Speech Production Effects on Word Learning and Memorization In Children

Maria Belén López

Supervisor: Dr. Tania S. Zamuner

A thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the Doctorate in Philosophy degree in Linguistics

Department of Linguistics

Faculty of Arts

University of Ottawa

Abstract

In this dissertation, I explore the impact of speech production on memory and learning in children, by investigating the Production Effect and the Reverse Production Effect. In the Production Effect, a memory advantage is observed for items that are produced aloud compared to items studied under other actions, such as reading silently or listening. Conversely, the Reverse Production Effect is a memory advantage for items studied under listening conditions, compared to producing aloud. Despite extensive research showing the Production Effect with adults, its manifestation and underlying mechanisms in children remains largely unexplored. This dissertation focuses on children, covering an age range from 2-to-6-years-old across all experiments combined. Different methodologies were used to study memory and learning in monolingual English-speaking children, specifically free recall tasks, identification tasks, and recognition tasks using the Visual World Paradigm with eye-tracking.

This dissertation presents four articles that investigate different aspects of the Production Effect and Reverse Production Effect, with a larger goal of examining how children's developing cognitive and linguistic abilities may influence the effect of speech production on memory and learning. The first article used a developmental approach to investigate whether the Production Effect changes across development, using a free recall task with real words. While older children (4-to-6-years-old) were able to benefit from speech production, younger children (2-to-3-years-old) showed a Reverse Production Effect, suggesting that the developmental and linguistic stage of the learner interacts with the Production Effect. The second study investigated whether the Production Effect found in children (5-to-6-years-old) remains after a 1-week-delay. Results show that for children who did show the Production Effect initially , this effect was not present when tested on the same words after 1 week, suggesting that the memory advantage is restricted in time. The third article investigated whether the Reverse Production Effect found in children (5-and-6-year-olds) is specific to speech production, or whether other actions would also disrupt learning. All actions patterned similarly, showing a Reverse Production Effect, indicating that this is not solely triggered by speech production, but by performing any action while learning. The fourth article investigated whether a more familiar context (storybook reading) and multiple exposures to the novel items would increase the chance of the Production Effect emerging in a word learning task with children aged 4-and-6-years. Results patterned with previous Reverse Production Effect findings, with further effects of children's age and the number of exposures during training.

These findings deepen our understanding of how speech production impacts memory encoding and retrieval across different stages of development. By examining the Production Effect and Reverse Production Effect in children, this dissertation contributes to psycholinguistic theories of language acquisition and memory, offering valuable insights into learning strategies from early childhood. Results from this dissertation show that production and perception influence memory in distinct ways. Practical implications include potential applications in educational settings.

Keywords: Production Effect, speech production, memory, learning, cognitive development, language development, eye-tracking

Co-Authorship Statement

The chapters of this dissertation are manuscripts that have been prepared for submission to scientific journals. The presented data are based on a series of collaborative research projects; however, all manuscripts have been primarily written by Belén López Assef. Chapter 2 was published by *Child Development* in 2021, authored by Belén López Assef, Félix Desmeules-Trudel, Amélie Bernard and Tania S. Zamuner. Chapter 3 was published by *Applied Psycholinguistics* in 2023, authored by Belén López Assef, Margarethe McDonald, Amélie Bernard, and Tania S. Zamuner. Chapter 4 has been accepted by *Journal of Child Language*, authored by Belén López Assef and Tania S. Zamuner. Chapter 5 has been written in preparation as a submission to *Developmental Psychology*.

Acknowledgements

I'd first like to thank my wonderful supervisor, Dr. Tania Zamuner, your unwavering support, patience, and dedication have been instrumental in my academic journey. You have not only provided me with invaluable opportunities but have also shared your knowledge, expertise and time with me, guiding me with understanding and kindness through every challenge. Doing my PhD as an international student during the pandemic was not an easy feat, but your mentorship made this journey not only bearable but rewarding (and fun!). Your encouragement kept me motivated during the toughest times, and your feedback pushed me to achieve more than I ever thought possible. I couldn't have asked for a better mentor. Thank you for all the wonderful times, hours long Zoom calls and joint suffering studying statistics.

I would like to express my deepest gratitude to everyone who made my time at uOttawa an unforgettable and amazing experience. Theresa Rabideau (could not have asked for a better office buddy), Scarlet Li and Margarethe McDonald - Your friendship and support made this so much more enjoyable. I am incredibly grateful to have shared this journey with the most amazing friends. Can't wait to see what the future has for us! Thank you to all of the grad students, post-docs, research assistants and friends at the Centre for Child Language Research and uOttawa Living Lab. To Rachel Miller, Daniela Gallardo and Nell Ramesh – thank you for being the best research assistants ever and putting so much effort into my research. To my wonderful roommates at 33, thank you for being my family away from home. Your support, love, and countless moments of laughter and comfort helped make my time in Canada truly special. I am so grateful to have shared this chapter of my life with such amazing people.

También me gustaría agradecer a todos los amigos y seres queridos que apoyaron mi decisión de venir a Canadá para realizar estudios de posgrado. No habría podido viajar (ni quedarme) sin su ayuda, apoyo e inmenso amor. Y por último, mis gracias más grandes en el universo van para mi Mamá, Papá, Pancho y Monse, por su apoyo, amor y agüante eterno, por darme fuerza y valentía , y por siempre escuchar todas mis quejas. No podría haber logrado nada sin ustedes.

Table of Content

1 GENERAL INTRODUCTION 1

1.1 THE PRODUCTION EFFECT 4

1.1.1 EXPLANATIONS FOR THE PRODUCTION EFFECT 4

1.2 THE REVERSE PRODUCTION EFFECT 7

1.3 PRODUCTION EFFECT AND REVERSE PRODUCTION EFFECT IN CHILDREN 9

1.4 COGNITIVE AND DEVELOPMENTAL FACTORS 11

1.5 THE PRESENT DISSERTATION 13

1.5.1 FIRST AIM OF THE CURRENT THESIS (CHAPTER 2): CHANGE ACROSS DEVELOPMENT 13

1.5.2 SECOND AIM OF THE CURRENT THESIS (CHAPTER 3): IMMEDIATE VS DELAYED TESTING 14

1.5.3 THIRD AIM OF THE CURRENT THESIS (CHAPTER 4): NATURE OF THE REVERSED PRODUCTION EFFECT 14

1.5.4 FOURTH AIM OF THE CURRENT THESIS (CHAPTER 5): EXPANDING THE LEARNING TASK 15

REFERENCES 16

2 A SHIFT IN THE DIRECTION OF THE PRODUCTION EFFECT IN CHILDREN AGED 2-6 YEAR 22

2.1 INTRODUCTION 23

2.2 EXPERIMENT 1 32

2.2.1 METHOD 34

2.2.2 RESULTS 40

2.2.3 DISCUSSION 44

2.3 EXPERIMENT 2 45

2.3.1 METHOD 45

2.3.2 RESULTS 46

2.3.3 DISCUSSION 49

2.4 GENERAL DISCUSSION 50

REFERENCES 57

3 EFFECTS OF SPEECH PRODUCTION TRAINING ON MEMORY ACROSS SHORT AND LONG DELAYS IN 5- AND 6- YEAR-OLDS: A PRE-REGISTERED STUDY 65

ABSTRACT 65

3.1 INTRODUCTION 66

3.1.1 THE PRODUCTION EFFECT: ADULTS 67

3.1.2 DELAYED TESTING: ADULTS 69

3.1.3 THE PRODUCTION EFFECT: CHILDREN 69

3.1.4 DEVELOPMENTAL DIFFERENCES IN MEMORY 71

3.2 METHOD 74

3.2.1 PARTICIPANTS 74

3.2.2 STIMULI 76

3.2.3 DESIGN 76

3.3 RESULTS 79

3.3.1 CODING 79

3.3.2 FREE RECALL TASK 80

3.3.3 OLD/NEW RECOGNITION TASK 85

3.3.4 EXPLORATORY ANALYSIS: VOCABULARY SCORE 88

3.4 DISCUSSION 92

REFERENCES 99

4 TASK EFFECTS IN CHILDREN'S WORD RECALL: EXPANDING THE REVERSE PRODUCTION EFFECT 107
ABSTRACT 107
4.1 INTRODUCTION 108
4.2 METHODS 113
4.2.1 PARTICIPANTS 113
4.2.2 STIMULI 113
4.2.3 DESIGN 114
4.2.4 PROCEDURE 116
4.3 RESULTS 116
4.3.1 CODING 116
4.3.2 FREE RECALL TASK 117
4.3.3 RECOGNITION TASK 120
4.4 DISCUSSION 123
REFERENCES 127
5 WORD LEARNING IN A STORYBOOK SETTING: RECOGNITION IS BETTER FOR YOUNGER CHILDREN WHEN THEY SIMPLY LISTEN RATHER THAN PRODUCE ALOUD 134
5.1 INTRODUCTION 136
5.1.1 FACTORS IN THE PRODUCTION EFFECT AND REVERSE PRODUCTION EFFECT 137
5.1.2 THE PRODUCTION EFFECT AND REVERSE PRODUCTION EFFECT WITH CHILDREN 138
5.1.3 STORYBOOK METHODOLOGY 140
5.1.4 REPEATED EXPOSURE DURING WORD-LEARNING 143

5.2 METHOD **145**

5.2.1 PARTICIPANTS 145

5.2.2 STIMULI 146

5.2.3 DESIGN 147

5.2.4 PROCEDURE 147

5.3 RESULTS **149**

5.3.1 TRANSPARENCY AND OPENNESS 149

5.3.2 CODING 149

5.3.3 RECOGNITION TASK 149

5.3.4 POINTING TASK 154

5.4 DISCUSSION **156**

REFERENCES **161**

6 GENERAL DISCUSSION AND FUTURE DIRECTIONS **172**

6.1 SUMMARY OF THE FINDINGS **172**

6.2 MODELLING THE PRODUCTION EFFECT **173**

6.1 LIMITATIONS OF THE CURRENT RESEARCH **187**

6.1 IMPLICATIONS AND FUTURE DIRECTIONS **188**

REFERENCES **191**

APPENDIX A – STUDY INSERTS **197**

List of Tables

Table 1 *Summary of training conditions in studies with children on the Production Effect* 3

Table 2 *Average percentages of recalled words (number of recalled words over the number of correctly trained words) and standard deviation (in parentheses) by word-training condition for each age group [in years].* 41

Table 3 *Numerical output of the GAMM for recall data in children aged three to six years.* 44

Table 4 *Average percentages of recalled words (number of recalled words over the number of correctly trained words) and standard deviation (in parentheses) by word-training condition for children aged 2 years.*........................47

Table 5 *Numerical output of the GAMM for recall data in children aged 2 years.* 47

Table 6 *Results from Linear Mixed Model Estimating Free Recall by Training Condition (Listen, Say), Test Session (Immediate, Delayed), And Training Order (Listen In Block 2, Say In Block 2).* 82

Table 7 *Post-Hoc Tests of the Estimated Marginal Means for Significant 2-Way Interaction from the Model Predicting Proportion of Recall by Training Condition (Listen, Say) and Training Order (Listen in Block 2, Say in Block 2).* 83

Table 8 *Post-Hoc Tests of the Estimated Marginal Means for Significant 3-Way Interaction from the Model Predicting Proportion of Recall by Training Condition (Listen, Say), Test Session (Immediate, Delayed), and Training Order (Listen in Block 2, Say in Block 2).*........................85

Table 9 *Results from Linear Mixed Model Estimating Recognition Accuracy by Training Condition (Listen, Say) and Training Order (Listen in Block 2, Say in Block 2).* 87

Table 10 *Recalls and proportion of target fixations average and standard deviation by Training Condition (Verbal/Speech– Heard, Verbal/Non-Speech– Heard, Non-Verbal/Non-Speech– Heard).* 117

Table 11 *Results from model estimating free recall by Training Condition (Verbal/Speech– Heard, Verbal/Non-Speech– Heard, Non-Verbal/Non-Speech– Heard) and Training Order (Heard in Block 1, Action in Block 1).* 119

Table 12 *Results from model estimating proportion of target fixations by Training Condition (Heard – Verbal/Speech, Verbal/Non-Speech, Non-Verbal/Non-Speech) and Training Order (Heard in Block 1, Action in Block 1).* 122

Table 13 *Model estimating number of fixations to target by Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations) and Age (4yrs, 6yrs).* 151

Table 14 *Post-hoc tests of the estimated marginal means for significant three-way interaction from the model predicting Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations) and Age (4yrs, 6yrs).* 153

Table 15 *Model estimating pointing accuracy by Training Condition (Heard, Say), Number of Exposures (4 Presentations, 1 Presentation) and Age (4yrs, 6yrs).* 156

Table 16 *Summary of studies with children on the Production Effect (PE) and Reverse Production Effect (RPE)* 178

List of Figures

Figure 1 *Effect of TRAINING condition and AGE (months) on percentages of correct recall for 3- to 6-year-old children in Experiment 1.* 42

Figure 2 *Difference curves (GAMM output) for probability of recall through AGE by TRAINING condition.* 43

Figure 3 *Effect of TRAINING condition on percentages of correct recall for 2-year-old children in Experiment 2* 47

Figure 4 *Production Effect Scores by age (in months), i.e., number of recalled words that were SAY minus that were LISTEN, for all children in Experiments 1 and 2, and regression line (linear).* 49

Figure 5 *Proportion of Recall by Training Condition (Listen, Say) and Test Session (Immediate, Delayed).* 81

Figure 6 *Mean Recall by Training Condition, Test Session, and Training Order.* 84

Figure 7 *Proportion of Correct Responses During Old/New Recognition Task by Training Condition (Listen, Say).* 87

Figure 8 *Proportion of Recall by Training Condition (Listen, Say) and Standardized Vocabulary Score.* 89

Figure 9 *Example of training and testing trials for an item under the Heard training condition.* 115

Figure 10 *Proportion of Recall by Training Condition (Verbal/Speech – Heard, Verbal/Non-Speech – Heard, Non-Verbal/Non-Speech – Heard).* 118

Figure 11 *Proportion of Looks to Target by Training Condition (Verbal/Speech – Heard, Verbal/Non-Speech – Heard, Non-Verbal/Non-Speech – Heard).* 121

Figure 12 *Predictions from model estimating number of fixations to target by Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations), and Age (4yrs, 6yrs), showing the three-way interaction.* 152

Figure 13 *Mean pointing Accuracy By Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations) and Age (4yrs, 6yrs).* 155

Figure 14 *Framework for the Production Effect.* 180

Figure 15 *Representation of Production Effect for a child memorizing familiar words* 185

Figure 16 *Representation of attenuation or reversal of Production Effect for a child memorizing familiar words* 187

1 General Introduction

One common strategy for memorizing information is to say it aloud. While this seems intuitive, it is also backed by empirical research showing that reading a word aloud improves recall and recognition compared to reading silently. This phenomenon is known as the Production Effect (MacLeod et al., 2010). While numerous studies have documented the positive impact of speech production (e.g., Forrin & MacLeod, 2016; Icht & Mama, 2015; Krishnan et al., 2017; Mama & Icht, 2016; Ozubko & MacLeod, 2010; Pritchard et al., 2019; Zamuner et al., 2016), other studies have yielded opposite results in which speech production does not lead to a memory advantage, labeled the Reverse Production Effect. From the literature, a picture is emerging that the presence and direction of the Production Effect can be mediated by many factors, such as the age of the participant (children vs. adults, Zamuner et al., 2016; 2018) and the characteristics of the stimuli (Baese-Berk & Samuel, 2016; Kaushanskaya & Yoo, 2011). Although multiple studies have investigated the Production Effect with adults, there are only a handful of studies that have been conducted with children. This dissertation examines the Production Effect and Reverse Production Effect in children, focusing on the effectiveness of production in different scenarios and in different age groups.

