doi.org/10.1017/s0142716409090109

Humphreys, M. S., Bain, J. D., & Pike, R. (2000). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *107*(1), 41-67.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, *17*(1), 23–59. https://eric.ed.gov/?id=EJ689121

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review*, *2*, 105-112.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. *Distinctiveness and memory*, 3-25.

Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science*, *22*(1), 10-15.

Hunt, R. R., & Worthen, J. B. (Eds.). (2006). *Distinctiveness and memory*. Oxford University Press.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Ibrahim, A. I. (2010). Noun Formation in Standard English and Modern Standard Arabic: A Contrastive Study. *Journal of Language Teaching & Research*, *1*(5).

Icht, M., & Mama, Y. (2015). The production effect in memory: A prominent mnemonic in children. *Journal of Child Language*, *42*(5), 1102–1124. https://doi.org/10.1017/s0305000914000713

Icht, M., & Mama, Y. (2022). The effect of vocal production on vocabulary learning in a second language. *Language Teaching Research*, *26*(1), 136216881988389. https://doi.org/10.1177/1362168819883894

Icht, M., Ben-David, N., & Mama, Y. (2020). Using vocal production to improve long-term verbal memory in adults with intellectual disability. *Behavior Modification*, *45*(5), 715–739. https://doi.org/10.1177/0145445520906583

Icht, M., Mama, Y., & Algom, D. (2014). The production effect in memory: Multiple species of distinctiveness. *Frontiers in psychology*, *5*, 886.

Illés, É., & Akcan, S. (2016). Bringing real-life language use into EFL classrooms. *ELT Journal*, *71*(1), 3–12. https://doi.org/10.1093/elt/ccw049

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences*, *369*(1651), 20130298.

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS One*, *10*(2), e0116494.

Indrisano, R., & Squire, J. R. (Eds.). (2000). *Perspectives on writing: research, theory, and practice*. Taylor & Francis.

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 325. https://psycnet.apa.org/buy/2011-22378-001

Ipek, H. (2009). Comparing and contrasting first and second language acquisition: Implications for language teachers. *English Language Teaching*, *2*(2), 155–163. https://eric.ed.gov/?id=EJ1082388

Irish, M., & Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in Behavioral Neuroscience*, *7*, 27. https://doi.org/10.3389/fnbeh.2013.00027

Irwin, D. E., & Thomas, L. E. (2008). Visual sensory memory. *Visual memory*, *1*(9), 9–43. https://doi.org/10.1093/acprof:oso/9780195305487.003.0002

Ishikawa, Y. (2019). Incidental and intentional learning of grammatical structures: The role of awareness and individual differences. *Language Awareness*, *28*(1), 62-81.

Jackson, E., Leitao, S., Claessen, M., & Boyes, M. (2019). Fast mapping short and long words: Examining the influence of phonological short-term memory and receptive vocabulary in children with developmental language disorder. *Journal of Communication Disorders*, *79*, 11–23. https://doi.org/10.1016/j.jcomdis.2019.02.001

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, *30*(5), 513-541

Jaeger, T. F., Graff, P., Croft, W., & Pontillo, D. (2008). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, *15*(2), 245–280. https://doi.org/10.1515/lity.2011.021

Jamieson, R. K., & Spear, J. (2014). The offline production effect. *Canadian Journal of Experimental Psychology*, *68*(1), 20–28. https://doi.org/10.1037/cep0000009

Jamieson, R. K., Hannah, S. D., & Crump, M. J. C. (2010). A memory-based account of retrospective revaluation. *Canadian Journal of Experimental Psychology*, *64*(3), 153–164. https://doi.org/10.1037/a0020886

Jaradat, A. A., & Al-Khawaldeh, N. N. A. (2015). Teaching modern standard Arabic for non-native speakers as a lingua franca. *Mediterranean Journal of Social Sciences*, *6*(5), 490–499. https://doi.org/10.5901/mjss.2015.v6n5p490

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (2000). Verbal short-term memory deficits in Down syndrome: A consequence of problems in rehearsal?. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(2), 233-244.

Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia*, *10*(3), 196–214. https://doi.org/10.1002/dys.278

Jescheniak, J. D. and Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: further evidence from the coactivation of near-synonyms. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *24*(5), 1256-1274. https://doi.org/10.1037/0278-7393.24.5.1256

Jescheniak, J. D., & Schriefers, H. (1998). Representation and selection of phonological information in speech production: Evidence from picture-word interference experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1256-1274.

Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000). The concreteness effect: Evidence for dual coding and context availability. *Brain and language*, *74*(1), 103-112.

Jiang, C. (2021). Critical period hypothesis in language acquisition. *Lecture Notes in Education Psychology and Public Media*, *1*, 1–5. https://doi.org/10.54254/2753-7048/1/ICEIPI_152

Jiang, N. (2023). *The study of bilingual language processing*. Oxford University Press.

Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *Modern Language Journal*, *104*(3), 550-566.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, *21*(1), 60-99.

Jonassen, D. H., & Grabowski, B. L. (1993). *Handbook of individual differences, learning, and instruction*. Lawrence Erlbaum Associates.

Jones, A. C., & Pyc, M. A. (2014). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 300–305. https://doi.org/10.1037/a0033337

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*(1), 193–224. https://doi.org/10.1146/annurev.psych.59.103006.093615

Jonker, T. R., & MacLeod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 831.

Jonker, T. R., Levene, M., & MacLeod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 441.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language teaching*, *44*(2), 137-166.

Kahn, J. B., Ward, R. D., Kahn, L. W., Rudy, N. M., Kandel, E. R., Balsam, P. D., & Simpson, E. H. (2012). Medial prefrontal lesions in mice impair sustained attention but spare maintenance of information in working memory. *Learning & Memory*, *19*(11), 513-517.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2-3), 199–225. https://doi.org/10.1016/0001-6918(94)90003-5

Kan, P., and Kohnert, K. (2011). A growth curve analysis of novel word learning by sequential bilingual preschool children. *Bilingualism Language and Cognition*, *15*(3), 452–469. https://doi.org/10.1017/s1366728911000356

Kandel, E. R., Schwartz, J., Jessell, T., Siegelbaum, S., Hudspeth, A. J., & Mack, S. (2000). *Principles of neural science* (Vol. 4, pp. 1227–1246). McGraw-Hill Medical.

Kantak, S. S., Mummidisetty, C. K., & Stinear, J. W. (2012). Primary motor and premotor cortex in implicit sequence learning - evidence for competition between implicit and explicit human motor memory systems. *European Journal of Neuroscience*, *36*(5), 2710–2715. https://doi.org/10.1111/j.1460-9568.2012.08175.x

Kapnoula, E. C., & Samuel, A. G. (2022). Reconciling the contradictory effects of production on word learning: Production may help at first, but it hurts later. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(3), 394.

Karpicke JD, Butler AC, Roediger HL 3rd. (2009). Metacognitive strategies in student learning: do students practise retrieval when they study on their own? *Memory*, *17*(4):471-9. https://doi.org/10.1080/09658210802647009

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*(4), 469-486.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborate studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. Science, 319(5865), 966-968.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471-479.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2014). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, 22(5), 671-679. https://doi.org/10.1080/09658211.2013.837925

Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*(1), 121–148. https://doi.org/10.1080/01690965.2010.486579

Kaushanskaya, M., Yoo, J., & Van Hecke, S. (2013). Word learning in adults with second-language experience: Effects of phonological and referent familiarity.

Keene, O. N. (1995). The log transformation is special. *Statistics in Medicine*, *14*(8), 811–819. https://doi.org/10.1002/sim.4780140810

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, *1*(1), 1-26.

Kersten, S. (2010). *The mental lexicon and vocabulary learning: Implications for the foreign language classroom* (Vol. 43). BoD–Books on Demand.

Khamis-Dakwar, R. (2010). The impact of diglossia on the development of Arabic literacy skills: A review of the literature. International Journal of Bilingual Education and Bilingualism, 13(3), 307-321.

Kharrat, Y., Alosh, M., & Clark, A. (2018). Review of Ahlan wa Sahlan: Functional modern standard Arabic for Beginners, 2nd ed. *Al-'Arabiyya*, *51*, 115–117. https://www.jstor.org/stable/26578122

Kinginger, C. (2011). Enhancing language learning in study abroad. *Annual Review of Applied Linguistics*, *31*, 58-73.

Klein, S. B. (2013). The temporal orientation of memory: It's time for a change of direction. *Journal of Applied Research in Memory and Cognition*, *2*(4), 222–234. https://doi.org/10.1016/j.jarmac.2013.08.001

Klimesch, W. (2013). *The structure of long-term memory: A connectivity model of semantic processing*. Psychology Press.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, *16*(8), 1412-1425.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language teaching and Research*, *4*(5), 900.

Kolb, N. (2022). *Child second language development in immersion education: A study on generic determiner phrases in L2 German and L2 French (p. 298)*. Peter Lang International Academic Publishers.

Konstantinou, N., Beal, E., King, J., & Lavie, N. (2014). Working memory load and distraction: Dissociable effects of visual maintenance and cognitive control. *Attention, Perception & Psychophysics*, *76*(7), 1985-1997. https://doi.org/10.3758/s13414-014-0742-z

Kormos, J. (2014). *Speech production and second language acquisition*. Routledge.

Krashen, S. D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. The Modern Language Journal, 73(4), 440-464.

Krashen, S. D. (2004). The power of reading: Insights from the research (2nd ed.). Libraries Unlimited.

Krashen, S. D. (2011). Free voluntary reading. Libraries Unlimited.

Krashen, S. D., & Terrell, T. (1983). *Natural approach*. Pergamon.

Krishnan, S., Watkins, K. E., & Bishop, D. V. M. (2017). The effect of recall, reproduction, and restudy on word learning: A pre-registered study. *BMC Psychology*, *5*(1), 28–42. https://doi.org/10.1186/s40359-017-0198-8

Kroll, J. F., & Bialystok, E. (2013). The bilingual advantage: Effects on mind and brain. *Trends in Cognitive Sciences*, *17*(6), 240–250.

Kroll, J. F., & Sunderman, G. (2003). Cognitive processes in second language learners and bilinguals: The development of lexical and conceptual representations. *The handbook of second language acquisition*, 104-129.