The present chapter (Chapter 1) provides a general overview of the Production Effect and Reverse Production Effect and also provides a review of the cognitive development literature relevant for future discussions. It concludes with a presentation of the research questions addressed in the following chapters.

A note on terminology: it is important to note that different studies use different labels to refer to the same training conditions. For instance, *Say, Produced, Speak* and *Verbal/Speech* all refer to actions in which individuals produce words aloud. Similarly, *Listen* and *Heard* are also equivalent, participants only listen to the items. Table 1 provides a summary of the training conditions used in studies with children for ease of understanding.

Table 1. *Summary of training conditions in studies with children on the Production Effect*

Authors	Stimuli	Training conditions	Description
Icht & Mama, 2015	Familiar words + images	Look	Look at image in silence
		Look + Listen	Look at image and listen to auditory label
		Look + Say	Look at image and produce the label aloud
Icht & Mama, 2015	Unfamiliar words + images	Look + Listen	Look at image and listen to auditory label
		Look + Say	Look at image, listen to auditory label and repeat label aloud
Junttila & Ylinen, 2020	English words, pseudowords + animal images	Intentional task with production	Look at image, listen to auditory label and repeat label aloud
		Intentional task without production	Look at image and listen to auditory label
López Assef et al., 2021 (Ch. 2)	Familiar words + images	Look	Look at image in silence
		Listen	Look at image and listen to auditory label
		Say	Look at image and produce the label aloud
López Assef et al., 2023 (Ch. 3)	Familiar words + images	Heard	Look at image and listen to auditory label
		Produced	Look at image and produce the label aloud
López Assef et al. (Ch. 4)	Novel words + nonce animals	Heard	Look at image and listen to auditory label
		Verbal/Speech (Produced)	Look at image, listen to auditory label and repeat label aloud
		Verbal/Non-Speech	Look at image, listen to auditory label and stick out tongue
		Non-Verbal/Non-Speech	Look at image, listen to auditory label and touch nose
López Assef et al. (Chapter 5)	Novel words + nonce images	Heard	Look at image and listen to auditory label
		Produced	Look at pictures, listen to auditory label and repeat label aloud
Pritchard et al., 2019	Familiar written words.	Read-Silent	Read words silently
		Read-Aloud	Read words aloud
Pritchard et al., 2019	Novel written words	Read-Silent	Read words silently
		Read-Aloud	Read words aloud
Richtsmeier et al., 2024	Novel words + nonce aliens	Heard	Look at image and listen to auditory label
		Produced	Look at image, listen to auditory label and repeat label aloud
Zamuner et al., 2018	Novel words + nonce animals	Heard	Look at image and listen to auditory label
		Produced	Look at image, listen to auditory label and repeat label aloud

1.1 The Production Effect

While the "Production Effect" is a fairly recent subject of study, researchers have been investigating the effects of speech production on learning for decades. Hopkins and Edwards (1972) compared the recognition of words read silently versus words read aloud in pure and mixed list conditions, finding an advantage for reading aloud words in a within-subjects, mixed list condition. A similar advantage for produced items was found by Conway and Gathercole (1987) and by Gathercole and Conway (1988). Conway and Gathercole (1987) specifically tested recognition and recall along with three conditions: read silently, mouthing, and producing, and they found gradience in performance: read-aloud was better than read silently, with mouthing being in between the two conditions: read aloud > mouthing > read silently. These previous studies proposed that the advantage for produced items could be caused by distinctiveness, such that produced items are more distinctive than read-only items. This distinctive feature is encoded in memory and used during recall and recognition.

MacLeod et al. (2010) is typically cited as the foundational work on the effect of speech production on memory. They showed a robust memory benefit for words that had been produced (read aloud) compared to words presented under different training conditions, such as silently read and silently mouthed. Across eight experiments with adults, MacLeod et al. included multiple manipulations, such as varying the stimuli (real words and nonwords), varying tasks (vocal production and mouthing), design (within-subjects and between-subjects) and list type (mixed lists and pure lists). Overall, there was a systematic beneficial effect from production; however, the effect was restricted to within-subjects and mixed lists designs.

1.1.1 Explanations for the Production Effect

In memory research, there seem to be two definitions of 'distinctiveness' (Hunt, 2006). On the one hand, we find the Isolation Paradigm in which distinctiveness is attached to the item being encoded. Within a group of many items, some of them have different characteristics from others, making them ‘distinctive’ and resulting in enhanced memory. The second definition for ‘distinctiveness’ stems from the Levels of Processing Framework (Craik, 2002; Craik & Lockhart, 1972). In this definition of distinctiveness, the focus shifts from the item being learned, to the processing involved during encoding. In this approach, added levels of encoding creates a more robust representation of an item, which facilitates memorization and retrieval when compared to items studied under conditions with fewer encoding levels.

One explanation for this difference in encoding is that speech production creates articulatory representations that are encoded within the lexical representation of the item being learned or memorized (Hickok, 2012; 2014; Hickok et al., 2011; Kapnoula & Samuel, 2022, Zamuner et al. 2017). This difference in encoding levels is thought to affect how easily we retrieve items, with a higher number of encodings facilitating retrieval due to a more robust representation of the item and more information available to aid the retrieval process. While different, both approaches to distinctiveness attribute the enhanced memory for produced items in the Production Effect to additional information stemming from production.

Studies comparing actions with differences in the number of encoding levels, such as Forrin and MacLeod (2016) and Mama and Icht (2016), provide support for the distinctiveness approach for the Production Effect, and specifically, the definition for distinctiveness where each encoding level creates a distinct representation. For example, Mama and Icht (2016) compared recall rates across different production actions. Their results indicated an advantage for the *'Listen and Write'* task over *'Listen and Speak*'. While participants were exposed to the auditory input (*'Listen'*) in all conditions, the *'Listen and Write'* task generated more distinct representations due to the number of encoding processes involved in each activity. This is because participants were not only listening and then producing the word in writing (auditory and motor information), but also reading them as they wrote, adding an extra layer of visual encoding. In contrast, the *'Listen and Speak'* condition involved only auditory and articulatory encoding, lacking this visual component. Computational models of the Production Effect have also successfully predicted results using the distinctiveness account (Jamieson et al. 2016; Kelly et al., 2022; Saint-Aubin et al., 2021), though to various degrees of success depending on the model (MINERVA2 in Jamieson et al. 2016; REM in Kelly et al., 2022; Feature Model in Saint-Aubin et al., 2021; see Kelly et al., 2024 for a comparison).

While distinctiveness has been the most widely used framework for the Production Effect, another recent approach is the strength-based account (see review in Baese-Berk, et al., 2024), in which speech production enhances the familiarity and memory of the item being learned by creating stronger memory traces (Bodner & Taikh, 2012; Bodner et al., 2014). Unlike the distinctiveness account, which emphasizes the contrast between produced and non-produced items, the strength-based account suggests that production increases memory strength regardless of other study conditions, such as within/pure-list or between/mixed-list designs (MacLeod & Bodner, 2017). In contrast, predictions from the distinctiveness account restrict the Production Effect to between/mixed-list design as these conditions create the comparison between distinctive and non-distinctive training condition. However, the underlying mechanisms that cause this "stronger" representation for produced items for the strength-based account is not clear (Kelly et al., 2024). While results from studies like MacLeod et al. (2010) and Ozubko et al. (2014) have challenged the predictions of a pure strength-based model, it is possible that the Production Effect involves contributions from both distinctiveness and strength-based mechanisms, and are not necessarily mutually exclusive (Baese-Berk, et al., 2024).

So far, the Production Effect appears to stem from a difference in encoding depending on the type of training actions. Multiple studies have replicated the results from MacLeod et al. (2010). Combined, these studies support the idea of the Production Effect as a robust tool for recall and recognition.

1.2 The Reverse Production Effect

While most of the literature with adults shows a consistent Production Effect, some studies with adults have found an attenuation of the effect or disadvantage for produced items. This arises when various factors are manipulated, such as stimuli (phonologically familiar or unfamiliar stimuli in Baese-Berk & Samuel, 2016; Kaushanskaya & Yoo, 2011; Lopez Assef et al., 2023, familiar vs unfamiliar accent in Cho & Feldman, 2016), testing methodology (e.g., list-discrimination task vs recognition task (Bodner & Taikh, 2012), and when comparing pure and mixed lists (Bodner & Taikh, 2012; Ozubko & MacLeod, 2010). These factors are essential because they demonstrate that the Production Effect is more complex than initially thought.

One factor that influences the Production Effect is the novelty of the stimuli. For example, Kaushanskaya and Yoo (2011) found contrasting results when testing adults learning novel words that contained phonologically familiar or unfamiliar sounds. A Production Effect was found with stimuli with phonologically familiar sounds, whereas a Reverse Production Effect (advantage for items studied silently), was found with phonologically unfamiliar stimuli (also see Lopez Assef et al., 2023) Kaushanskaya and Yoo argued that producing novel words caused learners to pay attention to the novel words' phonological structures. Familiar sounds highlighted the similarities between the participants' native language and the novel words, allowing for knowledge of the native language to support and facilitate word learning. For unfamiliar sounds, the learner not only had to learn a novel word but also had to produce unfamiliar sounds during training.

Baese-Berk and Samuel (2016) found a similar effect in which production disrupted Spanish-speaking adults' learning of a new sound contrast (/sa/-/ʃa/ from Basque). Learning was successful in a listening-only condition, but involving production during training led to a learning disruption. A follow-up experiment found that this learning disruption was not specific to producing the sound contrast while learning but was also found when adults named random letters during training. Their results suggest that engaging the production system overall can cause a disruption in learning, and it is not necessary to produce the critical item for the interruption to appear, although producing the new items did show a greater disruption than the letter production. The authors argued that the disruption was caused by task-related factors (production) and cognitive mechanisms, specifically attention and task switching, which agrees with the previously discussed Kaushanskaya and Yoo (2011) study. This is important because it illustrates that cognitive processing plays a role in the Production Effect for adults.

Another way that the Production Effect is impacted is by the amount of training. Although adults may initially benefit from producing speech, increasing the number of production training trials, and thus, the number of repetitions per item, can diminish the positive impact of speech production on learning novel words (Kapnoula & Samuel, 2022). The timing of speech production also appears to play a role in the Production Effect. Studies have shown that producing non-native sound contrasts (Baese-Berk & Samuel, 2022) and novel words (Kapnoula & Samuel, 2023) immediately after exposure can disrupt learning. However, this disruption can be lowered by introducing a 4-second delay between exposure to the item and speech production during the learning task (Kapnoula & Samuel, 2023).

A recent review by Baese-Berk et al. (2024) suggests that the Production Effect is influenced not only by previously discussed factors but also by the level of spoken language being targeted. For instance, studies focusing on the lexical (words) and sub-lexical (phonemes and syllables) levels tend to show a Reverse Production Effect in cognitively demanding situations, while research at the supra-lexical level (morphology and syntax) typically demonstrates a positive effect from production. The authors propose that speech production may impact "high" and "low" levels of language differently. They call for further research that accounts for variables such as task difficulty, the linguistic characteristics of the target, cognitive factors of the learner, and other aspects like the timing, repetition, and level of engagement during speech production.

The previous studies involving adults show that while the Production Effect is present in many contexts, there are a number of contexts in which the effect is reversed or attenuated. Interestingly, work on the Production effect in children shows even less beneficial findings from speech production than the literature with adults. Thus, I now move on to a brief discussion of Production Effect literature on children. While the next section presents a general picture of the literature with children, a more extensive discussion of studies with children can be found for each article in the following chapters.

1.3 Production Effect and Reverse Production Effect in children

Research on the Production Effect in children, though limited, provides valuable insights into how producing words can influence learning and memory. Overall, a handful of studies explored this phenomenon in different age groups and under varying conditions, revealing an intricate pattern on the potential benefit of speech production on learning.

Icht and Mama (2015) were the first to investigate this phenomenon in 5-year-olds, finding that children recalled familiar and unfamiliar words better when they produced the words compared to just looking at images or listening to the words being named during training. Similarly, Pritchard et al. (2019) found a Production Effect in 7-10-year-olds using a reading task with familiar and novel words. Both studies align with the distinctiveness and encoding levels approach for the Production Effect: speech production involves more levels of information to be processed compared to other actions, and thus, creates a distinctive representation that is more accessible to retrieve from memory.

While the previous studies found a Production Effect in children, Zamuner et al. (2018) found the opposite effect. They used an eye-tracking paradigm to investigate 4.5-to-6-year-olds learning of novel words paired with novel animals. They found better recognition for words that had been only heard during training, indicating a Reverse Production Effect. This reversal was attributed to the cognitive demands of the task, arguing that producing novel words while learning was cognitively demanding, which hindered children's ability to learn the mapping between novel words and referents. Additionally, studies with children have also reported equal learning across training condition: Richtsmeier et al. (2024) conducted a novel word learning study with 3- and 4-year-olds, investigating the effect of semantic depth (visual semantic cues) and training condition (producing aloud vs. listening). While semantic depth improved children's ability to correctly identify the novel word-picture pairings, no significant difference was observed between the two training conditions in terms of learning outcomes. Junttila and Ylinen (2020) found similar results: children aged 5-to-8-years-old showed equal learning of novel words that were said aloud or only listened to during training.

Together, these studies suggest that the Production Effect in children is possibly a more nuanced phenomenon or more fragile of an effect compared to adults. For children, the Production Effect can be influenced by the factors that influence adults' performance; however, children's performance is further influenced by the fact that their cognitive and linguistic systems are still developing. While older children and adults generally benefit from the distinctiveness provided by producing words, younger children may find producing while learning too demanding, leading to a reverse or attenuated Production Effect. This suggests that the effectiveness of speech production as a memory aid depends on the children's cognitive and linguistic abilities.

1.4 Cognitive and Developmental Factors

As previously discussed, the enhanced memory for produced items is largely attributed to distinctiveness (MacLeod et al., 2010; Forrin & MacLeod, 2016; Mama & Icht, 2016). This distinctiveness is argued to arise because produced items undergo multiple encoding processes—such as articulatory and auditory—providing more layers of encoding for later retrieval compared to items that are, for example, only heard or only seen. While this explanation seems straightforward, the process encoding is complex and influenced by various factors. This section will explore this complexity and review some cognitive factors that are proposed to affect the Production Effect.

Both adults and children have limited cognitive resources, which restrict participants' abilities to perform simultaneous tasks. In multitasking situations, it is often necessary to prioritize one task over another to effectively manage cognitive load (Courage et al., 2015). This cognitive strain is even more pronounced in infants and preschoolers, who typically struggle with multitasking (Courage et al., 2015). Because distinctiveness involves processing multiple levels of information simultaneously—such as the articulatory and auditory aspects of produced items—this can place a high cognitive demand on young children, who may find the learning task to be challenging due to their still developing cognitive and language related skills.

Cognitive development is particularly relevant to the understanding of the Production Effect. As children grow older, their cognitive abilities improve, ultimately reaching adult-like levels. For example, Davidson et al. (2006) examined memory, inhibition, and task-switching performance in children aged 4 to 13 years and young adults. Participants were tested on memory and inhibition tasks, including those involving retention demands, maintaining and inhibiting information, and task switching vs. single-task performance. The results showed that younger children could successfully complete tasks when the difficulty was minimal. However, as task demands increased, younger children struggled more, and even older children (around 13 years old) exhibited reduced performance, suggesting that adult-like cognitive performance is achieved later in development. These findings indicate that developmental stage and executive functioning significantly influence how participants manage task difficulty and attention demands, which may help explain differences in the Production Effect across age groups and experimental tasks.

However, the challenge of processing multiple encoding layers is not exclusive to children and can also be seen at different points across the lifespan. Research with adults has shown that the recognition and retrieval of distinctive items can also decline with age. Younger adults have better recollection of distinctive items than older adults, suggesting a reduced ability for distinctiveness to enhance memory in older individuals (Carr et al., 2015). This decline is believed to result from diminished cognitive resources in older age, which may impact working memory efficiency and the processing of distinctiveness (Smith, 2006). Additionally, older adults have been observed to exhibit a smaller Production Effect compared to younger adults (Lin & MacLeod, 2012).

1.5 The Present Dissertation

In this dissertation, I explore the impact of speech production on memory and learning in children, by investigating the Production Effect and the Reverse Production Effect. Despite extensive research showing the Production Effect with adults, its manifestation and underlying mechanisms in children remain largely unexplored. Children's developing cognitive and linguistic abilities impact the potential effect of speech production on memory, making the Reverse Production Effect a more likely outcome.

In order to further investigate the Production Effect in children, the studies in this dissertation cover an age range of children from 2-to-6-years-old, who were monolingual, native speakers of English. Children were trained on different stimuli, such as familiar words (Chapter 2 and 3) and novel words (Chapter 4 and 5). Different methodologies were used to study memory and learning: free recall tasks, Yes/No recognition tasks, and the Visual World Paradigm (VWP). All three methods have been used previously to test learning and memorization in Production Effect studies and are also appropriate for studies with children.

The research questions addressed in this dissertation are the following:

1. Does the Production Effect change across development? (Chapter 2)
2. Does the Production Effect with children persist in time, after a 1-week delay? (Chapter 3)
3. Does the learning disruption found in word learning studies with children stem from linguistic and/or general cognitive factors? (Chapter 4)
4. Will the Production Effect surface in a word learning task with children in a richer and more familiar learning context? (Chapter 5)

1.5.1 First Aim of the Current Thesis (Chapter 2): Change across development

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

Chapter 2 presents an article that investigated the Production Effect from a developmental perspective and addressed the first research question: Does the Production Effect change across development? Results indicated a relationship between age and the direction of the Production Effect. Younger participants showed a Reverse Production Effect, recalling more words that were heard during training, while older children showed the classic Production Effect, showing a memory advantage for produced words.

1.5.2 Second Aim of the Current Thesis (Chapter 3): Immediate vs delayed testing

López Assef, B., McDonald, M., Bernard, A., & Zamuner, T. S. (2024). Effects of speech production training on memory across short and long delays in 5-and 6-year-olds: A pre-registered study. *Applied Psycholinguistics*, 1-24. https://doi.org/10.1017/S0142716424000109

Chapter 3 extends the investigation of the Production Effect in children by looking at whether the effect persists after a 1-week delay. Results indicated that at immediate testing there was a recall advantage for produced words; however, this no longer held after a 1-week-delay in either the recall or recognition task. Thus, unlike with adults, the Production Effect did not persist to delayed testing.

1.5.3 Third Aim of the Current Thesis (Chapter 4): Nature of the Reversed Production Effect

López Assef, B. & Zamuner, T. S. (Accepted). Task effects in children's word recall: Expanding the reverse Production Effect. *Journal of Child Language.*

Chapter 4 addresses the third research question: Does the learning disruption found in word learning studies with children stem from linguistic or general cognitive factors? Results indicate that the learning disruption found in previous studies is not specific to speech production, and that word learning can be disrupted when children perform different types of concurrent actions.

1.5.4 Fourth Aim of the Current Thesis (Chapter 5): Expanding the Learning Task

López Assef, B. & Zamuner, T. S. (In prep). Word learning in a storybook setting: Recognition is better for younger children when they simply listen rather than produce aloud. *Developmental Psychology.*

Chapter 5 addressed the fourth research question; Will the Production Effect surface in a word learning task with children in a richer and more familiar learning context? Results showed a Reverse Production Effect across all manipulations, indicating that even when using a more naturalistic learning task, speech production continues to disrupt young children's mapping novel words to novel referents.

The details regarding each study, along with the review of the related literature are presented in the subsequent chapters.

Chapter 6 provides a summary of the results from the four articles. It also highlights the connections between the articles and proposes a model to frame the results for studies with children. Lastly, the chapter includes discussion of the limitations for the studies and future directions.

References

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

Baese-Berk, M. M., & Samuel, A. G. (2022). Just give it time: Differential effects of disruption and delay on perceptual learning. *Attention, Perception, & Psychophysics*, *84*(3), 960-980. https://doi.org/10.3758/s13414-022-02463-w

Baese-Berk, M.M., Kapnoula, E.C. & Samuel, A.G. (2024) The relationship of speech perception and speech production: It's complicated. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-024-02561-w

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., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review*, *21*, 149-154. https://doi.org/10.3758/s13423-013-0485-1

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. https://doi.org/10.1037/cep0000082

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

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

Courage, Bakhtiar, Fitzpatrick, Kenny, & Brandeau. (2015). Growing up multitasking: The costs and benefits for cognitive development. *Developmental Review*, 35, 5-41. https://doi.org/10.1016/j.dr.2014.12.002

Craik, F. I. M. (2002). Levels of processing: Past, present... and future? *Memory*, *10*(5–6), 305–318. https://doi.org/10.1080/09658210244000135

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

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Fawcett, J. M. (2013). The Production Effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*(1), 1-5. https://doi.org/10.1016/j.actpsy.2012.10.001

Forrin, N. D., & MacLeod, C. M. (2016). Auditory presentation at test does not diminish the Production Effect in recognition. *Canadian Journal of Experimental Psychology*, *70*(2), 116–124. https://doi.org/10.1037/cep0000092

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition, 16*, 110–119. https://doi.org/10.3758/BF03213478

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135–145. https://doi.org/10.1038/nrn3158

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29*(1), 2–20. https://doi.org/10.1080/01690965.2013.834370

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

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

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and Memory* (pp.3–25). New York, NY: Oxford University Press.

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

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the Production Effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology*, *70*(2), 154. https://doi.org/10.1037/cep0000081

Junttila, K., & Ylinen, S. (2020). Intentional training with speech production supports children's learning the meanings of foreign words: A comparison of four learning tasks. *Frontiers in Psychology*, *11*, 1108. https://doi.org/10.3389/fpsyg.2020.01108

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. https://doi.org/10.1037/xlm0001129

Kapnoula, E. C., & Samuel, A. G. (2023). Wait long and prosper! Delaying production alleviates its detrimental effect on word learning. *Language, Cognition and Neuroscience*, *38*(5), 724-744. https://doi.org/10.1080/23273798.2022.2144917

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

Kelly, M. O., Ensor, T. M., Lu, X., MacLeod, C. M., & Risko, E. F. (2022). Reducing retrieval time modulates the Production Effect: Empirical evidence and computational accounts. *Journal of Memory and Language*, *123*, 104299. https://doi.org/10.1016/j.jml.2021.104299

Kelly, M. O., Lu, X., Ensor, T. M., MacLeod, C. M., & Risko, E. F. (2024). Productions need not match study items to confer a production advantage, but it helps. *Experimental Psychology, 71*(1), 2–13. https://doi.org/10.1027/1618-3169/a000600

Krishnan, S., Watkins, K.E. & Bishop, D.V. (2017). The effect of recall, reproduction, and restudy on word learning: a pre-registered study. *BMC Psychology, 5*(28). https://doi.org/10.1186/s40359-017-0198-8

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). https://doi.org/10.16995/glossa.8582

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., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The Production Effect: delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

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), 1-16. https://doi.org/10.1080/09658211.2014.986135

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

Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2019). 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

Richtsmeier, P. T., Gladfelter, A., & Moore, M. W. (2024). Contributions of Speaking, Listening, and Semantic Depth to Word Learning in Typical 3-and 4-Year-Olds. *Language, Speech, and Hearing Services in Schools*, 1-14. https://doi.org/10.1044/2024_LSHSS-23-00105

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the Production Effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language, 118,* 104219. https://doi.org/10.1016/j.jml.2021.104219

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. (2018). 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

2 A Shift in the Direction of the Production Effect in Children Aged 2-6 year[1]

Abstract

Research has found mixed evidence for the Production Effect in childhood. Some studies have found a positive effect of production on word recognition and recall, while others have found the reverse. This paper takes a developmental approach to investigate the Production Effect. Children aged 2–6 years (n = 150) from a predominantly white population in Ottawa, Canada were trained on familiar words which were either seen, heard or produced, followed by a recall task. Results showed a developmental shift: younger participants showed a reverse Production Effect, recalling more words that were heard during training, while older children showed the typical Production Effect, recalling more produced words. The effect of production on recall is not unidirectional and varies by age.

Keywords: Production Effect, production, language development, cognitive development, recall

[1] 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.

2.1 Introduction

When a child learns a word, does it help them encode or remember the word if they say it aloud? Many of us will answer, ‘Yes, of course!’ based on our intuition of the benefit of producing words. This advantage for items said aloud is known as the Production Effect. But the picture is far from simple during early language development, and the effect of production on language learning is not as straightforward as we might assume. The influence of production starts in early infancy: production-related actions have been found to influence infants’ perception of sound contrasts (Bruderer et al., 2015; Yeung & Werker, 2013). With children and adults, speech production is both beneficial and detrimental to learning, and the direction of the effect has been argued to depend on task-, attentional-, linguistic-, and experience-related factors (Zamuner et al., 2017). The impact of production on word recognition and recall may vary across development, with different effects of production for a child who is just learning how to speak, compared to an older child who is already proficient in producing language. Although production skills begin in infancy and continue to develop throughout childhood, no work on the Production Effect has been done with children under 4½-5 years of age, and no studies have compared the potentially-changing impact of production across different ages. To further our understanding of the effect of production across development, the current study investigates the Production Effect with children across a wide age range, i.e., children aged 2-6 years old, evaluating its influence in a continuous manner. Children were trained on images of objects depicting familiar words in different conditions: look (children see the image in silence), listen (children see the image and hear the corresponding word), and say (children see the image and produce the corresponding word). In line with theories and models that predict an interaction between task demands and performance on tasks, such as PRIMIR (Werker & Curtin, 2005), and consistent with prior work by Icht and Mama (2015) and Pritchard et al. (2019), we predicted that older children would recall more words from the say than from the look or listen training conditions. However, we also expected the effect of production could vary as a function of age, e.g., younger children might either show an attenuation or a reversal of the Production Effect, since young children's less mature system may have more difficulty processing words that were produced aloud (Zamuner et al., 2018). By using an approach focused on development as a continuum, we aim to detail for the first time the specifics of the age-dependent shift in the effect of production on memory.

Adults display better recognition and recall for words that are overtly produced compared to words that are read silently (Hopkins & Edwards, 1972; MacLeod et al., 2010). Multiple studies have replicated the facilitative effect of production with adults under various circumstances (see MacLeod & Bodner, 2017, for a brief review). For example, the Production Effect is found when adults mouth, spell, write or type words compared to when they read words silently (Forrin et al., 2012; MacLeod et al., 2010). While the Production Effect is strongest when adults generate the words themselves, it is also reported when an experimenter or another participant generates the word (MacLeod, 2011). The Production Effect is seen under different training conditions such as repeating or recalling novel words (Kan et al., 2014; Krishnan et al., 2017), and it also persists after delayed testing (Icht & Mama, 2019; Kaushanskaya & Yoo, 2011; Ozubko et al., 2012).

Originally, the Production Effect was explained in terms of *distinctiveness* – the act of producing items offers additional information about a given word and makes it more distinct from other words trained under conditions that provided less information. This distinctiveness then aids later remembering (MacLeod et al., 2010; Ozubko et al., 2014). For example, when comparing a SAY word to a LISTEN word, the information that is obtained from the action of saying is both auditory (feedback from saying the word) and articulatory (from producing the word), whereas the action for listening involves just auditory information. This additional level of encoding, i.e., adding both auditory and articulatory information to the memory record, provides more details to the word representation in memory, which then contributes to a retention benefit for the SAY word over the LISTEN word.