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, *9*(2), 74–78. https://doi.org/10.1111/cdep.12110

Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, *67*(5), 713–727. https://doi.org/10.1016/j.neuron.2010.08.038

Kurtz, K. H., & Hovland, C. I. (1953). The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *Journal of Experimental Psychology*, *45*(3), 157–164. https://doi.org/10.1037/h0061022

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Labanieh, K. (2019). *Heritage language learners of Arabic in Islamic schools: Opportunities for attaining arabic language proficiency* [Doctoral dissertation, The University of Wisconsin-Milwaukee]. ProQuest Dissertation Publishing. https://www.proquest.com/openview/2cf515bab9dd6ad878d8b904550dcdf2/1?pq-origsite=gscholar&cbl=18750&diss=y

Lambert, A. M., Bodner, G. E., & Taikh, A. (2016). The production effect in long-list recall: In no particular order? *Canadian Journal of Experimental Psychology*, *70*(2), 165–176. https://doi.org/10.1037/cep0000086

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). Academic Press.

LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2014). Decoding attended information in short-term memory: An EEG study. *Journal of Cognitive Neuroscience*, 26(9), 1883-1894.

Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, *75*(4), 440-448.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*(4), 567-587.

Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla yearbook*, *5*(1), 223-250.

Laufer, B. (2005). Instructed second language vocabulary learning: The fault in the 'default hypothesis'. *Investigations in Instructed Second Language Acquisition*, 286-303.

Laufer, B. (2006). Comparing focus on form and focus on forms in second-language vocabulary learning. *Canadian Modern Language Review*, *63*(1), 149-166.

Laufer, B. (2016). The three "I" s of second language vocabulary learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 343–354). https://doi.org/10.4324/9781315716893

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. Language Learning, 54(3), 399-436.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399–436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

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

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*(6), 687–711. https://doi.org/10.1177/1362168814559797

Lavis, L., & Mahy, C. E. V. (2021). "I'll remember everything no matter what!": The role of metacognitive abilities in the development of young children's prospective memory. *Journal of Experimental Child Psychology*, *207*, 105117. https://doi.org/10.1016/j.jecp.2021.105117

Le Pelley, M. E., & McLaren, I. P. L. (2001). Retrospective revaluation in humans: Learning or memory? *The Quarterly Journal of Experimental Psychology Section B*, *54*(4b), 311–352. https://doi.org/10.1080/713932762

Legault, J., Zhao, J., An, Y., Chen, W., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages*, *4*(1), 13. https://doi.org/10.3390/languages4010013

Lei, L., & Shuyuan, X. (2005). A comparison of contributions of explicit and implicit memory to the results of two kinds of tests with three kinds of remembering methods. *Psychological Science (China)*, *28*(6), 1410–1414.

Lenneberg, E. H. (1967). The study of language and language acquisition. *PDF document*.

Lenth, R. V. (2022, August 20). *Estimated marginal means, aka least-squares means [R package emmeans version 1.6.3]*. Cran.R-Project. https://cran.r-project.org/package=emmeans

Lenth, R. V. (2022). Emmeans: Estimated marginal means, aka least-squares means (R package version 1.7.5).

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-j

Levine, L. J., Lench, H. C., & Safer, M. A. (2009). Functions of remembering and misremembering emotion. *Applied Cognitive Psychology*, *23*(8), 1059–1075. https://doi.org/10.1002/acp.1610

Li, D., Hu, Y., Qi, M., Zhao, C., Jensen, O., Huang, J., & Song, Y. (2023). Prioritizing flexible working memory representations through retrospective attentional strengthening. *NeuroImage*, *269*, 119902. https://doi.org/10.1016/j.neuroimage.2023.119902

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, *97*(3), 634-654.

Li, W., Yu, J., Zhang, Z., & Liu, X. (2022). Dual coding or cognitive load? exploring the effect of multimodal input on english as a foreign language learners' vocabulary learning. *Frontiers in Psychology*, *13*, 833487. https://doi.org/10.3389/fpsyg.2022.834706

Lichtenberger, E. O., & Kaufman, A. S. (2012). *Essentials of WAIS-IV assessment* (Vol. 96). John Wiley & Sons.

Lieven, E., & Tomasello, M. (2008). Children's first language acquistion from a usage-based perspective. In *Handbook of cognitive linguistics and second language acquisition* (pp. 178-206). Routledge.

Lightbown, P. M., Spada, N., Ranta, L., & Rand, J. (1999). *How languages are learned* (Vol. 2). Oxford university press.

Lin, O. Y. H., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology*, *66*(3), 212–216. https://doi.org/10.1037/a0028309

Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *The Mental Lexicon*, *3*(3), 349–374. https://doi.org/10.1075/ml.3.3.06lin

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 608–622. https://doi.org/10.1037/a0029243

Liu, Y. (2013). Toward a comprehensive model of foreign language vocabulary learning: To integrate instructed learning with incidental learning. *Asian Social Science*, *9*(10), 142–150. https://doi.org/10.5539/ass.v9n10p142

Lockhart, R. S., Craik, F. I., & Jacoby, L. (1976). Depth of processing, recognition and recall.

Long, M. (1991). Focus on Form: A Design Feature in Language Teaching Methodology. In K. De Bot, R. Ginsberg, & C. Kramsch (Eds.), Foreign Language Research in Cross-Cultural Perspectives (pp. 39-52). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/sibil.2.07lon

Long, M. H., De Bot, K., Ginsberg, R., & Kramsch, C. (2001). Focus on form: A design feature in language teaching methodology. *English language teaching in its social context*, 180-190.

López Assef, B., Desmeules-Trudel, F., Bernard, A., & Zamuner, T. S. (2021). A shift in the direction of the production effect in children aged 2–6 years. *Child Development*, *92*(6), 2447–2464. https://doi.org/10.1111/cdev.13618

López Assef, B., Strahm, S., Boyce, K., Page, M., & Zamuner, T. (2023). Production benefits recall of novel words with frequent, but not infrequent sound patterns. *Glossa: A Journal of General Linguistics*, *8*(1), 1–21. https://doi.org/10.16995/glossa.8582

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147–158). Cambridge University Press.

Ludewig, U., Hübner, N., & Schroeder, S. (2023). Vocabulary, text coverage, word frequency and the lexical threshold in elementary school reading comprehension. *Reading and Writing*, *36*(9), 2409–2431. https://doi.org/10.1007/s11145-022-10385-0

Luthra, S., Guediche, S., Blumstein, S. E., & Myers, E. B. (2018). Neural substrates of subphonemic variation and lexical competition in spoken word recognition. *Language, Cognition and Neuroscience*, *34*(2), 151–169. https://doi.org/10.1080/23273798.2018.1531140

Luz, R. M. D., Marinho, D. C. D. B., Lima, A. P. E., & Coriolano-Marinus, M. W. L. (2022). Educational interventions in child development and health literacy assumptions: An integrative review. *Revista brasileira de enfermagem*, *76*(1), e20220116.

Lwin, M. O., Morrin, M., & Krishna, A. (2010). Exploring the superadditive effects of scent and pictures on verbal recall: An extension of dual coding theory. *Journal of Consumer Psychology*, *20*(3), 317–326. https://doi.org/10.1016/j.jcps.2010.04.001

Maamouri, M. (1998). Language education and human development: Arabic diglossia and its impact on the quality of education in the Arab region. The World Bank.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica*, *98*(2-3), 291–310. https://doi.org/10.1016/s0001-6918(97)00047-4

MacLeod C. M, Dunbar K. (1998). Training and Stroop-like interference: evidence for a continuum of automaticity. *Journal of Exp Psychol Learn Mem Cogn*, *14*(1):126-35. https://doi.org/10.1037//0278-7393.14.1.126

MacLeod, C. M. (2007). The concept of inhibition in cognition. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 3–23). American Psychological Association. https://doi.org/10.1037/11587-001

MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*, 1197-1202.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*(4), 390–395. https://doi.org/10.1177/0963721417691356

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 132–142. https://doi.org/10.1037/0278-7393.22.1.132

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Deliberate production enhances memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(3), 671.

MacNeilage, P. F., & Davis, B. L. (1993). Motor explanations of babbling and early speech patterns. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage & J. Morten (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 341-352). Springer Netherlands.

MacWhinney, B. (2005). Extending the competition model. *International Journal of Bilingualism*, *9*(1), 69–84. https://doi.org/10.1177/13670069050090010501

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural processes*, *66*(3), 309-332.

Magallón, S., Narbona, J., & Crespo-Eguílaz, N. (2016). Acquisition of motor and cognitive skills through repetition in typically developing children. *PLOS ONE*, *11*(7), e0158684. https://doi.org/10.1371/journal.pone.0158684

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–156. https://doi.org/10.1080/03640210709336987

Mainela-Arnold, E., Evans, J. L., & Coady, J. A. (2010). Explaining lexical–semantic deficits in specific language impairment: The role of phonological similarity, phonological working memory, and lexical competition. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1742–1756. https://doi.org/10.1044/1092-4388(2010/08-0198)

Majerus, S., & Cowan, N. (2016). The nature of verbal short-term impairment in dyslexia: The importance of serial order. *Frontiers in Psychology*, *7,* 214259. https://doi.org/10.3389/fpsyg.2016.01522

Makita, K., Yamazaki, M., Tanabe, H. C., Koike, T., Kochiyama, T., Yokokawa, H., Yoshida, H., & Sadato, N. (2013). A functional magnetic resonance imaging study of foreign-language vocabulary learning enhanced by phonological rehearsal: The role of the right cerebellum and left fusiform gyrus. *Mind, Brain, and Education*, *7*(4), 213–224. https://doi.org/10.1111/mbe.12029

Malaia, E. A., Krebs, J., Roehm, D., & Wilbur, R. B. (2020). Age of acquisition effects differ across linguistic domains in sign language: EEG evidence. *Brain and Language*, *200*, 104708. https://doi.org/10.1016/j.bandl.2019.104708

Mama, Y., & Icht, M. (2016). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory*, *24*(1), 98–113. https://doi.org/10.1080/09658211.2014.986135

Mama, Y., & Icht, M. (2016). The effects of listening and speaking versus listening and writing on memory. *Memory & Cognition*, *44*(5), 856-864.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. https://doi.org/10.1037/0033-295x.87.3.252

Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, *3*(4), 386–396. https://doi.org/10.1037/0278-7393.3.4.386

Mangina, C. A., & Sokolov, E. N. (2006). Neuronal plasticity in memory and learning abilities: Theoretical position and selective review. *International Journal of Psychophysiology*, *60*(3), 203–214. https://doi.org/10.1016/j.ijpsycho.2005.11.004

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Marchman, V. A., Ashland, M. D., Loi, E. C., Adams, K. A., Fernald, A., & Feldman, H. M. (2019). Predictors of early vocabulary growth in children born preterm and full term: A study of processing speed and medical complications. *Child Neuropsychology*, *25*(7), 943–963. https://doi.org/10.1080/09297049.2019.1569608

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2000). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Association for Supervision and Curriculum Development.