Uncovering the precise mechanisms behind distinctiveness, however, is still an ongoing topic of investigation in memory research. For instance, as reviewed by Hunt (2006), distinctiveness can be explained as resulting from item-specific information or from using more than one level of processing for creating a record in memory. In the item-specific approach, a word is 'tagged' with visual, auditory and/or articulatory information, and is thus more distinct from others, regardless of the type of tagged information or the processing level that is solicited. This approach is consistent with the proposed mechanism for the Production Effect in some studies (e.g., Conway & Gathercole, 1987; Gathercole & Conway, 1988; MacLeod et al., 2010). However, it has also been suggested instead that memory encoding can be strengthened based on processing across a variety of modalities, also referred to as levels of processing (for example, Hopkins & Edwards, 1972). It is the processing of that same word on different levels that contributes to stronger encoding, e.g. when reading aloud a word, which involves auditory and articulatory processing, versus when reading silently, which involves only visual processing. This has the effect of creating a more robust record in memory, and further retrieval is facilitated (e.g.**,** Icht & Mama, 2019**;** Jones & Pyc, 2013; Mama & Icht, 2016; Ozubko et al., 2014; Pritchard et al., 2019; Zamuner et al., 2016).

Building on this, Icht and Mama (2015) were the first to investigate the Production Effect in children. In line with the distinctiveness account, they proposed that children's performance on word recognition and recall tasks would be related to the number of unique ways in which words were encoded during training. More distinct encodings (i.e., seen and/or heard and/or produced) would result in greater memory enhancement because there would be more ways in which words could be retrieved from memory. In their first experiment, 5-year-old Hebrew-speaking children were trained on familiar objects with additive encoding levels: LOOK (visual = seeing the image), LISTEN (visual + auditory = seeing the image and hearing the word produced by the experimenter) and SAY (visual + auditory + articulatory = seeing the image, hearing the word by saying it, and pronouncing it). Note that we use the terms LOOK, LISTEN and SAY (labels adapted from Icht & Mama, 2015) to be consistent with the terminology used to present previous research. During training, each participant was presented with thirty images depicted on cue-cards, divided across three boxes that represented each learning action/category (LOOK, LISTEN, SAY). On each trial, the experimenter randomly chose a cue-card from one of the three boxes (i.e., LOOK, LISTEN, SAY). The image from a cue-card from the LOOK box was presented in silence, the one for the LISTEN box was presented and labelled by the experimenter, and one from the SAY box was presented and labelled by the child. After training, children had a three-minute break, followed by a free recall task. Recall rates where highest for words from the SAY condition, followed by words from the LISTEN condition, followed by words from the LOOK condition. In a second experiment Icht and Mama (2015) extended this finding to unfamiliar objects in two conditions: REPEAT (experimenter saying the word and participant repeating it) and LISTEN (experimenter saying the word twice). Again, they found higher recall rates for REPEAT over LISTEN. Note that conditions in which participants produced the word in Experiment 1 were modified in Experiment 2 (from SAY to REPEAT), adding an initial labelling by the experimenter, as children were not familiar with the unfamiliar name corresponding to each unfamiliar object (e.g., *cuff*, *pestle*). Results across Icht and Mama's two experiments showed that the Production Effect extended to children (i.e., children recalled more of the words that they produced rather than the words that they had seen or heard) for both known and unfamiliar words. These findings were argued to support the distinctiveness account: the memory benefit, in children, for words in the SAY and REPEAT conditions (in Experiment 1 and 2, respectively) was caused by increased distinctiveness compared to words studied under less distinctive conditions (i.e., SAY and REPEAT, in Experiment 1 and 2, respectively). The greater number of encoding levels resulted in better memory because children had more information to use during retrieval.

The beneficial effect of production, however, is not always straightforward and may be affected by a variety of factors. For instance, in two experiments with children of the same age as in Icht and Mama (4½-5 years), Zamuner et al. (2018) found a reverse Production Effect. Children were trained on auditory novel words paired with visual nonce animals and were asked either to silently listen to the novel words (LISTEN) or to repeat the novel words (REPEAT). Training was followed by a recognition test using a preferential-looking paradigm. Contrary to Icht and Mama (2015), children showed better recognition for the novel words in the LISTEN condition than in the REPEAT condition during training. This was unexpected under the distinctiveness account because novel words from the LISTEN condition were encoded on fewer levels than those from the REPEAT condition, and distinctiveness proposes that words associated with a higher number of encoding levels will have a memory advantage. Zamuner et al. (2018) posited that the effect of using nonce stimuli might have been the source of the Production Effect reversal. When adults completed the same experiment, they showed a classic Production Effect (Zamuner et al., 2016), with better recognition for the novel words that were said aloud versus heard during training. However, for young participants, the effect of repeating a novel word could take away resources that would otherwise favour recall, thus possibly inverting the Production Effect. Methodological differences could thus help explain the changing direction of the Production Effect across studies; for instance, type of stimuli or variation in the duration of exposure to each image (which differed across studies) could, in itself, affect the quality of the memory record.

Further work with adults has also documented that production does not always benefit recognition and recall (Leach & Samuel, 2007). An attenuation or reversal of the Production Effect has been found with phonologically unfamiliar stimuli (e.g., Baese-Berk & Samuel, 2016). For example, Kaushanskaya and Yoo (2011) found a reversal of the effect for phonologically unfamiliar stimuli, but a classic Production Effect for phonologically familiar items. Changing whether items are produced with familiar vs. unfamiliar accents (Cho & Feldman, 2016) also led to a reversal of the Production Effect for words that had been heard with unfamiliar accents, but not for those heard with familiar accents. Even the type of production (e.g., writing or repeating out loud) has been shown to influence the Production Effect. Mama and Icht (2016) compared recall between two types of production conditions contrasting a writing (LISTEN AND WRITE) to a speaking condition (REPEAT). Their results showed an advantage for words that were produced via writing over speaking, supporting the idea that a variety of factors, such as the type of actions being performed and how they compare, may be at play in determining the effect of production on memory. Finally, differences in testing methodology, such as use of list-discrimination task, also produced a reversal or attenuation of the Production Effect in adults, as opposed to a word recognition task which yielded a classic Production Effect (Bodner & Taikh, 2012).

In summary, production has been found to be either beneficial or detrimental to word recognition or recall, and the direction of the effect depends on various factors, such as the task, attentional resources, developmental stage, linguistic stimuli, and experience-related factors such as the age and the linguistic background of the participants (Zamuner et al., 2017). Given the potential influence of multiple variables on the Production Effect, Zamuner et al. (2018) posited that the reversal of the effect in children stemmed from the cognitive demands of the learning task, which are exacerbated in the early stages of language development. In their interpretation of the results, they suggested that engaging the production system made it more difficult to learn the mapping between the novel word form and its referent. For instance, Zamuner et al. (2018) suggested that in their LISTEN condition, the production system was not engaged, allowing children to direct more attention and processing resources to mapping, which created stronger representations and improved recognition and recall. However, in their REPEAT condition, Zamuner et al. (2018) posited that engaging the articulatory system diverted cognitive resources away from encoding, thus the memory trace was not as robust.

Thus, the cognitive demands hypothesis is not solely based on the maturation of the cognitive system, but also on language development: production experience and lexical development could reduce the cognitive strain from production. The influence of cognitive demands across development is in line with predictions made by the Articulatory Filter Hypothesis (Vihman et al., 2014; Vihman, 2017). According to this hypothesis, sounds in the child's current articulatory repertoire are more salient and words that contain these sounds will be encoded more easily. Thus, strings that contain sounds in the child's repertoire will have an encoding advantage over words with new sounds, thereby exerting fewer cognitive demands (Vihman et al., 2014; Vihman, 2017). Moreover, the effect of the Articulatory Filter has been argued to depend on elements such as attentional resources, task difficulty, and developmental stage of the learner. For example, if learners are recalling newly learned words, advantages can be seen for words that contain sounds in children's existing production repertoires (Vihman et al., 2014; also see Richtsmeier & Moore, 2020; Zamuner & Thiessen, 2018). However, when testing infants in a less demanding task (such as passive listening), infants may show listening preferences for patterns outside of their production repertoires rather than production patterns that they have mastered (e.g., DePaolis et al., 2011). Similarly, children's production experience is related to word segmentation abilities (Hoareau et al., 2019) and their ability to learn words (Kay-Raining Bird & Chapman, 1998; Keren-Portnoy et al., 2010), supporting the influence of production on learning. Another factor from language development that is beneficial is lexical development, as it can further facilitate word learning and interact with the cognitive demands from the learning task: language processing skills and vocabulary growth are inter-related (e.g., Fernald & Marchman, 2012; von Koss Torkildsen et al., 2009).

While the majority of work on the Production Effect in adults has used reading paradigms, two of the three child-based studies on the topic (Icht & Mama, 2015; Zamuner et al., 2018) used images and auditory stimuli because their participants were too young to read (i.e., 4.5-6 years). Using a reading paradigm, Pritchard et al. (2019) addressed whether the Production Effect held in children aged 7-10 years. In their study, children were trained to read words printed in one colour aloud and to read silently words printed in another colour. Children then saw words printed in black (40 trained, 20 untrained) and were asked to verbally indicate whether they remembered the words from training. Across two experiments with real words and novel words, Pritchard et al. (2019) found a recognition advantage for items that were read aloud over those that were read silently. This is consistent with the adult literature using reading paradigms and extends the previous findings from children using images and auditory stimuli (Icht & Mama, 2015) but diverges from Zamuner et al. (2018), who found a reverse Production Effect. Pritchard et al. (2019) propose that the difference in their results could also be explained by the cognitive demands for the tasks used. They suggest that their reading-task and a yes/no recognition task were less cognitively demanding, whereas, in Zamuner et al. (2018), children had to learn new word forms while simultaneously producing them, map novel words to referent nonce animals, then recognize the nonce animals that corresponded to the novel words. Thus Pritchard et al.'s (2019) results are in line with a distinctiveness approach for SAY items, in which production resulted in enhanced learning for school-aged children, while also being consistent with the cognitive demand hypothesis when compared to the results of Zamuner et al. (2018).

In sum, research on the Production Effect with children has shown mixed results. Some studies report a production advantage (Icht & Mama, 2015; Pritchard et al., 2019), whereas others found the reverse effect (Zamuner et al., 2018). These differences stem from various interconnected factors: task (learning and testing methodologies), attentional resources (cognitive development), linguistic stimuli (characteristics of the words participants are learning), and experience (age, language experience). Attribution of cognitive resources can change across development, as children's language and cognitive skills improve with age. Yet, to date, no studies have taken a developmental approach, i.e., investigating the effect in a wide age range and continuously comparing performance at different ages. Here we report on a study that aims to fill this gap. Previous research has studied older children with analyses focusing on restricted age groups (i.e., 4.5-6 years in Zamuner et al., 2018; 5 years in Itch & Mama, 2015; 7-10 years in Pritchard et al., 2019). We were interested in establishing whether the production advantage changes across development. Thus, we investigated children aged 2-6 in order to thoroughly capture the phenomenon over a wide age range, from children who are just learning how to speak to older children who are proficient in comprehending and producing language. Our analyses were confirmatory in nature as there is a clear point-like direction for each of our hypotheses. If children show the same pattern of results across a wide range, this would suggest continuity of the effect of production throughout childhood and would thus support the distinctiveness account across development. In other words, it would indicate a constant memory advantage for produced items caused by distinctiveness for children across a wide age range. However, if the direction of the Production Effect changes (e.g., attenuation or reversal of the effect in younger children and a classic Production Effect in older children), this would show a developmental shift in the Production Effect and offer support to the cognitive demands hypothesis, in which children's allocation of cognitive resources changes with age in part due to maturation of the system, as well as with the correlated growth in lexical development (though we did not take measures of vocabulary size and cannot comment on this directly). It is important to note that these two accounts (distinctiveness and cognitive demands hypothesis) are not mutually exclusive but rather complementary. While additional levels of encoding should be beneficial under the distinctiveness account, they may only be beneficial when the cost of cognitive demands are not too high. For instance, if we observe a shift in the direction of the Production Effect within the same general task, this would provide further support for the cognitive demands hypothesis, which, combined with the distinctiveness approach, gives a more complete picture of the Production Effect. Experiment 1 was conducted with children aged 3-6 years, and Experiment 2 with 2-year-old children.

2.2 Experiment 1

The current experiment investigates the Production Effect in children aged between 3-6 years. We adapted Experiment 1 from Icht and Mama (2015) with familiar objects. They used a mixed-training design, and tested Hebrew-speaking children aged 5 years. We used a blocked-training design so that the task would be more suitable to a wider age range, and tested English-speaking children aged 3-6 years. Children were first trained on familiar words in blocks of three different conditions: LOOK (children only see the image in silence), LISTEN (children see the image and hear a recording of the corresponding word) and SAY (children see the image and produce the corresponding word – there is no prior auditory presentation of the word in this condition).

After a short break, children were tested on a free-recall task, and analyses compared the number of words recalled from each training condition. We predicted that, as in Icht and Mama (2015), older children would recall more SAY words, followed by LISTEN words, followed by LOOK words. According to the distinctiveness account, we expected that the more encoding levels that were involved in each action performed at training, the better the performance at test. In the current case, we would expect a gradient increase in recall performance depending on the number of encoding levels: LOOK: 1 level – visual information from picture; LISTEN: 2 levels – visual information from picture and auditory information; SAY: 3 levels – visual information from picture, articulatory information, and auditory feedback.

However, because the same task may be more cognitively demanding for younger children, we could also expect an attenuation or reversal of the Production Effect for younger children (more LISTEN words recalled compared to SAY words). In this case, producing the word could make the task more difficult for younger children, which would require greater cognitive resources during training and thus affect recall performance. If a reverse Production Effect is observed with our youngest participants (LISTEN > SAY), there are different possibilities as to how children might recall words from the LOOK condition with respect to the LISTEN and SAY conditions. Although Zamuner et al., (2018) observed a reverse Production Effect with LISTEN > REPEAT, their study did not include a LOOK condition from which we could draw predictions for the current study. Words in the LOOK condition may show the lowest recall, as silent looking involves the fewest encoding levels (SAY > LISTEN > LOOK; distinctiveness hypothesis). However, when also considering cognitive demands, we might find that recall for LOOK items could pattern in one of three ways. LOOK words may show the highest recall, as silent looking may be the least cognitively demanding task for younger children (LOOK > LISTEN > SAY; distinctiveness and cognitive demand hypothesis: possibility 1). On the other hand, LOOK words may require more cognitive resources for younger children: When an image is provided without an auditory label, children must independently retrieve the word's phonological representation (Ngon & Peperkamp, 2016). This retrieval needed in the LOOK condition may turn out to be more demanding for young children compared to the LISTEN condition, where the auditory label is provided with the image (LISTEN > LOOK & SAY; distinctiveness and cognitive demand hypothesis: possibility 2). Silent observation and retrieval of a word's phonological representation may also create a less robust memory as compared to retrieving a word and producing it (SAY & LISTEN > LOOK; combining the distinctiveness and cognitive demand hypotheses: possibility 3).

2.2.1 Method

2.2.1.1 Participants

Participants were 120 English-speaking children aged 3-6 years. Age was measured continuously in months, but the breakdown by years was as follows: 3-year-olds *(n* = 30, 16 males, 14 females, *M* = 42 months, *SD* = 3.3, *range* 36-47); 4-year-olds (*n* = 30, 15 males, 15 females, *M* = 53 months, *SD* = 3.7, *range* 48-59); 5-year-olds (*n* = 30, 14 males, 16 females, *M* = 66 months, *SD* = 3.4, *range* 61-71); 6-year-olds (*n* = 30, 18 males, 12 females, *M* = 78 months, *SD* = 3.3, *range* 72-83). Children were tested in a sound-attenuated room at a campus-based lab (*n* = 11) or at a museum-based lab (*n* = 109). Participants were recruited from web-based ads or from the museum floor at the uOttawa Living Lab at the Canada Science and Technology museum in Ottawa, Ontario, Canada, and received a sticker for participating. The demographics of the population in Ottawa is predominantly white, which was reflected in our sample. Participants were required to have a minimum lifetime average of 70% exposure to English (*M* = 91%, *SD* = 9.0*, range* = 70-100), to have learned English from birth and to have no more than two consecutive years of more than 30% exposure to another language as estimated from parental reports. All children were also reported to have normal hearing, normal-to-corrected vision, and no history of language impairment. Twenty-eight additional children were tested but not included in the analyses for the following reasons: equipment error (*n* = 2), failure to perform the task properly, e.g., produced all words in the LOOK condition (*n* = 10), no video for off-line coding (*n* = 3), no recall (*n* = 8), failure to complete the experiment (*n* = 3), and errors on training trials that resulted in the lack of data in a condition (*n* = 2). Our original design included 2-year-old participants. Data collection with 2-year-olds started (*n* = 12) but was stopped because of high attrition rates: only one 2-year-old (aged 35 months) completed the task successfully and recalled words. Thus, Experiment 1 reports on data from children aged 3-6 years, and we return to 2-year-olds in Experiment 2.

2.2.1.2 Stimuli

The test stimuli consisted of 30 monosyllabic English words, which were divided into 3 sets of 10: Set 1– *ball, boat, cat, cheese, duck, hair, horse, key, moon, train;* Set 2 – *bath, bear, book, chair, cow, fish, pants, shoes, spoon, truck;* Set 3 – *bed, bird, car, door, frog, hat, milk, pig, sock, tree.* There were 6 practice words: *cookie*, *flower*, *apple*, *monkey*, *airplane*, *baby*. All words were controlled for familiarity based on lexical norms from parental reports from the MacArthur-Bates CDI (Fenson et al., 1993; Frank et al., 2017). According to these norms, the experimental stimuli are produced on average by 83% (*range* 67-96) of children at 24 months, and 97% (*range* 85-99) of children by 30 months. All stimuli were pre-recorded by a female native speaker of English and normalized for amplitude (70 dB). The visual stimuli consisted of coloured clipart. Images and sound files can be found in the dedicated OSF repository: https://osf.io/8rtxu/?view_only=10c727be2f8f40ca8343d8408421705f. To engage the children, we used a stuffed bear (Mr. Wiggles), who wore a removable sleep mask. Twelve WEDGIT™ building blocks were used for the three-minute break between training and the recall task.

2.2.1.3 Design and procedure

There were three training conditions (LOOK, LISTEN, SAY), each presented in separate blocks. Each block consisted of 2 practice trials and 10 training trials, for a total of 36 trials. The order of the blocks was counterbalanced across participants by using six different lists (each presenting the block conditions in a different order).

The experiment was presented using PowerPoint. At the beginning of the experiment, the experimenter introduced a bear named Mr. Wiggles, who was very tired and was going to sleep while the experimenter and participant played a game. Children were told that they were going to learn words, and when Mr. Wiggles had woken up they would be asked to tell him the words they had learned. Children were assigned to a list that determined the order in which each training condition would occur. At the beginning of each block, an image was displayed on the screen indicating the condition for the block (i.e., image of eyes for the LOOK condition, an image of an ear for the LISTEN condition, and an image of an open mouth for the SAY condition) and the experimenter referred to it while explaining the task for that block. Then the presentation included a display of 'Let's Practice!', followed by the 2 practice trials, then a display of 'Let's go!', followed by the 10 training trials. The practice and training trials had the same structure. On each trial, a blank screen was displayed, and when participants looked at the screen, the experimenter triggered the start of the trial. The image of the condition task appeared at the top of the screen (e.g., eyes for the LOOK condition). The experimenter then pressed the keyboard to trigger the stimulus image, which appeared at the bottom of the screen (e.g., pig). After 3 seconds, the slide automatically transitioned to a blank screen. On LOOK trials, the stimulus image appeared in silence. On LISTEN trials, an auditory token of the word played simultaneously with the appearance of the stimulus image. On SAY trials, the stimulus image appeared in silence and the child produced the word.

After all the training trials were presented, a GIF image of a counting down analog clock appeared on the screen to indicate the start of the break period, during which children were engaged in a block-building task. After three minutes, the timer went off which indicated the start of the recall task.

Children were then told that Mr. Wiggles had woken up and were asked to teach Mr. Wiggles the words they had learned. Once their first attempt at trying to recall the words ended (child stopped recalling words), participants were further prompted and asked to indicate if they did not remember any other words and were encouraged to think and try to remember as many more words as they could. If the child was having difficulties or was too shy, the experimenter could ask questions like “Do you remember any words from the game?” “Do you remember any of the things from the game on the computer screen?”. The experiment ended once participants indicated that they did not remember any more words. Children were individually tested in one session, lasting approximately 10 minutes.

2.2.1.4 Coding

Each session was video recorded and coded off-line for children’s responses during training and recall. Each training trial was coded to ensure that the child gave the correct response given the training condition (e.g., saying the word aloud for SAY condition, remaining silent during LISTEN and LOOK conditions). Words from the recall tasks were coded for their training condition. Variation was allowed for recalled items, e.g., the responses *teddy bear* for *bear*, *kitty* for *cat*, *plane* for *airplane*, etc., and coded as correct (n = 54). Most instances (n = 50) were examples were variations that contained the word in whole or in part (e.g., *horse~horsey, ball~beach ball, bear~teddy bear*). There were four examples with a greater difference between the target word and the recall item. In one instance the participant heard *cat* during the LISTEN condition and recalled *kitty*, and in the second instance the participant produced *cat* and recalled *kitty.* The other two instances belonged to the LOOK condition (participant saw a picture of a *bird* and recalled *chick*). In the LOOK condition, participants were not exposed to an auditory token, thus, the participants liked activated the alternative word during training (silently named the image as *chick*). We excluded words from the recall analysis if they had training errors (e.g., LISTEN item that was produced during training), were practice items, or were extra words (words recalled that were not in training). Words that were recalled more than once were counted as a single recall.

2.2.1.5 Statistical analysis

For analysis, we used logistic generalized additive mixed-effects models (GAMMs; Wood, 2017), a type of regression model that can include random as well as (potentially nonlinear) fixed effects. Significance of factors is established through model comparisons, and differences across levels of (significant) factors can also be computed. Just like logistic linear mixed-effects models (Barr, 2008), logistic GAMMs are models that generate (log) probabilities of an outcome response (here, the probability of recalling a word) depending on discrete (e.g., TRAINING condition) or continuous (e.g., AGE) factors, but they also offer the flexibility of accounting for nonlinear trends without overfitting the data (i.e., by using penalized splines). They are thus robust against overconfidence of estimates and can handle missing data points (e.g., if the data set does not contain points for each and every specific AGE). GAMMs have been used at numerous occasions to analyze and model time-dependent linguistic data such as eye movements (Desmeules-Trudel, 2018; Desmeules-Trudel et al., 2020; Desmeules-Trudel & Zamuner, 2019; Porretta et al.,, 2018; Porretta et al., 2016; van Rij et al., 2016) and electroencephalography data (Tremblay & Newman, 2015). They have also been used in other fields where time-dependent data are commonly analyzed, such as ecology and geography (Simpson, 2018). We implemented GAMMs using *R* with the *mgcv* package (Wood, 2017), and model comparisons (chi-square on maximum-likelihood scores of two models) as well as difference curves assessing within-factor level differences through time (or in the current case, AGE) with *itsadug* (van Rij et al., 2017).

We were primarily interested in the effects of TRAINING condition and AGE (in months), but we also included the effect of BLOCK ORDER presentation to ensure that there were no effects or interactions of block recency or primacy on word recall (Murdock, 1962; Tan & Ward, 2000). For determining which random effects structure to use for analysis, a (maximal) model was fitted with flat random intercepts by PARTICIPANTS and ITEMS. A second model was fitted to the data using random flat intercepts by PARTICIPANTS and random linear slopes by ITEMS through AGES. Finally, a third model was built with flat intercepts by PARTICIPANTS and nonlinear smooths by ITEMS through AGES. Since each participant contributed only one AGE value, no random slope or smooth could be included per PARTICIPANT by AGE, thus flat intercepts were included in all models for participants. Pairwise model comparisons on all three possible combinations (chi-square of maximum-likelihood scores) determined that the best fit included nonlinear random smooths by ITEMS through AGES, balancing fit and model complexity. Consequently, the statistical model presented below includes random intercepts per PARTICIPANTS and random nonlinear smooths by ITEMS through AGES. Once the random effects structure had been established, we proceeded to determine whether BLOCK ORDER was significant (comparison of the maximal model with another that did not contain BLOCK ORDER). Following Porretta et al. (2017), the best model was refitted using restricted maximum-likelihood method for the current presentation. Analysis code and data can be found on OSF project.

2.2.2 Results

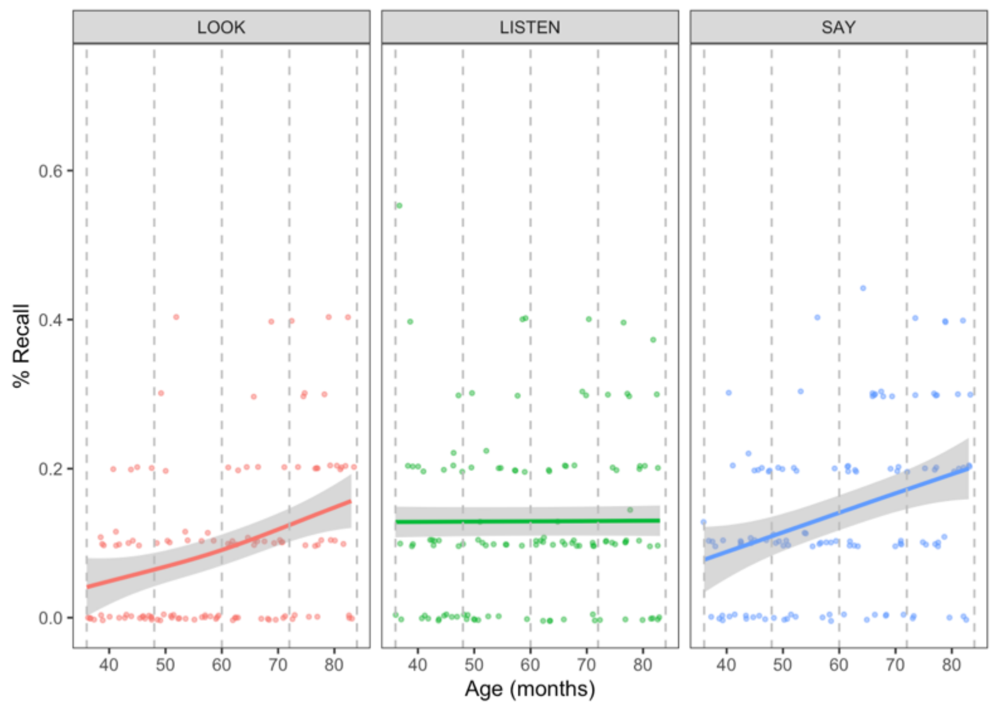
As already mentioned, in Experiment 1, we were interested in the effects of TRAINING condition and AGE (in months) on word recall, in addition to BLOCK ORDER presentation. Importantly, we modeled AGE as a continuous variable, enabling us to determine whether the number of recalled words changed as children aged depending on the TRAINING condition. Rates of recalled words by TRAINING condition through apparent time (i.e., AGE in months) are shown in Figure 1, and descriptive statistics (mean percentages of recalled words, which correspond to the number of recalled words over the number of trained words for each child excluding errors in training trials, and standard deviations by TRAINING condition and AGE in years) are presented in

Table 2. The results displayed in Figure 1 and Table 2 show that for the SAY and LOOK conditions, there was an increase in percentages of recalled words as children aged. However, in the LISTEN condition, percentages of correct recall hovered around the same value (between 11.9 and 14.4% of trained words were recalled) which, considering the fairly large standard deviations, does not suggest an increase or decrease in performance with AGE.