Masek, L. R., McMillan, B. T. M., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review*, *60*, 100961. https://doi.org/10.1016/j.dr.2021.100961

Matlin, M. W. (2016). Pollyanna principle. In M. W. Matlin (Ed.), *Cognitive illusions* (pp. 315-335). Psychology Press.

Matsukura, M., Luck, S. J., & Vecera, S. P. (2007). Attention effects during visual short-term memory maintenance: Protection or prioritization? *Perception & Psychophysics*, *69*(8), 1422–1434. https://doi.org/10.3758/bf03192957

Mattson, M. P. (2014). Superior pattern processing is the essence of the evolved human brain. *Frontiers in Neuroscience*, *8*(265), 265. https://doi.org/10.3389/fnins.2014.00265

Matychuk, P. (2005). The role of child-directed speech in language acquisition: A case study. *Language Sciences*, *27*(3), 301–379. https://doi.org/10.1016/j.langsci.2004.04.004

Maule, J., Skelton, A. E., & Franklin, A. (2023). The development of color perception and cognition. *Annual Review of Psychology*, *74*(1), 87–111. https://doi.org/10.1146/annurev-psych-032720-040512

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/cbo9780511816819.004

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*(8), 760–769. https://doi.org/10.1037/0003-066x.63.8.760

Mayer, R. E. (2014). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1–24). Cambridge University Press. https://doi.org/10.1017/cbo9781139547369.002

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423. https://doi.org/10.1111/jcal.12197

Mayer, R. E. (2021). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 145–157). Cambridge University Press. https://doi.org/10.1017/9781108894333.015

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). Cambridge University Press. https://doi.org/10.1017/cbo9781139547369.015

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, *12*(1), 107–119. https://doi.org/10.1016/s0959-4752(01)00018-4

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McCombs, B. (2012). *Developing responsible and autonomous learners: A key to motivating students*. American Psychological Association. https://www.apa.org/education/ k12/learners

McCreath, G. A., Linehan, C. M. J., & Mar, R. A. (2017). Can differences in word frequency explain why narrative fiction is a better predictor of verbal ability than nonfiction? *Discourse Processes*, *54*(5-6), 373–381. https://doi.org/10.1080/0163853x.2017.1289794

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*(2), 237–255. https://doi.org/10.3758/pbr.15.2.237

McDaniel, M. A., & Einstein, G. O. (1993). The importance of cue familiarity and cue distinctiveness in prospective memory. *Memory*, *1*(1), 23–41. https://doi.org/10.1080/09658219308258223

McKenzie, W. A., & Tiberghien, G. (2004). Context effects in recognition memory: The role of familiarity and recollection. *Consciousness and Cognition*, *13*(1), 20–38. https://doi.org/10.1016/s1053-8100(03)00023-0

McKinney, L. C., & Woodward, A. E. (2004). Remembering what one intended to forget: The lack of directed forgetting effects in implicit memory. *The American Journal of Psychology*, *117*(2), 169–190. https://doi.org/10.2307/4149021

McMurray, B., Klein-Packard, J., & Tomblin, B. (2018). Preprint: Submitted 4/21/2018. *PsyArXiv (OSF Preprints)*. https://doi.org/10.31234/osf.io/jvrcy

McMurray, B., Samelson, V. M., Lee, S. H., & Bruce Tomblin, J. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language*, *43*(4), 652-679.

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching*, *13*(3-4), 221-246.

Medina, J. J. (2008). *The biology of recognition memory Archived 2016-03-04 at the Wayback Machine*. Psychiatric Times. https://www.semanticscholar.org/paper/the-biology-of-recognition-memory-medina/c30f4e96151a7 44e4bb359e46395 7eef87 ec4f37

Meganathan, P., Thai Yap, N., Paramasivam, S., & Jalaluddin, I. (2019). Incidental and intentional learning of vocabulary among young ESL learners. *3L the Southeast Asian Journal of English Language Studies*, *25*(4), 51–67. https://doi.org/10.17576/3l-2019-2504-04

Melby-Lervåg, M., Lervåg, A., Lyster, S.-A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-Repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science*, *23*(10), 1092–1098. https://doi.org/10.1177/0956797612443833

Menghini, D., Finzi, A., Carlesimo, G. A., & Vicari, S. (2011). Working memory impairment in children with developmental dyslexia: Is it just a phonological deficity? *Developmental Neuropsychology*, *36*(2), 199–213. https://doi.org/10.1080/87565641.2010.549868

Mercado, E. (2015). Implicit and explicit category learning by macaques (Macaca mulatta) and humans (Homo sapiens). *Journal of Comparative Psychology*, *129*(2), 131-141.

Merriam-Webster. (2024). *Lingua franca definition & meaning*. https://www.merriam-webster.com/dictionary/lingua%20franca

Meyerhoff, H.S., Jaggy, O., Papenmeier, F. *et al.* Long-term memory representations for audio-visual scenes. *Mem Cogn* **51**, 349–370 (2023). https://doi.org/10.3758/s13421-022-01355-6

Miao, S., Han, J., Gu, Y., Wang, X., Song, W., Li, D., Liu, Z., Yang, J., & Li, X. (2017). Reduced prefrontal cortex activation in children with attention-deficit/hyperactivity disorder during go/no-go task: A functional near-infrared spectroscopy study. *Frontiers in Neuroscience*, *11*, 367. https://doi.org/10.3389/fnins.2017.00367

Mierop, A., Maurage, P., & Corneille, O. (2020). Cognitive load impairs evaluative conditioning, even when individual CS and US stimuli are successfully encoded. *International Review of Social Psychology*, *33*(3), 1-8. https://doi.org/10.5334/irsp.339

Mihalca, L., Salden, R. J., Corbalan, G., Paas, F., & Miclea, M. (2011). Effectiveness of cognitive-load based adaptive instruction in genetics education. *Computers in Human Behavior*, *27*(1), 82-88.

Millar, L. (2006). Touchstones: Considering the relationship between memory and archives. *Archivaria*, *61*, 105–126. https://www.archivaria.ca/index.php/archivaria/article/view/12537

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working memory 2.0. *Neuron*, *100*(2), 463–475. https://doi.org/10.1016/j.neuron.2018.09.023

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Mirandola, C., Del Prete, F., Ghetti, S., & Cornoldi, C. (2011). Recollection but not familiarity differentiates memory for text in students with and without learning difficulties. *Learning and Individual Differences*, *21*(2), 206–209. https://doi.org/10.1016/j.lindif.2010.12.001

Mirman, D., McClelland, J. L., & Holt, L. L. (2011). Interactive activation and competition models of speech perception: A tutorial review. *Language and Cognitive Processes*, *26*(11), 1692–1722.

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, *117*(2), 53–68. https://doi.org/10.1016/j.bandl.2011.01.004

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140.

Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 452.

Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research*, *21*(3), 199–249. https://doi.org/10.1191/0267658305sr247oa

Montrul, S. (2010). Current issues in heritage language acquisition. *Annual Review of Applied Linguistics*, *30*, 3-23.

Montrul, S. A. (2016). *The acquisition of heritage languages*. Cambridge University Press.

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback.

Educational Technology Research and Development, *53*(3), 35-45. https://doi.org/10.1007/BF02504796

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46–60. https://doi.org/10.3758/s13423-010-0034-0

Moscati, A., Flint, J., & Kendler, K. S. (2020). Genetic influences on the timing of puberty: A developmental genetics approach. Behavior Genetics, 50(3), 151-161.

Moscati, V., Rizzi, L., Vottari, I., Chilosi, A. M., Salvadorini, R., & Guasti, M. T. (2020). Morphosyntactic weaknesses in developmental language disorder: The role of structure and agreement configurations. *Journal of Child Language*, *47*(5), 909–944. https://doi.org/10.1017/s0305000919000709

Mostari, H. A. (2009). What do mobiles speak in Algeria? Evidence from SMS language. *Current Issues in Language Planning*, *10*(4), 377-386.

Mostari, H. A. (2009). What do mobiles speak in Algeria? Evidence from SMS language. Current Issues in Language Planning, 10(4), 377-386.

Mroz, A. (2014). 21st century virtual language learning environments (vlles). *Language and Linguistics Compass*, *8*(8), 330–343. https://doi.org/10.1111/lnc3.12089

Muter, P. (1984). Recognition and recall of words with a single meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 198.

Myhill, D., & Brackley, M. (2004). Making connections: Teachers' use of children's prior knowledge in whole class discourse. *British Journal of Educational Studies*, *52*(3), 263-275.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, 10(5–6), 389–395.

Nalbantian, S. (2003). *Memory in literature from rousseau to neuroscience*. Palgrave Macmillan.

Namkung, J. M., & Fuchs, L. S. (2015). Children's acquisition of fraction knowledge from third through fifth grade: Race and gender differences and relations to mathematical achievement. *Journal of Educational Psychology*, *107*(3), 909-928.

Naphade, M. R., & Huang, T. S. (2002). Extracting semantics from audio-visual content: The final frontier in multimedia retrieval. *IEEE Transactions on Neural Networks*, *13*(4), 793–810. https://doi.org/10.1109/tnn.2002.1021881

Nashir, M., & Laili, R. N. (2018). The effect of metacognitive strategy on students' reading comprehension: A case study of second grade students at MA Darul Huda Wonodadi Blitar. Journal of English Language Teaching and Linguistics, 3(2), 157-168.

Nation, I. S. P. (1982). Beginning to learn foreign vocabulary: A review of the research. *RELC Journal*, *13*(1), 14–36. https://doi.org/10.1177/003368828201300102

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge University Press.

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2015). Principles guiding vocabulary learning through extensive reading. Reading in a Foreign Language, 27(1), 136-145.

Nation, I. S. P. (2016). Making and using word lists for language learning and testing.

Nation, I. S., & Nation, I. S. P. (2001). *Learning vocabulary in another language* (Vol. 10). Cambridge university press.

Nation, P. (2017). How vocabulary is learned. *Indonesian JELT*, *12*(1), 1-14. https://download.garuda.kemdikbud.go.id/article.php?article=641308&val=7524&title=How%20vocabulary%20is%20learned

Nation, P., & Meara, P. (2013). 3 Vocabulary. In N. Schimitt (Ed.), *An introduction to applied linguistics* (pp. 44–62). Routledge.

Navarrete, E., Pastore, M., Valentini, R., & Peressotti, F. (2015). First learned words are not forgotten: Age-of-acquisition effects in the tip-of-the-tongue experience. *Memory & Cognition*, *43*(7), 1085–1103. https://doi.org/10.3758/s13421-015-0525-3

Neath, I., & Cavanaugh, K. (2009). The neural basis of long-term memory storage. In P. J. Reinders & R. J. Connor (Eds.), *Levels of processing in perception: A tribute to Donald Broadbent* (pp. 105–123). Routledge.

Neath, I., & Surprenant, A. M. (2003). Human memory: An introduction to research, data, and theory. *Wadsworth*.

Nelson, D. L. (2014). Remembering pictures and words: Appearance, significance, and name. In L. S. Cermak & F. I. Craik (Eds.), *Levels of processing in human memory (PLE: Memory)* (pp. 45–76). Psychology Press.

Nemati, A. (2009). Memory vocabulary learning strategies and long-term retention. *International Journal of Vocational and Technical Education*, *1*(2), 14–24. http://www.academicjournals.org/ijvte

Nieuwenhuis-Mark, R. E. (2012). *Recall and the effect of repetition on recall*. Tilburg University Research Portal; Springer. https://research.tilburguniversity.edu/en/publications/recall-and-the-effect-of-repetition-on-recall

Nishiyama, R., & Ukita, J. (2013). Articulation of phonologically similar items disrupts free recall of nonwords. *Journal of Memory and Language*, *68*(1), 1-9.

Nor, N. M., & AB Rashid, R. (2018). A review of theoretical perspectives on language learning and acquisition. *Kasetsart Journal of Social Sciences*, *39*(1), 161–167. https://doi.org/10.1016/j.kjss.2017.12.012

Norris, J., & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis. Language Learning, 50, 417-528. http://dx.doi.org/10.1111/0023-8333.00136

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(04), 557-581. https://doi.org/10.1017/s027226310707043x

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr Opin Neurobiol*, *4*(6):769-76. https://doi.org/10.1016/j.conb.2004.10.016

Oberauer, K. (2009). Design for a working memory. In K. D. Federmeer & C. Lee (Eds.), *The psychology of learning and motivation* (Vol. 51, pp. 45–100). Elsevier Inc. https://doi.org/10.1016/s0079-7421(09)51002-x

Oh, T. M., Graham, S., Ng, P., Yeh, I. B., Chan, B. P. L., & Edwards, A. M. (2019). Age and proficiency in the bilingual brain revisited: Activation patterns across different l2-learner types. *Frontiers in Communication*, *4*, 39. https://doi.org/10.3389/fcomm.2019.00039

Olive, T. (2012). Writing and working memory: A summary of theories and findings. In E. L. Grigorenko, E. Mambrino, & D. D. Preiss (Eds.), *Writing: A mosaic of new perspectives* (pp. 125-140). Psychology Press.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*(2), 441–449. https://doi.org/10.2307/1130323

Olmsted, D. L. (2019). *Out of the mouth of babes: Earliest stages in language learning* (Vol. 117). Walter de Gruyter GmbH & Co KG.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2020). Competing activation in bilingual language production: Implications for models of lexical selection. *Bilingualism: Language and Cognition*, *23*(4), 849–861.

Oppenheim, G. M., Griffin, Z., Peña, E. D., & Bedore, L. M. (2020). Longitudinal evidence for simultaneous bilingual language development with shifting language dominance, and how to explain it. *Language Learning*, *70*(S2), 20–44. https://doi.org/10.1111/lang.12398

Ordan, N., Hershberg, N., & Shlesinger, M. (2010). Translational conflicts between cognate languages: Arabic into Hebrew as case in point. *Corpus Linguistics and Linguistic Theory*, *6*(2), 217–239. https://doi.org/10.1515/cllt.2010.008

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199-230.

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1543–1547. https://doi.org/10.1037/a0020604

Ozubko, J. D., Bamburoski, L. D., Carlin, K., & Fawcett, J. M. (2020). Distinctive encodings and the production effect: failure to retrieve distinctive encodings decreases recollection of silent items. *Memory*, *28*(2), 237-260.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1

Paas, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122-133.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/s15326985ep3801_1

Paisart, W., & Suriyatham, W. (2017). *A case study of learning vocabulary through pictures* [Doctoral dissertation, Thammasat University]. Ethesis Archive Libarary. http://ethesisarchive.library.tu.ac.th/thesis/2017/TU_2017_5521320019_8561_8336.pdf

Paivio, A. (1971). Imagery and language. In J. Segal (Ed.), *Imagery: Current cognitive approaches* (pp. 7–32). Academic Press. https://doi.org/10.1016/b978-0-12-635450-8.50008-x

Paivio, A. (1986). *Mental representations A dual coding approach*. Oxford University Press, Incorporated.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Lawrence Erlbaum Associates.

Palma, P., Lee, S., Hodgins, V., & Titone, D. (2023). From one bilingual to the next: an iterated learning study on language evolution in bilingual societies. *Cognitive Science*, *47*(5). https://doi.org/10.1111/cogs.13289

Pan, J., Zhang, L., & Yi, L. (2023). The role of morphological awareness in Chinese children's reading comprehension: A longitudinal study. Reading and Writing, 36(2), 367-389.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness Recall and Testimony. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 93–150). The Guilford Press.

Park, S. (2013). Immigrant students' heritage language and cultural identity maintenance in multilingual and multicultural societies. *Concordia Working Papers in Applied Linguistics*, *4*, 30–53. https://doe.concordia.ca/copal/documents/4_Park_Vol4.pdf

Parkinson, D. B. (2005). The Arabic Language and National Identity: A Study in Ideology. *Language*, *81*(1), 276-279.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing instruction and study to improve student learning. IES Practice Guide. NCER 2007-2004. National Center for Education Research.

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, *41*(4), 745-768.

Pelegrina, S., Mirandola, C., Martínez, R. L., & Carretti, B. (2023). Recollection contributes to children's reading comprehension: Using the process dissociation procedure in a working memory updating task. *Journal of Experimental Child Psychology*, *226*, 105550. https://doi.org/10.1016/j.jecp.2022.105550

Pellicer-Sánchez, A. (2015). Developing automaticity and speed of lexical access: The effects of incidental and explicit teaching approaches. *Journal of Spanish Language Teaching*, 2(2), 126-139.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97-130.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, *19*(3), 170–182. https://doi.org/10.1002/hbm.10110

Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). Oxford University Press.

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology*, *115*, 101238. https://doi.org/10.1016/j.cogpsych.2019.101238

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551-577.

Pham, G., Donovan, D., Dam, Q., & Contant, A. (2017). Learning words and definitions in two languages: What promotes cross-language transfer? *Language Learning*, *68*(1), 206–233. https://doi.org/10.1111/lang.12274

Pham, T., & Archibald, L. M. D. (2022). The role of phonological and semantic representations in verbal short-term memory and delayed retention. *Memory & Cognition*, *50*(2), 325–338. https://doi.org/10.3758/s13421-021-01216-8

Philip, J. (2010). Incidental acquisition of implicit and explicit knowledge of language. *TESOL Quarterly*, *44*(4), 665-679.

Poort, E. D., & Rodd, J. M. (2017). The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition. *Acta Psychologica*, *180*, 52–63. https://doi.org/10.1016/j.actpsy.2017.08.008

Pratte, M. S., & Rouder, J. N. (2012). Assessing the dissociability of recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1591–1607. https://psycnet.apa.org/buy/2012-12099-001

Prior, A., & van Hell, J. G. (2021). Including bilingual participants in cognitive research. In W. S. Francis (Ed.), *Bilingualism across the lifespan* (pp. 47–65). Routledge.

Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2020). The production effect improves memory in 7- to 10-year-old children. *Child Development*, *91*(3), 901–913. https://doi.org/10.1111/cdev.13247

Ptok, M., Thomson, S. J., Humphreys, K. R., & Watter, S. (2019). Congruency encoding effects on recognition memory: A stage-specific account of desirable difficulty. *Frontiers in Psychology*, *10*, 858. https://doi.org/10.3389/fpsyg.2019.00858

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *57*(s1), 155-199.

Punar-Özçelik, N., & Uzun, L. (2021). Comparing three different vocabulary teaching techniques for retention. In W. B. James, C. Cobanoglu, & M. Cavusoglu (Eds.), *Advances in global education and research* (pp. 1–10). USF M3 Publishing.

Quine, W.V.O. (1960). *Word and Object*. Cambridge.

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*(8), 904-915.

R Core Team. (2021). *R: The R project for statistical computing*. R-Project. https://www.r-project.org

Radić-Bojanić, B. B. (2021). Audiovisual media and the acquisition of efl vocabulary. *Nasleđe*, (48).

Radojevic, N. (2006). *Exploring the use of effective learning strategies to increase students' reading comprehension and test taking skills* [Doctoral dissertation, Brock University]. ProQuest Dissertation Publishing. http://hdl.handle.net/10464/1466

Rahman, M. M., Singh, M. K. M., & Pandian, A. (2021). The impact of mobile-assisted language learning (MALL) in developing countries: A systematic review. Education and Information Technologies, 26(4), 4515-4536.

Ramsook, K. A., Welsh, J. A., & Bierman, K. L. (2020). What you say, and how you say it: Preschoolers' growth in vocabulary and communication skills differentially predict kindergarten academic achievement and self-regulation. *Social Development*, *29*(3), 783–800. https://doi.org/10.1111/sode.12425

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology*, *61*(1), 129–141. https://doi.org/10.1080/17470210701508822

Ravid, D. (2019). First-language acquisition of morphology. In J. Fernández-Domínguez & M. Aronoff (Eds.), *Oxford research encyclopedia of linguistics*. Oxford University Press.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191-201.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595-626.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*(4), 829-856.