Table 2 *Average percentages of recalled words (number of recalled words over the number of correctly trained words) and standard deviation (in parentheses) by word-training condition for each age group [in years].*

	LOOK	LISTEN	SAY
3 years old	5.07% (6.88)	11.93% (13.28)	9.19% (8.59)
4 years old	6.41% (10.03)	13.16% (11.83)	11.11% (9.59)
5 years old	10.04% (9.83)	12.12% (9.60)	16.48% (11.93)
6 years old	16.00% (12.48)	14.39% (11.74)	19.37% (12.82)

Figure 1 *Effect of* TRAINING *condition and* AGE *(months) on percentages of correct recall for 3- to 6-year-old children in Experiment 1.*

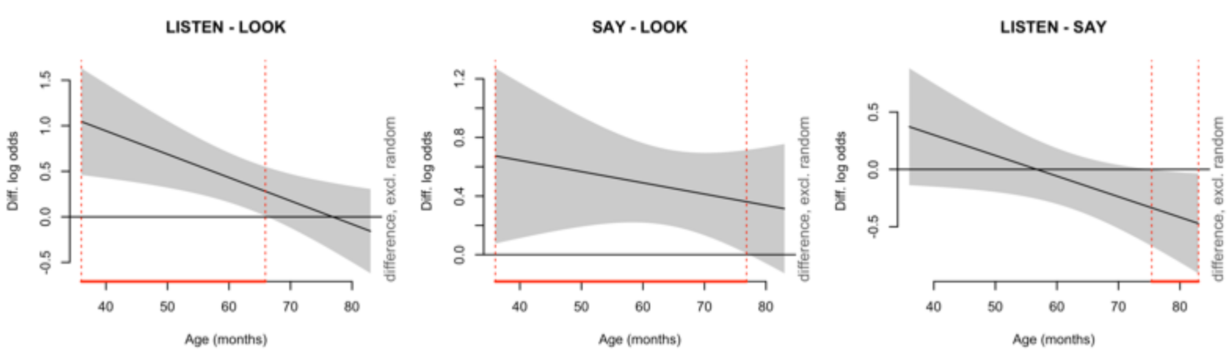


Note. Individual (jittered) dots represent individual children's scores. Coloured lines represent the mean proportion of recall for each TRAINING condition. Grey areas represent the standard error to the mean (0.95).

The GAMM comparison (maximum-likelihood scores) between the maximal model (containing all three factors of interest) and a model containing only TRAINING condition and AGE revealed that there was no evidence in favour of the most complex models (Δ in ML score (Δ *df* = 17) = 9.7, *p* = .31). This shows that BLOCK ORDER was not significant in Experiment 1, and thus this factor was rejected from the model for further analyses.

The numerical output of the final GAMM is presented in Table 3. One way of visualizing this output is by computing difference curves between two levels of a factor through apparent time. Here, Figure 2 shows all three possible difference curves (i.e., between TRAINING conditions through AGES). The leftmost panel shows that between 36 and 65.9 months, children recalled significantly more words that were trained in the LISTEN condition than in the LOOK condition (green curve minus blue curve in Figure 1). In the center panel, model results show that children aged between 36 and 76.8 months recalled significantly more words that were trained in the SAY than LOOK condition (red curve minus blue curve in Figure 1). These two results suggest that hearing or pronouncing familiar words, as opposed to looking, contributes to higher number of recall in children early in development, and that this advantage weakens with age. The rightmost panel shows that children aged 75.4 months or older recalled significantly fewer words that were trained in the LISTEN than in the SAY condition (green curve minus red curve in Figure 1), which is consistent with the 'classic' Production Effect. Combining this observation with the tendency observed in the raw data in Figure 1, the evidence indicates a shift of the Production Effect, where children on average recalled more LISTEN words than SAY words closer to 36 months, and inversely recalled fewer LISTEN words than SAY words later in development after 75 months.

Figure 2 *Difference curves (GAMM output) for probability of recall through AGE by TRAINING condition.*



Note. Significant intervals are displayed in red between dotted lines (i.e., where the 95% confidence interval does not overlap with 0)

Table 3 *Numerical output of the GAMM for recall data in children aged three to six years.*

Formula: recall ~ TRAINING + s(AGE) + s(AGE, by=TRAINING) + s(PARTICIPANT, bs="re") + s(AGE, ITEM, bs="fs", m=1)				
Parametric coefficients	Estimate	Std. Error	Z value	p-value
Intercept(TRAINING:SAY)	-1.905	0.123	-15.502	< 0.001
TRAINING:LISTEN	-0.058	0.124	-0.472	0.637
TRAINING:LOOK	-0.49	0.138	-3.543	< 0.001
Smooth terms	edf	Ref.df	χ^2	p-value
s(AGE)	1	1.001	1.062	0.303
s(AGE, by= TRAINING:LOOK)	1	1.001	6.835	0.009
s(AGE, by= TRAINING:LISTEN)	0.0003	0.0005	0	0.994
s(AGE, by= TRAINING:SAY)	1.001	1.001	4.134	0.042
s(PARTICIPANT) (RANDOM INTERCEPT)	1.911	118	1.937	0.441
s(AGE:ITEM) (RANDOM SMOOTH)	23.06	269	72.608	< 0.001

2.2.3 Discussion

Our results showed a reverse Production Effect for younger children (recall LISTEN > SAY), which shifted to a Production Effect for the oldest children in our sample (recall SAY > LOOK). This shift is in line with the prediction that the direction of the Production Effect is mediated by development-related factors, specifically changes in availability and attribution of cognitive resources during a task. Lastly, we observed that early in development, words from the LOOK condition were recalled less often than those in the LISTEN and SAY conditions, though this disadvantage weakened with age, consistent with the distinctiveness and the cognitive demands hypotheses.

2.3 Experiment 2

Our initial aim in Experiment 1 was to investigate the Production Effect across development, starting with children aged 2-years. However, when 2-year-olds were tested on the methodology of Experiment 1, there were high attrition rates and many children did not recall any words. In Experiment 2, we simplified the design to examine whether the pattern of the reverse Production Effect observed with the younger age group in Experiment 1 would extend to an even younger age group. Specifically, we tested 2-year-old's recall immediately after each of the three training blocks (rather than after a break period following all 3 training blocks).

2.3.1 Method

2.3.1.1 Participants

Participants were 30 English-speaking children aged 2-years (16 males, 14 females, *M* = 30 months, *SD* = 3.2, *range* 24-35). As in Experiment 1, testing occurred in one session. All children were tested in a museum-based lab and recruited from the museum floor. As in Experiment 1, all children were monolingual speakers of English (*M* = 92.8%, *SD* = 8.8 *, range* 70-100), with normal hearing and no diagnosed language or developmental delays, as estimated from parental reports. Thirty-seven additional children were tested but not included in the analyses for the following reasons: failure to perform the task properly (*n* = 8), no video for off-line coding (*n* = 2), no recall (*n* = 7), failure to complete the experiment (*n* = 14), and parental interference (*n* = 6).

2.3.1.2 Stimuli

Same stimuli as Experiment 1.

2.3.1.3 Design and procedure

The procedure replicated that of Experiment 1, with the exception that Mr. Wiggles was woken up after each block to test children's recall. Recall was probed immediately at the end of each of the three training blocks, rather than after a three-minute break, as in Experiment 1. For example, a participant assigned to List 1 would receive: 2 training items, 10 test items, and a recall test in the SAY condition (Block 1), followed by 2 training items, 10 test items, and a recall test in the LISTEN condition (Block 2), followed by 2 training items, 10 test items, and a recall test in the LOOK condition (Block 3).

2.3.1.4 Coding

Coding was the same as Experiment 1.

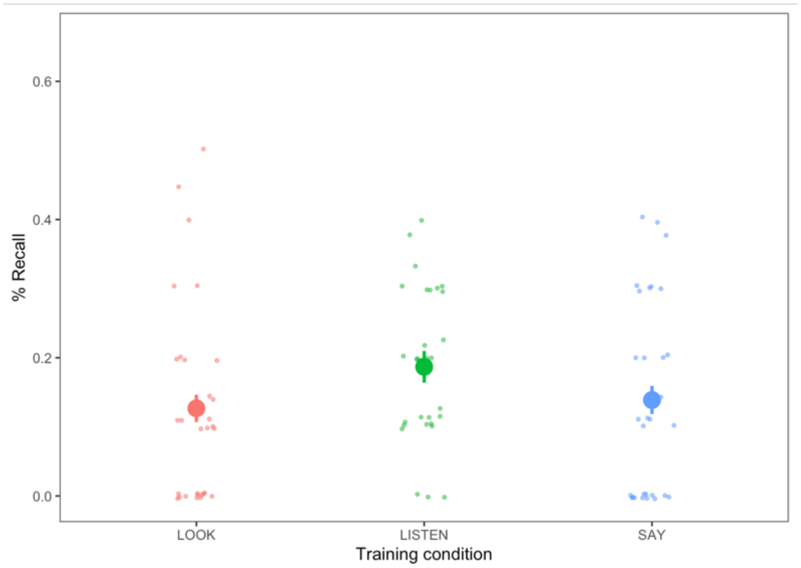
2.3.1.5 Statistical Analysis

For Experiment 2, we were interested in the effects of TRAINING condition, and BLOCK ORDER, without considering age (in months) in this group. Only one random effects structure was possible (i.e., no possibility of inclusion of slopes or smooths through age), which included flat random intercepts by PARTICIPANT and ITEM. We started by fitting a "maximal" model with both fixed factors (i.e., TRAINING condition and BLOCK ORDER). Then, a second model was fitted without the BLOCK ORDER factor and was compared with the maximal model to determine whether BLOCK ORDER was significant. Analysis code and data can be found on OSF project.

2.3.2 Results

Results of Experiment 2 are presented in Figure 3 and Table 4, which display rates of recalled words for 2-year-olds. Model comparisons between the maximal GAMM (containing both TRAINING condition and BLOCK ORDER) and the second model containing only TRAINING condition did not show any evidence in favour of the most complex model (Δ in ML score ($\Delta\ df = 5$) = 0.93, $p = .87$), suggesting that BLOCK ORDER was not significant. Because we were interested in differences across levels of TRAINING condition, this factor was retained. Results of the parametric section of the final model (see Table 5) revealed that neither LISTEN nor LOOK conditions yielded significantly more recall than the SAY condition, although the mean recall for LISTEN condition was numerically slightly greater than those of the SAY and LOOK conditions.

Figure 3 *Effect of* TRAINING *condition on percentages of correct recall for 2-year-old children in Experiment 2*



*Note***.** Small dots represent individual children's proportions of correct recall, and large dots represent the mean percentages of recall for each TRAINING condition.

Table 4 *Average percentages of recalled words (number of recalled words over the number of correctly trained words) and standard deviation (in parentheses) by word-training condition for children aged 2 years.*

	LOOK	LISTEN	SAY
2 years old	12.54% (14.30)	18.74% (11.02)	13.84% (14.06)

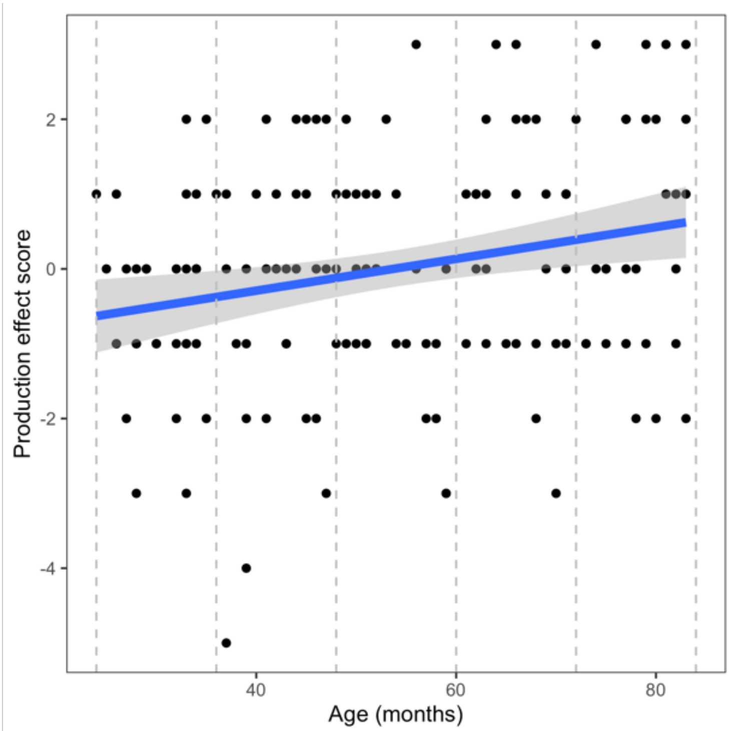
Table 5 *Numerical output of the GAMM for recall data in children aged 2 years.*

Formula: recall ~ TRAINING + s(PARTICIPANT, bs="re") + s(ITEM, bs="re")				
Parametric coefficients	Estimate	Std. Error	Z value	p-value

Intercept(TRAINING:SAY)	-1.988	0.241	-8.247	< 0.001
TRAINING:LISTEN	0.394	0.235	1.676	0.094
TRAINING:LOOK	-0.083	0.253	-0.327	0.744
Smooth terms	edf	Ref.df	χ^2	p-value
s(PARTICIPANT) (RANDOM INTERCEPT)	14.66	29	29.92	0.002
s(ITEM) (RANDOM INTERCEPT)	17.56	29	45.66	< 0.001

To compare the results of the two experiments and to further explore the development of the Production Effect between toddlerhood and childhood, we combined all children tested across Experiments 1 and 2 (i.e., 24-83 months old), and computed a difference score for each child, referred to as a Production Effect Score. This difference score was composed of the number of recalled words that were trained in the SAY condition, minus the number of recalled words that were trained in the LISTEN condition. A negative score indicates a reverse Production Effect (i.e., more recalled words that were LISTEN than SAY), while a positive score indicates a classic Production Effect (i.e., more recalled words that were SAY than LISTEN). This Production Effect Score was submitted to a Pearson's correlation test to assess whether child age (in months) had an effect (Figure 4). We found a positive correlation coefficient ($r = 0.24$), corresponding to a small effect size, which was significantly different from 0 ($t(148) = 2.98$, $p = .003$), indicating that the shift in the direction of the Production Effect was significant within the group of 24- to 83-months-old participants. Indeed, as shown in Figure 4, values of the Production Effect Score tended to be negative in younger children and positive in older children, with an overall crossover in direction occurring slightly before five years old. This is consistent with our hypothesis that the direction of the Production Effect shifts as children advance in age and can be explained by a shift in cognitive resource allocation over the childhood period.

Figure 4 *Production Effect Scores by age (in months), i.e., number of recalled words that were* *SAY minus that were LISTEN, for all children in Experiments 1 and 2, and regression line (linear).*



Note. Grey areas represent the standard error to the mean (0.95).

2.3.3 Discussion

In Experiment 2, children aged 2 years were trained on familiar words under three different conditions: LOOK, LISTEN and SAY. Although no significant differences emerged, the raw data suggests a reversal of the Production Effect, i.e., highest recall rates for LISTEN words. Of the 30 children aged 2 years that were tested, 19 recalled more LISTEN than SAY words, 3 had equal recall amount for LISTEN and SAY words, and 8 recalled fewer LISTEN than SAY words. The lack of a statistically significant effect could be due to reduced power in Experiment 2 or to overly conservative statistical models. Still, the data are numerically consistent with a shift in the Production Effect, displaying as a reverse Production Effect in 2-year-old children. Combining the results from both experiments supports the proposal of a developmental account of the

Production Effect, in which producing words does not benefit young children's recall as it benefits older children's recall.