Reed, S. K. (2006). Cognitive architectures for multimedia learning. *Educational Psychologist*, *41*(2), 87-98. https://doi.org/10.1207/s15326985ep4102_2

Renoult, L., Davidson, P. S. R., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: At the crossroads of semantic and episodic memory. *Trends in Cognitive Sciences*, *16*(11), 550–558. https://doi.org/10.1016/j.tics.2012.09.003

Rhoades, D., & Sippel, J. (2010). Serial and concurrent presentations of stimuli and their effects on items recalled. *The Journal of Undergraduate Research*, *8*(1), 6. https://openprairie.sdstate.edu/jur/vol8/iss1/6

Richards, B. J. (1994). Child-directed speech and influences on language acquisition: Methodology and interpretation. In C. Gallaway & B. J. Richards (Eds.), *Input and Interaction in Language Acquisition* (pp. 74–106). Cambridge University Press. https://doi.org/10.1017/cbo9780511620690.006

Richardson, J. T. E. (2007). Measures of short-term memory: A historical review. *Cortex*, *43*(5), 635–650. https://doi.org/10.1016/s0010-9452(08)70493-3

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Jhon Wiley & Sons.

Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. ITL - *International Journal of Applied Linguistics*, *171*(2), 191-220.

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

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20-27.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814.

Roediger. H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn Sci*, *15*(1):20-7. https://doi.org/10.1016/j.tics.2010.09.003

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of personality and social psychology*, *35*(9), 677.

Rogers, T. V., Ellison, V. L., & Gaskins, R. B. (2015). The impact of quadratic functions instruction on students' understanding of parabolas. *Journal of Mathematical Behavior*, *39*, 142-155.

Rosselli, M., Ardila, A., Matute, E., & Vélez-Uribe, I. (2014). Language development across the life span: A neuropsychological/neuroimaging perspective. *Neuroscience Journal*, *2014*(1), 1–21. https://doi.org/10.1155/2014/585237

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. Studies in Second Language Acquisition, 21(4), 589-619.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rumbaugh, C. M., & Landau, J. D. (2018). Recognition and recall performance both benefit from the production effect with content-specific information. *Reading Psychology*, *39*(1), 29–40. https://doi.org/10.1080/02702711.2017.1361494

Rummel, J., & Kvavilashvili, L. (2023). Current theories of prospective memory and new directions for theory development. *Nature Reviews Psychology*, *2*, 40–54. https://doi.org/10.1038/s44159-022-00121-4

Ryan, L., Hoscheidt, S. M., & Nadel, L. (2008). Perspectives on episodic and semantic memory retrieval. In M. A. Hofer (Ed.), *Handbook of behavioral neuroscience* (pp. 5–616). Elsevier BV. https://doi.org/10.1016/s1569-7339(08)00201-4

Said, F., & Zhu, H. (2017). “No, no maama! Say *‘shaatir ya ouledee shaatir*’!” children’s agency in language use and socialisation. *International Journal of Bilingualism*, *23*(3), 771–785. https://doi.org/10.1177/1367006916684919

Saiegh–Haddad, E. L. I. N. O. R. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics*, *24*(3), 431-451.

Salim, J. A., & Mehawesh, M. (2014). Stages in language acquisition: A case study. *English Language and Literature Studies*, *4*(4), 16. https://zu.edu.jo/MainFile/Profile_Dr_UploadFile/Researcher/Files/ResearchFile_548%20_30_6.pdf

San Martin, C., & Paz, G. (2014). *Neural correlates of prospective memory: an EEG and ICA approach* [Doctoral dissertation, University of Glasgow]. University of Glasgow Repository. https://theses.gla.ac.uk/5850/

Santos, F. H., Bueno, O. F. A., & Gathercole, S. E. (2006). Errors in nonword repetition: Bridging short- and long-term memory. *Brazilian Journal of Medical and Biological Research*, *39*(3), 371–385. https://doi.org/10.1590/s0100-879x2006000300008

Saxton, M. (2009). The inevitability of child directed speech. In S. Foster-Cohen (Ed.), Language acquisition: Palgrave advances in linguistics (pp. 62-86). Palgrave Macmillan. https://doi.org/10.1057/9780230240780_4

Schacter, D. L. (2002). *The seven sins of memory: How the mind forgets and remembers*. HMH.

Schacter, D. L. (2012). Constructive memory: Past and future. *Cognitive Systems*, *14*(1), 7–18. https://doi.org/10.31887/dcns.2012.14.1/dschacter

Scheele, A., Leseman, P., & Mayo, A. (2009). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31*(1), 117–140. https://doi.org/10.1017/s0142716409990191

Schjølberg, S., Eadie, P., Zachrisson, H. D., Øyen, A.-S., & Prior, M. (2011). Predicting language development at age 18 months: Data from the norwegian mother and child cohort study. *Journal of Developmental & Behavioral Pediatrics*, *32*(5), 375–383. https://doi.org/10.1097/dbp.0b013e31821bd1dd

Schmidt, H. G. (1983). Problem-based learning: Rationale and description. *Medical education*, *17*(1), 11-16.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Second Language Teaching & Curriculum Center.

Schmidt, S. R. (1991). Can repetition enhance prior knowledge: Effects of repetition on incidental memory for details of meaningful narrative text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 826–837.

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

Schmitt, N. (2010). Researching vocabulary: A vocabulary research manual. Palgrave Macmillan.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, *53*(1), 109-120.

Schneider, W., & Niklas, F. (2017). Intelligence and verbal short-term memory/working memory: Their interrelationships from childhood to young adulthood and their impact on academic achievement. *Journal of Intelligence*, *5*(2), 26.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research*, *12*(1), 40-72.

Shanks, D., Brydges, R., Brok, W., Nair, P., & Hatala, R. (2013). Are two heads better than one? comparing dyad and self-regulated learning in simulation training. *Medical Education*, *47*(12), 1215-1222. https://doi.org/10.1111/medu.12284

Shapira-Lichter, I., Oren, N., Jacob, Y., Gruberger, M., & Hendler, T. (2013). Portraying the unique contribution of the default mode network to internally driven mnemonic processes. *Proceedings of the National Academy of Sciences*, *110*(13), 4950–4955. https://doi.org/10.1073/pnas.1209888110

Sharma, V. (2022). Letting the struggling saudi EFL readers take lead: How teachers transform english language instruction. *Journal of Language Teaching and Research*, *13*(3), 533–540. https://doi.org/10.17507/jltr.1303.09

Sharmin, S., & Chow, S. Y. (2020). The influence of text difficulty on eye movements and comprehension processes of young and adult readers. Journal of Eye Movement Research, 13(6), 1-14.

Shevlin, H. (2020). Current controversies in the cognitive science of short-term memory. *Current Opinion in Behavioral Sciences*, *32*, 148–154. https://doi.org/10.1016/j.cobeha.2020.02.005

Siegel, L. S. (2003). Basic cognitive processes and reading disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), Handbook of learning disabilities (pp. 158-181). Guilford Press.

Silvina Montrul. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. John Benjamins Publishing Company.

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours.

Sinyashina, E. (2020). “Incidental + intentional” vs “intentional + incidental” vocabulary learning: Which is more effective?. *Complutense Journal of English Studies*, *28*, 81–96. https://doi.org/10.5209/cjes.66685

Skinner, B. F. (1957). The experimental analysis of behavior. *American Scientist*, *45*(4), 343-371.

Skinner, T., Taylor, J., Alam, A., & Ali, R. (2021, October). Does learning method matter in cyber security behaviour? spaced vs. massed e-learning. *In 2021 8th International Conference on Behavioral and Social Computing (BESC)* (pp. 1-6). IEEE. https://doi.org/10.1109/besc53957.2021.9635235

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Smith, C. F., & Border, S. (2019). The twelve cranial nerves of Christmas: Mnemonics, rhyme, and anatomy – seeing the lighter side. *Anatomical Sciences Education*, *12*(6), 673–677. https://doi.org/10.1002/ase.1846

Smith, J. D., & Church, B. A. (2018). Dissociable learning processes in comparative psychology. *Psychonomic Bulletin & Review*, *25*(5), 1565-1584.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology*, *81*(1), 1-23.

Sok, S., & Han, Z. (2020). A study of L2 vocabulary acquisition under incidental and intentional conditions. *Vigo International Journal of Applied Linguistics*, *17*, 113–140. https://doi.org/10.35869/vial.v0i17.1467

Solimando, C. (2019). Web-Arabic as lingua franca (WALF): Variation and standard in teaching arabic as foreign language (TAFL). *Linguæ & - Rivista Di Lingue E Culture Moderne*, *17*(2), 99–112. https://doi.org/10.7358/ling-2018-002-soli

Sommers, M. S., & Danielson, S. M. (1999). Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context. *Psychology and Aging*, *14*(3), 458–472. https://doi.org/10.1037/0882-7974.14.3.458

Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall?. *Cognitive psychology*, *107*, 1-21.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. Language Learning, 60(2), 263-308.

Sparks, R., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2008). Early first-language reading and spelling skills predict later second-language reading and spelling skills. *Journal of Educational Psychology*, *100*(1), 162-174. https://doi.org/10.1037/0022-0663.100.1.162

Speciale, G., Ellis, N., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*(2), 293-321. https://doi.org/10.1017/s0142716404001146

Spencer, R. J., Reckow, J., Drag, L. L., & Bieliauskas, L. A. (2016). Incidental learning: A brief, valid measure of memory based on the WAIS–IV vocabulary and similarities subtests. *Cognitive and Behavioral Neurology*, *29*(4), 206-211.

Spivey, M., Cardon, C., & Schwieter, J. W. (2015). Methods for studying adult bilingualism. *The Cambridge handbook of bilingual processing*, 108-132.

Squire L. R, & Dede A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harb Perspect Biol 7*(3):a021667. doi: 10.1101/cshperspect.a021667.

Squire L. R. (1987). The organization and neural substrates of human memory. *Int J Neurol*. 21-22:218-22.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*(3), 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Squire, L. R., & Dede, A. J. O. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*, *7*(3), a021667. https://doi.org/10.1101/cshperspect.a021667

Stackhouse, J., & Wells, B. (1997). *Children's speech and literacy difficulties: A psycholinguistic framework*. Wiley.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207-222.

Stangor, C., & Walinga, J. (2019). 8.1 Memories as types and stages. *Introduction to Psychology*. https://openpress.usask.ca/introductiontopsychology/chapter/memories-as-types-and-stages/

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education*, *189*(1-2), 23–55. https://doi.org/10.1177/0022057409189001-204

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (pp. 73–92). Elsevier Inc. https://doi.org/10.1016/b978-0-12-770601-6.50010-3

Steenbergen, B., van der Kamp, J., Verneau, M., Jongbloed-Pereboom, M., & Masters, R. S. W. (2010). Implicit and explicit learning: Applications from basic research to sports for individuals with impaired movement dynamics. *Disability and Rehabilitation*, *32*(18), 1509–1516. https://doi.org/10.3109/09638288.2010.497035

Stevens, C., & Bavelier, D. (2012). The role of selective attention on academic foundations: A cognitive neuroscience perspective. *Developmental Cognitive Neuroscience*, *2*(1), S30–S48. https://doi.org/10.1016/j.dcn.2011.11.001

Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density.

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 115.

Straube, B. (2012). An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. *Behavioral and Brain Functions*, *8*(1), 35. https://doi.org/10.1186/1744-9081-8-35

Surrain, S. (2018). "Spanish at home, English at school": How perceptions of bilingualism shape family language policies among Spanish-speaking parents of preschoolers. *International Journal of Bilingual Education and Bilingualism*, *24*(8), 1–15. https://doi.org/10.1080/13670050.2018.1546666

Swanson, H. L. (1999). What develops in working memory? A life span perspective. *Developmental psychology*, *35*(4), 986.

Swanson, H. L. (2008). Working memory and intelligence in children: What develops?. *Journal of Educational Psychology*, *100*(3), 581.

Swanson, H. L. (2011). Working memory, attention, and mathematical problem solving: a longitudinal study of elementary school children. *Journal of Educational Psychology*, *103*(4), 821.

Swanson, M. R. (2020). The role of caregiver speech in supporting language development in infants and toddlers with autism spectrum disorder. *Development and Psychopathology*, *32*(4), 1–10. https://doi.org/10.1017/s0954579420000838

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 Years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296.

Syarifuddin, S. (2017). Deliberate learning of second language (L2) vocabulary: Research-Based guidelines. *KnE Social Sciences*, *1*(3), 210. https://doi.org/10.18502/kss.v1i3.741

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278. https://doi.org/10.1016/j.neuroimage.2013.08.023

Tawfiq, H. H. (2020). Investigating the sociocultural factors that affect learning a second language (on the example of English). *Language Learning & Technology*, *22*(3), 1–19. https://doi.org/10125/44651

Teng, M. F. (2023). Effectiveness of captioned videos for incidental vocabulary learning and retention: The role of working memory. *Computer Assisted Language Learning*, 1–29. https://doi.org/10.1080/09588221.2023.2173613

Theodoridou, K. (2011). Learning spanish withlaura: the effects of a pedagogical agent. *Educational Media International*, *48*(4), 335-351. https://doi.org/10.1080/09523987.2011.632280

Thomas, N. (2020). Incidental L2 vocabulary learning: Recent developments and implications for future research.

Thomure, H., Tamim, R., & Griffiths, M. (2021). *A report on the effect of Arabic language diglossia on teaching and learning* (pp. 1–97). Queen Rania Foundation for Education and Development. https://www.qrf.org/sites/default/files/2021-11/report_on_the_effect_of_arabic_language_diglossia_on_teaching_and_learning.pdf

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E. (2017). Implementing evidence based practice with limited evidence: the case of language intervention with bilingual children. *Revista de logopedia, foniatría y audiología*, *37*(4), 164-171.

Thorin, J., Sadakata, M., Desain, P., & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *Journal of the Acoustical Society of America*, *144*(1), 92–103. https://doi.org/10.1121/1.5044415

Thorn, A. S., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 729.

Tiffin-Richards, M. C., Hasselhorn, M., Woerner, W., Rothenberger, A., & Banaschewski, T. (2008). Phonological short-term memory and central executive processing in attention-deficit/hyperactivity disorder with/without dyslexia – evidence of cognitive overlap. *Journal of Neural Transmission*, *115*(2), 227–234. https://doi.org/10.1007/s00702-007-0816-3

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1412.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language learning*, *59*(3), 623-646.

Trembath, D., and Iacono, T. (2016). Standardized assessment of prelinguistic communication. In D. Keen, H. Meadan, N. Brady, J. Halle (Eds.), Prelinguistic and minimally verbal communicators on the autism spectrum. Springer. https://doi.org/10.1007/978-981-10-0713-2_5Y

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research:"Clozing" the gap. *Studies in Second Language Acquisition*, *33*(3), 339-372.

Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.

Tulving, E. (1984). Relations among components and processes of memory. *Behavioral and Brain Sciences*, *7*(2), 257–268. https://doi.org/10.1017/s0140525x00044617

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology*, *53*(1), 1-25.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of verbal learning and verbal behavior*, *5*(4), 381-391.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, *87*(1), 116–124. https://doi.org/10.1037/h0030186

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Tulving, E., & Watkins, M. J. (1973). Continuity between recall and recognition. *The American Journal of Psychology*, *86*(4), 739–748. https://doi.org/10.2307/1422081

Tunmer, W. E., & Hoover, W. A. (2017). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 175–214). Routledge. https://doi.org/10.4324/9781351236904

Twomey, K. E., Morse, A. F., Cangelosi, A., & Horst, J. S. (2016). Children's referent selection and word learning: Insights from a developmental robotic system. *Interaction Studies*, *17*(1), 101–127. https://doi.org/10.1075/is.17.1.05two

Tyler, A. E., and Ortega, L. (2018). Usage-inspired L2 instruction: An emergent, researched pedagogy. In A. E. Tyler, L. Ortega, M. Uno, & H. I. Park (Eds.), *Usage-inspired L2 instruction: Researched pedagogy* (pp. 3–26). John Benjamins.

Tymofiyeva, O., & Gaschler, R. (2021). Training-Induced neural plasticity in youth: A systematic review of structural and functional MRI studies. *Frontiers in Human Neuroscience*, *14*, 497241. https://doi.org/10.3389/fnhum.2020.497245

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559-599.

Uchihara, T., Webb, S., & Yanagisawa, A. (2023). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *73*(1), 183-224.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1-2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Underwood, B. J. (1957). Proactive inhibition as a function of time and degree of prior learning. *Journal of Experimental Psychology*, *54*(4), 297-302.

Uner, O., & Roediger, H. L. (2022). Do recall and recognition lead to different retrieval experiences? *The American Journal of Psychology*, *135*(1), 33–43. https://doi.org/10.5406/19398298.135.1.03

Unger, L., & Sloutsky, V. M. (2022). Selective attention and vocabulary development in early childhood. *Developmental Psychology*, *58*(4), 701-712.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104.

Valdés, G. (2005). Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized?. *The modern language journal*, *89*(3), 410-426.

Van Lier, L. (2004). The semiotics and ecology of language learning: Perception, voice, identity and democracy. *Utbildning & Demokrati*, *13*(3), 79–103. https://journals.oru.se/uod/article/download/783/772

van Riesen, S. A., Gijlers, H., Anjewierden, A., & de Jong, T. (2019). The influence of prior knowledge on experiment design guidance in a science inquiry context. *International Journal of Science Education*, *41*(11), 1528-1548.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System*, *41*(3), 609-624.

VanPatten, B. (2020). Input processing in adult L2 acquisition. In *Theories in second language acquisition* (pp. 105-127). Routledge.

Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: How to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *30*(6), 1322–1337. https://doi.org/10.1037/0278-7393.30.6.1322

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217-234.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. Language Learning, 61(1), 219-258.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219-258.

Vigliocco, G., Tranel, D., & Druks, J. (2012). Language production. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 443–465). Cambridge University Press.

Vihman, M. M. (2016). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*(1), 1–27. https://doi.org/10.1111/bjop.12207

Vinci-Booher S, Cheng H, James KH. (2019). An Analysis of the Brain Systems Involved with Producing Letters by Hand. *Journal of Cogn Neurosci*, *31*(1):138-154. https://doi.org/10.1162/jocn_a_01340

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld.*Psychologische Forschung*, 18, 299–342.

Voss, J. L, Paller, K. A. (2008). Brain substrates of implicit and explicit memory: the importance of concurrently acquired neural signals of both memory types. *Neuropsychologia*, *46*(13):3021-9. https://doi.org/10.1016/j.neuropsychologia.2008.07.010

Vyas, P., & Sharma, S. (2022). Socio-Demographic and psychological predictors of second language achievement: A systematic review. *TESL-EJ*, *26*(1), 1–35. https://eric.ed.gov/?id=EJ1348813

Wager, T. D. (2015). *How your brain makes memories: A neuroscientist explains the biology of memory and how we can improve it*. Penguin Books.

Wang, L., Xu, Z., Si, Y., Cao, L., & Liu, Y. (2017). Context-Associative hierarchical memory model for human activity recognition and prediction. *IEEE Transactions on Multimedia*, *19*(3), 646–659. https://doi.org/10.1109/tmm.2016.2617079

Wang, Y., & Gennari, S. P. (2019). How language and event recall can shape memory for time. *Cognitive Psychology*, *108*, 1–21. https://doi.org/10.1016/j.cogpsych.2018.10.003

Wang, Y., Yin, L., & McBride, C. (2021). Unique predictors of early reading and writing: A one-year longitudinal study of Chinese kindergarteners. *Journal of Learning Disabilities*, 54(3), 170-179.

Watkins, M. J. (1975). Inhibition in recall with extralist "cues". *Journal of Verbal Learning and Verbal Behavior*, *14*(3), 294-303.

Watkins, M. J., & Todres, A. K. (1978). On the relation between recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *17*(5), 621–633. https://doi.org/10.1016/s0022-5371(78)90381-x

Wattenmaker, W. D. (1993). Incidental concept learning, feature frequency, and correlated properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(1), 203-222.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46-65.

Webb, S., & Chang, A. C-S. (2015a). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667-686.

Webb, S., & Chang, A. C-S. (2015b). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition, 37*(4), 651-675.

Webb, S., & Nation, P. (2017). How vocabulary is learned. In P. Nation & S. Webb (Eds.), *Researching vocabulary: A vocabulary research handbook* (pp. 1-20). Routledge.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

White, C. N., Ratcliff, R., & Starns, J. J. (2013). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology, 67*(3), 135-168.

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, *121*(4), 577-590.

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.

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

Wilkinson, S. (1998). Study abroad from the participants' perspective: A challenge to common beliefs 1. *Foreign Language Annals*, *31*(1), 23-39.