2.4 General Discussion

The current research investigated the Production Effect across development by testing children whose ages ranged from 2-6 years on a word recall task. We aimed to reconcile the variable results in the handful of existing child-based studies. Our results indicate a relation between age and the direction of the Production Effect. Older children were more likely to show better recall for words that they produced (SAY condition), whereas younger children showed the opposite pattern, with better recall for words that they heard (LISTEN condition). Words associated with pictures seen in silence (LOOK condition) were recalled the least, though the difference between the LOOK compared to the LISTEN and SAY conditions weakened with age. Overall, the direction of the word recall pattern for younger children was LISTEN > SAY > LOOK, whereas for older children it was SAY > LISTEN > LOOK. Below, we first consider the finding for SAY versus LISTEN conditions, then discuss the findings for the LOOK condition.

The development-based shift in the direction of the Production Effect for SAY and LISTEN conditions provides support for the cognitive demands hypothesis (Zamuner et al., 2017, 2018), which in combination with the distinctiveness account (MacLeod et al., 2010), gives a more complete picture of what drives the Production Effect. When cognitive demands are high, whether this is due to task difficulty, cognitive and language development factors, or familiarity with the stimuli, participants' processing cannot be predicted solely by relying on distinctiveness. This is evidenced by the fact that distinctiveness alone cannot account for the results from the younger children for whom the cognitive cost of producing speech hinders the emergence of the Production Effect. Our findings from younger children are also consistent with Zamuner et al.

(2018), who observed a reverse Production Effect with older children trained on auditory novel words paired with visual nonce animals. Children did not have existing lexical representations for the novel stimuli; thus, engaging the production system during learning likely made it more challenging to map the novel word form and referent. As a result, word-learning was detrimentally affected in the REPEAT condition, leading to lower recognition compared to the LISTEN condition, in which there were more resources available to create new lexical representations. The current results are also consistent with previous work on the Articulatory Filter, which has been argued to depend on cognitive factors, such as attentional resources, task difficulty and the developmental stage of the learner (DePaolis et al., 2011; Vihman et al., 2014; Vihman, 2017).

In the current experiments, we found that depending on the age of the child, production can also have an adverse effect on the recall of known words. This supports the idea that higher cognitive demands lead to a reversal of the Production Effect and extends the results to factors other than engagement of the production system during novel word learning. Thus, we suggest that older children were able to successfully use the additional information from SAY items (articulatory experience) to create a distinct representation for produced items because, as a result of their more mature processing systems and/or greater language proficiency, their cognitive resources were not taxed by the act of producing words. Thus, in this case the higher number of encodings led to more distinctiveness, which benefited recall. However, for younger children producing the word caused a higher cognitive demand during training, which resulted in lower recall for produced items. LISTEN items did not involve the cost of production, creating an advantage for LISTEN items over SAY items. Our proposal that the distinctiveness and cognitive demands hypothesis are complementary might appear contradictory; however, we must keep in mind that as language develops, language skills go from receptive to expressive. Words occur first in learners' passive (receptive) lexicons before moving into their active (productive) lexicons; this is termed the comprehension-expression gap (Schneider et al., 2015). Various factors influence the transition of a word from the passive to the active lexicon (e.g., Stokes et al., 2019). For instance, words with a higher frequency or phonological neighbourhood density value have a greater probability of being in the active lexicon and the retrieval of frequent and dense words is faster and more accurate. Thus, we would expect that predictions from distinctiveness and the Production Effect would depend on additional factors relating to lexical representations, such as whether the words are in children's active or passive lexicons and/or words' familiarity and phonological density.

We now turn to evaluating the predictions that were made as to how LOOK items might pattern with respect to words from the LISTEN and SAY conditions. One possibility was that recall would be the greatest for LOOK words, under the hypothesis that silent observation is the least cognitively demanding task. This was not borne out by the data, and instead the pattern of recall was LISTEN > SAY > LOOK in younger children and SAY > LISTEN > LOOK in older children. The lowest recall for LOOK words is consistent with the distinctiveness account (silent looking involves only one encoding process); however, this alone cannot account for the concurrent reversal of performance on recall for LISTEN versus SAY words. The last possibility was that the LOOK condition might be more cognitively demanding because children have to retrieve the word's phonological representation from memory. While this need to retrieve the word's phonological representation was also true for the SAY condition, it may be that silent retrieval in the LOOK condition also created a less robust memory encoding as compared to retrieving a word and producing it aloud because, in the latter, encoding benefited from additional strengthening from the auditory feedback as the child produced the word. Lastly, another possible interpretation is that children may have been less engaged/less attentive on the LOOK trials, leading to lower recall (the lazy reading hypothesis, e.g., MacLeod et al., 2010). Further work is needed to distinguish between these possibilities.

Other factors may have influenced our results, such as children's language or general cognitive abilities. Although our stimuli are familiar on average to 83% of 24-month-olds according to the MacArthur-Bates CDI, linguistic knowledge and experience accumulates with age. Thus, one possible explanation for the reverse Production Effect in younger children is that the stimuli were less familiar to this age group, and younger children's lexical representations were not as well established as were those of the older children. If the words were less familiar to the younger children (2- to 4-years-old group), more attention would have been needed to encode the words. With the addition of an active task like production, this could have led to a disruption in recall. In contrast, for older children, the benefit of production emerged as the words had well-established lexical representations, which subsequently made the task less cognitive demanding, allowing them to benefit from the additional form of encoding. Research with adults has also shown negative effects regarding stimuli familiarity: studies with phonologically unfamiliar stimuli have shown a reversal or attenuation of the Production Effect (Baese-Berk & Samuel, 2016; Kaushanskaya & Yoo, 2011) but a positive effect has been found for phonologically familiar novel words (e.g. Dahlen & Cladwell-Harris, 2013; Kaushanskaya & Yoo, 2011; Krishnan et al., 2017; MacLeod et al., 2010; Zamuner et al., 2016). More work is needed to distinguish between the effect of familiarity and general cognitive development on the direction of the Production Effect. Relatedly, our youngest participant was 24 months old while the oldest was 83 months old: our age group thus ranged from children who were still learning how to articulate to older children who were proficient in producing language. Thus, it is possible that the speech production task added a layer of articulatory complexity for the youngest age group. Additionally, our youngest participants could have also had smaller lexicons than their older counterparts, suggesting that they were less familiar with the sounds and phonological patterns from our words, further adding to the complexity of the production task. To test whether the cognitive demands were related to speech production, we could test the Production Effect for words with varying articulatory complexity. If cognitive demands are influenced by this factor, we would expect the reverse Production Effect to be strongest with words that are more difficult to articulate.

Another factor that may have influenced our results was the methodology. Recall rates in our study were lower compared to those of Icht and Mama (2015). For example, in our study the 5-year-olds recalled an average of 16.5% of the trained words available from the SAY condition, whereas 5-year-olds recalled about 29% of the 10 words from the SAY condition in Icht and Mama (2015). One possibility is that our computer task was less engaging than their cue-card task or that our blocked training condition design was less engaging than their mixed training condition design. The difference in recall rates may also be the result of differences across the studies in the duration of exposure to each image. In our study each trial (slide) had a strict timing that was controlled by the computer, whereas in Icht and Mama (2015) each flashcard was presented for 3 seconds, but presumably there was variability, given that this was presented and controlled by the experimenter. Although we had lower recall rates, our methodology does provide some advantages as the timing of exposure for each auditory and visual stimulus was controlled, and the same auditory tokens were used in the LISTEN condition across all participants (rather than variable across items and participants). Despite these differences in methodologies, 5-year-olds in both studies showed the same pattern: SAY > LISTEN > LOOK.

One important consideration is the equivalence (or lack thereof) of the testing procedures across the production-effect literature. Pritchard et al. (2019) compared the recall of printed words read aloud versus printed words read silently, while Icht and Mama (2015), Zamuner et al. (2018), and the current experiments used visual images in LOOK, LISTEN and SAY (REPEAT) conditions. One could argue that the LOOK condition (silent observation) is equivalent to Pritchard et al.'s (2019) silent reading condition. In our studies, children recalled more words from the SAY condition than from the LOOK condition (with significant differences up until 6 ½ years), using a free recall task, which would then be a replication of Pritchard et al.'s (2019) finding of better performance on a yes/no recognition task for words read aloud relative to words read silently, with school-aged children. Perhaps a more analogous reading-based comparison to the current study would be to test children's recognition and recall for printed words read aloud (SAY) versus read aloud by someone else (LISTEN) versus read silently (LOOK). We predict that for children with emerging or less proficient literacy skills, hearing someone read a word aloud would be a less cognitively demanding task, compared to having to read the words themselves, and that recall performance in a reading task would follow the same pattern as with our youngest participants (LISTEN > SAY > LOOK).

A variety of recall and recognition tasks (e.g., free recall, four alternative forced-choice, old/new recognition, visual recognition) have been used to investigate the Production Effect in children, with an even greater variety seen in the literature on adults. As our findings are in line with the previous proposal that cognitive demands could influence the direction of the Production Effect, it is thus also important to consider the cognitive demands of different testing methodologies. Free recall, our chosen method for testing and Icht and Mama's (2015) in

Experiment 1, could impose a higher cognitive demand for children, as there is no support or information to assist during retrieval. This is different from a recognition experiment, in which participants get to see or hear the word, which could facilitate retrieval. While the reversal of the Production Effect in our study could be attributed to the higher cognitive demand during free recall, a reversal effect was also found using visual recognition in Zamuner et al. (2018), and eyetracking is considered a less demanding task. Importantly, the current work illustrates that even when using the same task, the direction of the Production Effect shifts, indicating that different factors are at play.

To conclude, our findings contribute to reconciling the variable results on the Production Effect by demonstrating that development plays a role in the direction of the effect. When using the same task across a wide age range, younger children showed a reversed Production Effect, whereas older children were more likely to display a production advantage. Although we characterized the shift in the Production Effect as age-dependent, we included no additional measures, such as children's language skills, vocabulary size, executive functioning, and/or working memory capacity. As these skills are inter-related (e.g., Gathercole et al., 1992; Hoff et al., 2008; Swanson, 1996), future work with additional measures is needed to better understand what drives shifts in the direction of the Production Effect. Nevertheless, taken together with previous results in the literature, the current studies remind us that multiple factors can, and almost always do, interact across different levels of development. Moreover, our results contribute to a better understanding of the complex effect of the role of production on language learning.

References

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

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

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 13531–13536. https://doi.org/10.1073/pnas.1508631112

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*, 673-683. https://doi.org/10.1080/20445911.2016.1182173

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

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants?. *Infant Behavior and Development*, *34*, 590-601. https://doi.org/10.1080/15475441.2013.829740

Desmeules-Trudel, F. (2018). *Spoken word recognition in native and second language Canadian French: Phonetic detail and representation of vowel nasalization* (Unpublished doctoral dissertation). University of Ottawa, Ottawa, ON.

Desmeules-Trudel, F., Moore, C. E., & Zamuner, T. S. (2020). Monolingual and bilingual children's processing of coarticulation cues during spoken word recognition. *Journal of Child Language, 47*, 1189-1206. doi: 10.1017/S0305000920000100

Desmeules-Trudel, F., & Zamuner, T. S. (2019). Gradient and categorical patterns of spoken-word recognition and processing of phonetic details. *Attention, Perception, & Psychophysics, 81,*1654-1672. doi:10.3758/s13414-019-01693-9

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur communicative development inventories: User's guide and technical manual*. Singular.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child development*, *83*(1), 203-222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the Production Effect. *Memory & Cognition*, *40*, 1046-1055. https://doi.org/10.3758/s13421-012-0210-8

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*, 677. https://doi.org/10.1017/S0305000916000209

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition, 16*, 110–119. https://doi.org/10.3758/BF03213478

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887–898. https://doi.org/10.1037/0012-1649.28.5.887

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20-to 24-month-olds. *Journal of Child Language*, *35*(4), 903. https://doi.org/10.1017/S0305000908008751

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*, 534-537.

Hoareau, M., Yeung, H. H., & Nazzi, T. (2019). Infants' statistical word segmentation in an artificial language is linked to both parental speech input and reported production abilities. *Developmental Science*, *22*(4), e12803. https://doi.org/10.1111/desc.12803

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and Memory* (pp.3–25). New York, NY: Oxford University Press.

Icht, M., & Mama, Y. (2015). The Production Effect in memory: A prominent mnemonic in children. *Journal of Child Language*, *42*, 1102-1124. DOI: 10.1017/S0305000914000713

Icht, M., & Mama, Y. (2019). The effect of vocal production on vocabulary learning in a second language. *Language Teaching Research*. https://doi.org/10.1177/1362168819883894

Kan, P. F., Sadagopan, N., Janich, L., & Andrade, M. (2014). Effects of speech practice on fast mapping in monolingual and bilingual speakers. *Journal of Speech, Language, and Hearing Research*, *57*, 929-941. DOI: 10.1044/2013_JSLHR-L-13-0045

Kaushanskaya, M. & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*, 121-148. https://doi.org/10.1080/01690965.2010.486579

Kay-Raining Bird, E., & Chapman, R. S. (1998). Partial representations and phonological selectivity in the comprehension of 13-to 16-month-olds. *First Language*, *18*, 105-127. https://doi.org/10.1177/014272379801805204

Keren-Portnoy, T., Vihman, M. M., DePaolis, R. A., Whitaker, C. J., & Williams, N. M. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research, 53,* 1280-1293. https://doi.org/10.1044/1092-4388(2009/09-0003)

Krishnan, S., Watkins, K. E., & Bishop, D. V. (2017). The effect recall, reproduction, and restudy on word learning: A pre-registered study. *BMC Psychology*, *5*, 1-14. DOI: 10.1186/s40359-017-0198-8

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*, 306-353. https://doi.org/10.1016/j.cogpsych.2007.01.001

MacLeod, C.M. (2011). I said, you said: The Production Effect gets personal. *Psychonomic Bulletin & Review*, *18*, 1197-1202. https://doi.org/10.3758/s13423-011-0168-8

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., Gopie, N., Hourihan, K.L., Neary, K. R., & Ozubko, J.D. (2010). The Production Effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 671-685. https://doi.org/10.1037/a0018785

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*, 482–488. https://doi.org/10.1037/h0045106

Ngon, C., & Peperkamp, S. (2016). What infants know about the unsaid: Phonological categorization in the absence of auditory input. *Cognition*, *152*, 53-60. https://doi.org/10.1016/j.cognition.2016.03.014

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The Production Effect endures and improves memory for text. *Memory*, *20*(7), 717–727. https://doi.org/10.1080/09658211.2012.699070

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the Production Effect. *Memory*, *22*, 509-524. https://doi.org/10.1080/09658211.2013.800554

Porretta, V., Kyrölaïnen, van Rij, J., & Järvikivi, J. (2018). Visual World Paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. J. Howlett, & L. C. Jain (Eds.), *Intelligent Decision Technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies – Part II* (pp. 268-277). Springer International Publishing.