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*(1), 1–48. https://doi.org/10.1017/s0272263199001011

Willoughby, H. V. (2020). *The pupillometric production effect: Measuring attentional engagement during a production task* [Doctoral dissertation, Memorial University of Newfoundland]. MU Library. https://research.library.mun.ca/14757/

Wilson, M., & Fox, G. (2007). Working memory for language is not special: Evidence for an articulatory loop for novel stimuli. Psychonomic Bulletin & Review, 14(3), 470–473.

Witherby, A. E., & Carpenter, S. K. (2022). The rich-get-richer effect: Prior knowledge predicts new learning of domain-relevant information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(4), 483–498. https://doi.org/10.1037/xlm0000996

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, *114*(1), 152.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585. https://doi.org/10.1017/s0142716407070312

Worcester, G. (2013). Effective reading comprehension strategies for individuals with learning disabilities: A research synthesis. *LC Journal of Special Education*, *8*(1), 11. https://digitalshowcase.lynchburg.edu/lc-journal-of-special-education/vol8/iss1/11

Woumans, E., Clauws, R., & Duyck, W. (2021). Hands down: Cognate effects persist during written word production. *Frontiers in Psychology*, *12*, 647362. https://doi.org/10.3389/fpsyg.2021.647362

Wu, Y. J., & Thierry, G. (2010). Investigating bilingual processing: The neglected role of language processing contexts. *Frontiers in Psychology*, *1*, 178. https://doi.org/10.3389/fpsyg.2010.00178

Xiaoning, C., & Feng, T. (2017). Assessing the effects of word exposure frequency on incidental vocabulary acquisition from reading and listening. *Chinese Journal of Applied Linguistics, 40*(1), 56-73.

Yaccob, N. M., Noor, N. M., & Baharudin, H. (2022). The role of visual input enhancement in the acquisition of phrasal verbs among EFL learners. *System*, *4*, 102705.

Yaccob, N. S., Abd. Rahman, S. F., Azlan Mohamad, S. N., Abdul Rahim, A. A., Khalilah Abdul Rashid, K., Mohammed Abdulwahab Aldaba, A., Md Yunus, M., & Hashim, H. (2022). Gamifying ESL classrooms through gamified teaching and learning. *Arab World English Journal*, *8*(8), 177–191. https://doi.org/10.24093/awej/call8.12

Yang, H. (2022). Facilitating vocabulary learning by "3r"-- recycling, repetition and retrieve. *Learning & Education*, *10*(8), 175–176. https://doi.org/10.18282/l-e.v10i8.3111

Yang, J., & Li, P. (2012). Brain networks of explicit and implicit learning. *PLoS ONE*, *7*(8), e42993. https://doi.org/10.1371/journal.pone.0042993

Yaqubi, B., Rayati, R. A., & Gorgi, N. A. (2012). The involvement load hypothesis and vocabulary learning: The effect of task types and involvement index on l2 vocabulary acquisition. *Journal of Teaching Language Skills*, *2*(1594), 145–163. https://doi.org/10.22099/jtls.2012.404

Yazan, B., & Ali, I. (2018). Family language policies in a Libyan immigrant family in the U.S. *Heritage Language Journal*, *15*(3), 369–387. https://doi.org/10.46538/hlj.15.3.5

Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research*, *33*(4), 519–549. https://doi.org/10.1177/0267658317708009

Ylinen, S., Nora, A., & Service, E. (2020). Better phonological short-term memory is linked to improved cortical memory representations for word forms and better word learning. *Frontiers in Human Neuroscience*, *14*, 209. https://doi.org/10.3389/fnhum.2020.00209

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of experimental psychology: Learning, memory, and cognition*, *20*(6), 1341.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience*, *25*(11), 3002–3008. https://doi.org/10.1523/jneurosci.5295-04.2005

Yonelinas, A. P., Kroll, N. E., Dobbins, I., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: convergence of remember-know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, *12*(3), 323.

Yonelinas, A., Ramey, M., & Riddell, C. (2022). Recognition memory: The role of recollection and familiarity. In *The Oxford handbook of human memory* (pp. 1–41). Springer. https://hmlpubs.faculty.ucdavis.edu/wp-content/uploads/sites/214/2022/05/2022_Yonelinas.pdf

Young, J. Q., Van Merrienboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive load theory: Implications for medical education: AMEE guide no. 86. *Medical Teacher*, *36*(5), 371–384. https://doi.org/10.3109/0142159X.2014.889290

Young, T. A., Ricks, P. H., & MacKay, K. L. (2023). Engaging students with expository books through interactive read-alouds. *The Reading Teacher*, *77*(1), 6–15. https://doi.org/10.1002/trtr.2210

Zaidan, O. F., & Callison-Burch, C. (2014). Arabic dialect identification. *Computational Linguistics*, *40*(1), 171-202.

Zamuner, T. S., Fikkert, J. P. M., & Gick, B. (2007). Production facilitates lexical acquisition in young children. *The Journal of the Acoustical Society of America*, *122*(5), 3002–3036. https://repository.ubn.ru.nl/bitstream/handle/2066/44580/44580.pdf

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. A. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, *89*, 55–67. https://doi.org/10.1016/j.jml.2015.10.003

Zamuner, T. S., Strahm, S., Morin-Lessard, E., & Page, M. P. A. (2017). Reverse production effect: Children recognize novel words better when they are heard rather than produced. *Developmental Science*, *21*(4), e12636. https://doi.org/10.1111/desc.12636

Zhang, J. (2019, May 30). *Cognitive functions of the brain: Perception, attention and memory*. ArXiv.org. https://doi.org/10.48550/arXiv.1907.02863

Zhang, X., & Samuel, A. G. (2018). Is speech recognition automatic? Lexical competition, but not initial lexical access, requires cognitive resources. *Journal of Memory and Language*, *100*, 32–50. https://doi.org/10.1016/j.jml.2018.01.002

Zlotnik, G., & Vansintjan, A. (2019). Memory: An extended definition. *Frontiers in Psychology*, *10*, 1664–1078. https://doi.org/10.3389/fpsyg.2019.02523

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*(3), 340–352. https://doi.org/10.1080/09658211.2018.1510966

Zuber, S., & Kliegel, M. (2020). Prospective memory development across the lifespan. *European Psychologist*, *25*(3), 162–173. https://doi.org/10.1027/1016-9040/a000380

Zung, T. K., Plaut, D. C., & Perfetti, C. A. (2022). A computational model of the self-teaching hypothesis for learning to read. *Cognitive Science*, *46*(4), e13123.

Zychowicz, K., Biedroń, A., & Pawlak, M. (2018). Polish nonword span (PNWSPAN): A new tool for measuring phonological loop capacity. *Glottodidactica: An International Journal of Applied Linguistics*, *45*(2), 309–327. https://doi.org/10.14746/gl.2018.45.2.18

Appendices

Appendix 1. Arabic consonants with IPA symbols

Character	IPA	Character	IPA
ا	a, a:	ض	dˤ
ب	b	ظ	zˤ
ت	t	ع	ʕ
ث	θ	غ	ɣ
ج	d͡ʒ	ف	f
ح	ħ	ق	q
خ	x	ك	k
د	d	ل	l
ذ	ð	م	m
ر	r	ن	n
ز	z	ه	h
س	s	و	w, u:
ش	ʃ	ي	y, i, i:
ص	sˤ	ء	ʔ
ط	tˤ	ة	æ

Appendix 2. Participants' Information Sheet

Project Title: **The Production Effect on the Acquisition of Arabic Language**

Department of English Language and Linguistics

University of Birmingham, Edgbaston

Researchers

Nouf Alharbi

Dr Gareth Carrol

Dr Bene Bassetti

Thank you for taking an interest in this project which aims to understand how children learn language. More specifically, I am interested in whether children will learn a language better when they speak it out loud or write it, and what will help them remember what they have learned.

The project involves one visit in which the children will play a vocabulary game for 20 minutes. Due to the COVID-19 pandemic, it may be necessary to complete the study using ZOOM (video chat software).

If the session takes place in person, your child will be introduced to a vocabulary game on the computer. They will hear a word, then be asked to either not respond, say it out loud or write it down. Later, they will be asked which of the words they can remember by showing them pictures. During this, their eye movements will be recorded (a camera will record what they look at on the computer screen). During the visit, I will bring all of the equipment necessary with me, including a laptop and some toys (LEGO bricks) for the children to play with midway through the session. I will be present during all play sessions, and a teacher will be waiting outside the classroom, so the child is not distracted with her presence. If the session is conducted online via Zoom, their eye-movements will not be recorded; however, the Zoom session will be recorded on the researcher's secured laptop. During the session, I will provide an online game for the children to play with midway through the session. I will be present during all play sessions, and your presence should not be close to your child, so the child is not distracted. However, you must remain close to supervise your child at all times of the study.

Your child's name will not be recorded, and all children will be assigned a code number so that we can analyze the data. Please keep this ID Code in case you decided to withdraw from the study. Any personal information provided in the questionnaire will not be published in a way that will identify any of the participants. The data will only be accessible by myself, my supervisors, Dr Gareth Carrol, and Dr Bene Bassetti, and the marker of this project. The anonymized data collected in this project will be used in the researcher's PhD thesis, and may also be used in the future, including in published work. For you and the children's participation you will receive no direct compensation. However, a summary of results will be available to you and to the School on request. Your participation is voluntary, and you have the right to withdraw your consent or remove your child from the study at any time without any penalty. You and your child have the right to refuse to do particular tasks, and if your child decides not to complete the study then his/her data will not be used. You may decide to withdraw your child's data for up to two weeks following the date on which the study is completed. Please email the researcher if you wish to do this. It is also your right to ask questions or for clarifications before agreeing to take part in this study, or before the study begins.

Appendix 3. Consent Form

Project Title: **The Production Effect on the Acquisition of Arabic Language**

Department of English Language and Linguistics

University of Birmingham, Edgbaston

Researchers

Nouf Alharbi

Dr Gareth Carrol

Dr Bene Bassetti

Participant Details

Child's name:
Parent/Guardian's name:

Contact telephone number:

By signing below, you indicate that you have read and understood the project information sheet. Please tick the following boxes to indicate your consent:

I have explained to my child about the project and they agree to participate. ☐

I agree to allow my child to participate in this project. ☐

I have been given the opportunity to ask questions about the project. ☐

I understand that data from this session will be recorded. ☐

I have been fully informed about the study in the form of the participant information sheet and the contact details of researcher(s). ☐

I understand that my taking part is voluntary. I can withdraw from the study within 14 days from participating in the study, or when the child decides not to complete the game. I do not have to give any reasons for why I no longer want to take part. ☐

I understand that my data is and will remain anonymous, and that anonymized data may be used as part of a future scientific publication or shared with third-party researchers. ☐

________________________   ________________________

Signature of parent/ guardian   Signature of researcher

Appendix 4. Language Background Questionnaire

ID number:……………

- Are you the ☐ parent or ☐ caretaker of the child for whom you are completing this survey? (please select one)
  Email: …………………………………………………
- What is your child's gender?

  ☐ Male

  ☐ Female

  ☐ Non-binary

  ☐ Other (please specify): ………………………………
- What is your child's date of birth? ……….. / …………. / ……………
  Day Month Year
- What is your child's current grade in the British school? ……………….
- What is your child's current grade in the Arabic school? ……………….
- What are some reasons why your child should learn Arabic?

1. ………………….…………………….……………….
2. ………………….…………………….……………….
3. ………………….…………………….……………….

- How many hours per week does your child speak in Arabic? …………………
- How many hours per week does your child watch TV in Arabic? …………………

- Please select the options below that pertain to your child.

Questionnaire statement	**Yes**	**No**	**Percentage of time spent on this activity (compared to English and other languages)**
Can speak Arabic			
Can read Arabic			
Can write in Arabic			
Can understand instructions given in Arabic			

- Which languages are regularly spoken at home?

…………………………………………………………………………………………

- What is the **main** language spoken at home?

……………………………………………………………………………………

- Does your child have a language difficulty?

☐ Yes

☐ No

- If yes, please describe it: ……………………………………………………………

- Was this language difficulty diagnosed by a speech and language therapist?

☐ Yes

☐ No

- Has your child ever attended speech and language sessions?

☐ Yes

☐ No

- If yes, what was the duration of the sessions (e.g. the number of weeks attended)?

…………………………………………

- Is your child in a special education class at school?

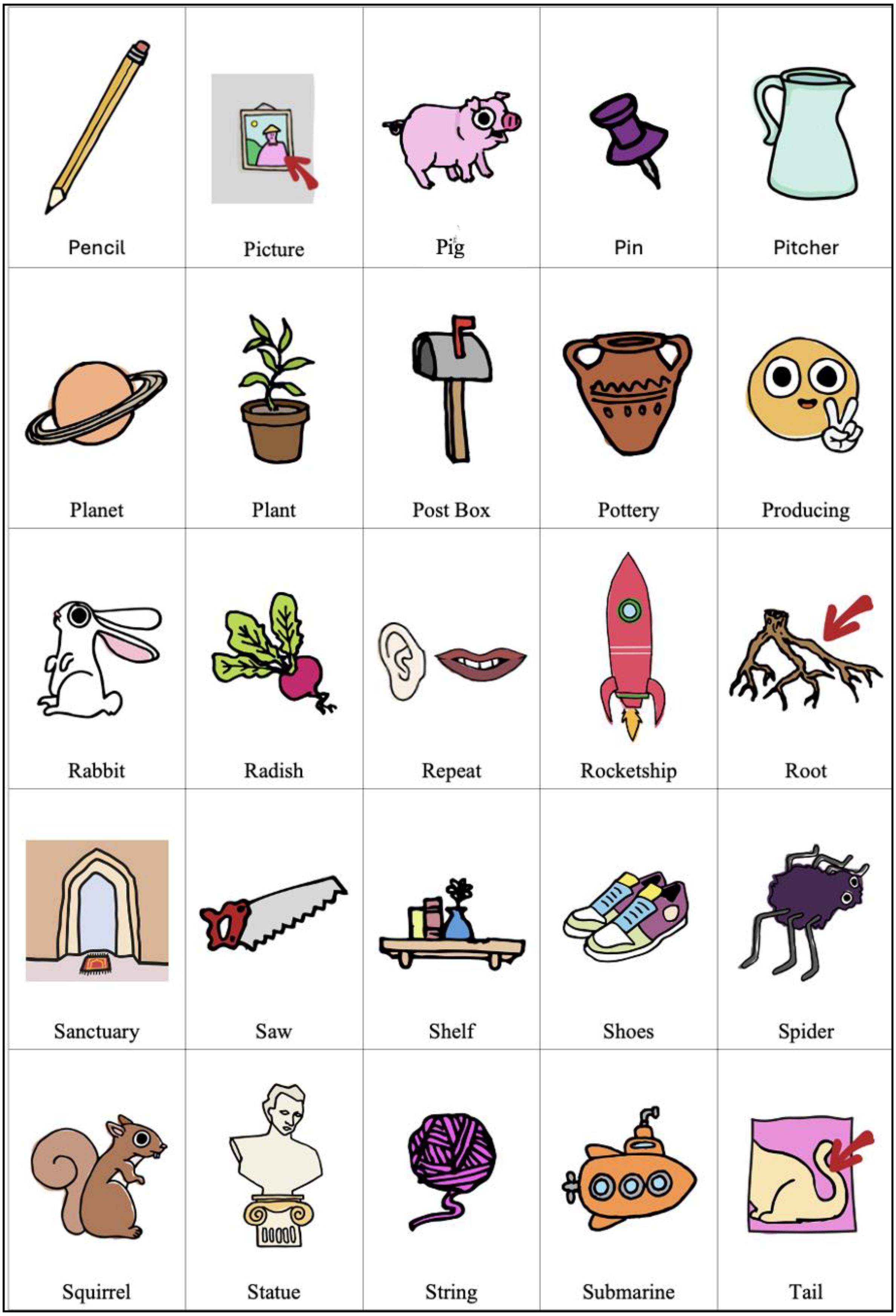
☐ Yes

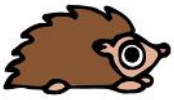
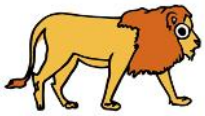
☐ No

- If yes, what type of class? ....

Thank you for your participation!

Appendix 5. Visual Stimuli



 Heart	 Hedgehog	 Honey	 House	 Hyena
 Kite	 Ladder	 Lion	 Listen	 Listen (1)
 Listen (2)	 Magnet	 Metal Plate	 Milk	 Mission Accomplished
 Mole	 Mud	 Needle	 Newspaper	 Oud
 Owl	 Ox	 Palm Tree	 Peacock	 Pen


Closet


Coals


Copy Book

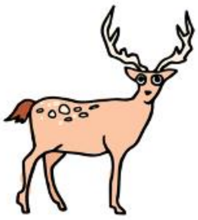

Cord


Cottage


Crocodile


Crown

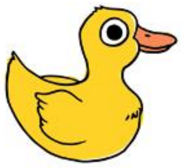

Cup


Deer


Dog


Dragonfly


Dress


Duck


Envelope


Fig


Fireplace


Flag


Flower


Frame


Frog


Fur


Garment


Goose


Grasshopper


Hearing


Air Balloon


Almond


Apple


Arm


Arrow


Axe


Bag


Banana


Bat

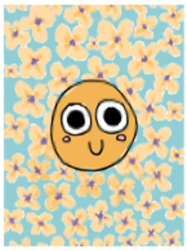

Bee


Blueberry


Book


Brace Drill

Break (1)


Break (2)

Break (3)


Butterfly


Cactus


Calf


Car


Cat


Chain


Chair


Cherry


Clock


Talking


Tambourine


Tap


Tent


Tree


Truck


Tube


Vein


Volcano


Wasp


Whale


Window

Appendix 6. A Report on Emotional Effects on Parents' Motivation and Learners' Engagement (Appeared in BAAL Newsletter (P.31))

PhD Report - I don't want to say it: The emotional effects of COVID-19 on parent's motivation and young learners' online engagement

By Nouf Alharbi (University of Birmingham)

Gamification has become an interactive method to teach children new vocabulary thanks to the ease of using online games in education, especially in a home learning environment. Despite the rise in implementing such methods with primary students, data collection has different stories to tell during the Covid-19 pandemic, in which less attention has been given to the emotional reasons driving both parents' and children's decision to withdraw from a study. In this light, the exclusionary factor that affected the interaction of child learners during a 6-minute online Arabic language game was examined. In particular, children were asked to listen only or listen to and repeat 20 unfamiliar and less frequent Modern Standard Arabic words. During the game, the parent's role was to assist their children in understanding the instructions that appeared on the screen without interfering in the learning process. However, the parents reported that children were bored and hesitant to play the game, and that the parents themselves were frustrated while instructing the children so they refused to complete the task. To gain insight into hesitancy in completing the short task, semi-structured interviews were conducted with the parents and their children. It became clear that the pressure of online learning during the pandemic enhanced frustration in children, and that spending more time on screen contributed to demotivating and distressing parents when assessing the children's distance learning. This finding shed light on the need to analyse several observable and not directly observable factors that influenced children and parents' emotional wellbeing. But the question here is *what role do I play as a researcher in supporting my little participants and their families?*

To answer this question, I need to first outline some of the challenges that I faced while conducting this online study. I should mention that finding the right time and being flexible about it helped a lot when recruiting participants. Also, not all families were comfortable having me explain the game via zoom. It was already sufficiently challenging for their children to attend daily lessons on the computer. Therefore, I offered the parents some guidance and ensured they understood the instructions and left it for them to explain the game and assess their children, and I remained available for any support they might need. One might argue that the role I played added a lot of pressure on myself as a PhD student. However, the feeling of being supportive and offering some fun to these families showed me that this study does not only offer a main chapter in the thesis but also represents a valuable time in the chapter of others' lives. I'm not a wellbeing expert, but I drew on these positive thoughts to keep me going. *What about those who were not happy to join the study?* The little ones who were crying and shouting that they do not want to play anymore but their parents were bribing them with a chocolate bar or extra TV time to telling me that they wanted to withdraw. While they had been informed of the right to withdraw from the study at any time, the guilt and shame of not offering me the help I needed stopped them from withdrawing. As a researcher, I talked to parents first showing them that I'm considerate and that their children's happiness is much more important that being a participant in my study. I offered them colouring sheets as a thank-you letter for trying to participate and organised a virtual teatime to chat with the parents as a mother not just as a researcher.

Conducting research with huma beings became much more complicated during the pandemic when everyone was thinking about their safety while trying to live a normal life. The lesson I took from the pandemic is about being more human, with myself as a researcher, and with my participants. Being compassionate is equally important, if not more, to obtain the necessary ethical approval to conduct your study!