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1-21. doi:10.1016/j.wocn.2016.05.006

Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2019). The Production Effect improves memory in 7-to 10-year-old children. *Child Development*, *91*, 901-913. https://doi.org/10.1111/cdev.13247

Richtsmeier, P., & Moore, M. (2020). Order effects in the perception and production of new words. *Journal of Speech, Language, and Hearing Research*, *63*, 3432-3442. doi: 10.1044/2020_JSLHR-20-00077

Schneider, R., Yurovsky, D., & Frank, M. (2015). Large-scale investigations of variability in children's first words. In D. C. Noelle, Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2110-2115): Cognitive Science Society. Retrieved from https://cogsci.mindmodeling.org/2015/papers/0364/

Simpson, G. L. (2018). Modelling palaeoecological time series using generalized additive models. *Frontiers in Ecology and Evolution*, *6*. doi: 10.3389/fevo.2018.00149

Stokes, S. F., de Bree, E., Kerkhoff, A., Momenian, M., & Zamuner, T. (2019). Phonology, semantics, and the comprehension–expression gap in emerging lexicons. *Journal of Speech, Language, and Hearing Research*, *62*, 4509-4522. DOI: 10.1044/2019_JSLHR-19-00177

Swanson, H. L. (1996). Individual and age-related differences in children's working memory. *Memory & Cognition*, *24*, 70-82. DOI: 10.3758/bf03197273

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1589–1625. https://doi.org/10.1037/0278-7393.26.6.1589

Tremblay, A., & Newman, A. J. (2015) Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *52*, 124-139. doi: 10.1111/psyp.12299

van Rij, J., Hollebrandse, B., & Hendriks, P. (2016) Children's eye gaze reveals their use of discourse context in object pronoun resolution. In A. Holler & K. Suckow (eds), *Empirical Perspectives on Anaphora Resolution* (pp.267-293), Berlin: De Gruyter. DOI: 10.1515/9783110464108-103 van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). *itsadug: Interpreting time series and autocorrelated data using GAMMs* [R package]. Retrieved from https://rdrr.io/cran/itsadug/man/itsadug.html

von Koss Torkildsen, J., Hansen, H. F., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, *108*(2), 73-88. https://doi.org/10.1016/j.bandl.2008.09.005

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, *108*, 1–27. https://doi.org/10.1111/bjop.12207

Vihman, M. M, DePaolis, R. A., & Keren-Portnoy, T. (2014). The Role of Production in Infant Word Learning. *Language Learning*, *64*, 121–140. https://doi.org/10.1111/lang.12058

Werker, J.F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*, 197-234. https://doi.org/10.1207/s15473341lld0102_4

Wood, S. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Boca Raton: Chapman and Hall/RC.

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *245*, 603–612. https://doi.org/10.1177/0956797612458802

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. (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., Yeung, H. H., & Ducos, M. (2017). The many facets of speech production and its complex effects on phonological processing. *British Journal of Psychology*, *108*, 37-39. https://doi.org/10.1111/bjop.12220

Zamuner, T. S., Strahm, S., Morin-Lessard, E., & Page, M. P. (2018). Reverse Production Effect: Children recognize novel words better when they are heard rather than produced. *Developmental Science, 21*, 1-13. https://doi.org/10.1111/desc.12636

Zamuner, T. S., & Thiessen, A., (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics, 63*, 609-632. https://doi.org/10.1017/cnj.2018.10

3 Effects Of Speech Production Training On Memory Across Short And Long Delays In 5- And 6- Year-Olds: A Pre-Registered Study[2]

Abstract

Studies on the role of speech production on learning have found a memory benefit from production labelled the "Production Effect". While research with adults has generally shown a robust memory advantage for produced words, children show more mixed results, and the advantage is affected by age, cognitive, and linguistic factors. With adults, the Production Effect is not restricted to the immediate context but also found after a delay. So far, no studies have investigated the effect of delayed recall on the Production Effect with children. Children aged 5-and-6 years old (*n*=60) participated in two sessions. Children were trained on familiar words and images, which were heard (*Listen*) or produced aloud (*Say*). Children then performed a free recall task. One week later, children repeated the recall task and an additional recognition task. At immediate testing, there was a recency effect on words recalled from the different training conditions and a recall advantage for words produced over words heard; however, this no longer held after a 1-week-delay in either the recall or recognition task. Exploratory analysis showed that vocabulary did not predict the Production Effect. Findings indicate that unlike adults, the Production Effect is not as robust in children after a delay.

Keywords: Production Effect, long-term memory, speech production, delayed testing, recall

---

[2] Lopez Assef, B., McDonald, M., Bernard, A., & Zamuner, T. S. (2024). Effects of speech production training on memory across short and long delays in 5-and 6-year-olds: A pre-registered study. *Applied Psycholinguistics*, 1-24.

3.1 Introduction

For years researchers from many areas have been interested in how different factors could potentially help or improve memorization and learning. Actions have been found to be helpful during learning, for example, pairing gestures with new words that were heard during learning improves later recognition (Mayer et al., 2015). Another action that has been identified as potentially beneficial is speech production: saying words aloud. The intuition that saying words aloud will somehow help memorization and learning has been confirmed by a large body of research dedicated to what is labelled as "The Production Effect" (term coined by MacLeod et al., 2010). The Production Effect refers to a memory advantage for produced items compared to items that have been only heard or only seen. Previous studies have consistently found that the Production Effect generalizes across different conditions or populations, such as: different age groups (children, Icht & Mama, 2015; adults, MacLeod et al., 2010), language familiarity (L1, Bodner & Taikh, 2012; L2, Icht & Mama, 2019), testing methodologies (free recall task, Cho & Feldman, 2016; recognition task, Zamuner et al., 2016; translation task, Kaushanskaya & Yoo, 2011), and stimuli characteristics (images, Icht et al., 2020; auditory stimuli, Forrin & MacLeod, 2016).

Since the Production Effect has been presented as a memory tool for children (Icht & Mama, 2015), a key question to ask is whether the Production Effect is restricted to the immediate experimental setting or whether this memory advantage persists over time. While studies with adults have found that the Production Effect persists after a delay in adults (1- and 2- weeks) (Icht & Mama, 2019; Kaushanskaya & Yoo, 2011; Ozubko et al., 2012), to our knowledge, no studies to date have investigated this with children. Delay effects in children, as opposed to adults, are of special interest as the Production Effect in children has been argued to vary depending on many factors, including age and task difficulty (Zamuner et al., 2017). Investigating the effect of a delay on the Production Effect in children not only informs our theories on the Production Effect and the multiple factors that possibly affect it, but it also informs theories about long-term memory and encoding in children, along with potentially informing psycholinguistic theories on the relationship between perception and production. To further our understanding of the Production Effect in long-term memory, the current study extends previous research with children by looking at whether the production advantage is maintained after a 1-week delay. The study's methodology, hypotheses and analysis plan were preregistered, see https://osf.io/3guay/?view_only=eb042fdb21874e35913f37465acbbcd4. The study was divided into 2 sessions: Immediate testing and Delayed testing (1-week later). Children aged 5 and 6 years were trained on images of familiar objects in two conditions: *Listen* (participants see the image and then hear a recording of the corresponding word) and *Say* (participants see the image and say the corresponding word aloud). After training, children were tested on a free recall task. One week later, participants were tested again on the free recall task, and they also completed an additional old/new recognition test. We predicted that if the Production Effect creates distinct encodings for items that are being memorized, the memory trace for *Say* items would be stronger than for *Listen* items. Moreover, we predicted that if this distinctiveness is not susceptible to children's less developed cognitive skills, this would lead to better recall for *Say* over *Listen* words at both immediate and delayed testing, similar to the effect seen with adults.

3.1.1 The Production Effect: Adults

The Production Effect was first proposed by MacLeod et al. (2010). They found a memory advantage for items that were read aloud over items that were read silently in a series of experiments. The foundational work by MacLeod and colleagues built upon previous research that had found an advantage for production when compared to silent reading and mouthing (Conway & Gathercole, 1987; Gathercole & Conway, 1988; Hopkins & Edwards, 1972). MacLeod and colleagues proposed distinctiveness as the reason behind the advantage: produced items are more distinctive than read-only items and this distinctive feature is encoded in memory and used when performing a recall or recognition task.

While many studies have shown a robust memory advantage for produced items (e.g., Hopkins & Edwards, 1972; Icht et al., 2020; Forrin et al., 2012; MacLeod et al., 2010; Ozubko & MacLeod, 2010), some studies have found the opposite: an attenuation of the effect or a disadvantage for produced items. This change in direction of the effect for produced items arises when various factors are manipulated, such as the level of familiarity with the stimuli being used in the study (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Cho & Feldman, 2016; Kaushanskaya & Yoo, 2011, López Assef et al. 2023) and design of the study (list discrimination task, Bodner & Taikh, 2012; between- vs. within-subjects, Jones & Pyc, 2014; mixed vs. pure lists, Ozbuko & MacLeod, 2010). Another factor that has been found to interact with the Production Effect is the order in which the stimuli are presented (Cyr et al., 2021; Gionet et al., 2022; Saint-Aubin et al., 2021). For example, Saint-Aubin et al. (2021) found that items read silently showed better recall when they appeared earlier in the experiment, compared to produced items, which showed better recall when they appeared towards the end.

Furthermore, the Production Effect is also impacted by the amount of training: while participants might benefit from production initially, increasing the number of production training trials (and thus, increasing the number of repetitions per item) can reverse the positive effect from speech production when learning novel words (Kapnoula & Samuel, 2022). Timing of production seems to also affect the Production Effect. Studies have found learning disruptions when producing non-native sound contrasts (Baese-Berk & Samuel, 2022) and new words (Kapnoula & Samuel, 2023) immediately after exposure, but this disruption can be reduced by including a 4s delay between perception and production during the training trials (Baese-Berk & Samuel, 2022; Kapnoula & Samuel, 2023).

3.1.2 Delayed testing: Adults

Previous research has investigated the effect of time on the Production Effect with adults (Grohe & Weber, 2018; Icht & Mama, 2019; Kaushanskaya & Yoo, 2011; Ozubko et al., 2012). Overall, the Production Effect has been found to be present at both immediate and delayed testing. There are lasting effects for the Production Effect in adults for different types of stimuli and different lengths of delays: the Production Effect is found for known words at immediate testing, after a 1-day-delay and 1-week-delay (Ozubko et al., 2012); for words in familiar and unfamiliar accented speech at both immediate testing and a 1-week-delay (Grohe & Weber, 2018); in a second language vocabulary learning task at immediate testing, after 1 week and 2 weeks (Icht & Mama, 2019); and for novel words that were phonologically familiar at immediate testing and after 1 week (Kaushanskaya & Yoo, 2011).

3.1.3 The Production Effect: Children

Currently, only four studies have investigated the Production Effect with children: Icht and Mama (2015), who tested 5-year-old children on familiar and unfamiliar objects (using real words but less frequent and not familiar to children); Zamuner et al. (2018), who tested 4.5 to 6-year-old children using eye-tracking and novel word learning; Pritchard et al. (2019), who tested 7- to 10-year-old children on familiar and novel words using a reading paradigm; and López

Assef et al., (2021), who looked at the Production Effect with familiar objects across development in children 2- to 6-years-old.

Some of these studies observed a Production Effect in children. Pritchard et al. (2019) found an advantage for produced words and novel words when comparing reading aloud to reading silently. Icht and Mama (2015) also found an advantage for produced items using familiar objects when compared to a "*Look*" condition (silently observing the picture) and "*Look and Listen*" condition (looking at the picture and hearing the experimenter say the corresponding word). Icht and Mama described their results in reference to the number of encoding processes for each item: the higher the number of encoding processes involved, the more likely it is to be remembered. More specifically, '*Look'*, silently observing the picture, involved only one encoding process: visual; '*Look and Listen'*, looking at the picture and hearing the experimenter say the corresponding word, involved in two encoding processes: visual and auditory; lastly '*Look and Say*', looking at the picture and saying the corresponding word aloud, was defined as implicating three encoding processes (visual, auditory, and articulatory). Consistent with Icht and Mama's hypotheses, children's recall rates showed a gradient-like pattern, paralleling the number of encoding processes during training: recall was the lowest for '*Look',* followed by '*Look and Listen',* and *'Look and Say'* showed the highest recall. Thus, the memory benefit for produced words was attributed to increased distinctiveness, stemming from the higher number of encoding processes for produced items compared to '*Look*' or '*Look and Listen'* items. Icht and Mama argued that the greater number of encoding processes resulted in better memory for produced items because children had more information to use during activation and retrieval of the items from memory.

However, as with adults, other studies with children have also shown a reversal of the Production Effect. This was found when testing children on their recognition of novel words (Zamuner et al., 2018) and in younger children with familiar words (López Assef et al. 2021). López Assef and colleagues, using the same paradigm as Icht and Mama (2015) and the current paper, found a reversed Production Effect (advantage for heard items over produced items) in their younger children (2- to 3-years old), and a typical Production Effect for older children (5- to 6-years old). While production did not result in a memory advantage for younger children, as children grow older, their cognitive skills develop, thus allowing them to make use of the extra information provided by production.

3.1.4 Developmental differences in memory

In the discussion up until now, we have reviewed the findings where children were tested on recall immediately after training or exposure. In contrast with the adult literature, no studies to date have investigated the Production Effect and delayed testing with children. Thus, we will inform our predictions for delayed testing and the Production Effect with children based on the existing literature looking at the effect of delayed testing on word learning with children.

Previous research has shown that children can remember information across different delays: 5-minutes (Sakhon et al., 2018), 1-hour (Wang et al., 2018), 1-week (Holland et al., 2015; Scarf et al., 2013), 1-month (Markson & Bloom, 1997), and multiple months (Kan, 2014; Wang et al., 2015), with memory usually declining as the delay increases (i.e., Lawson & London, 2015). While many studies suggest that children's memory peaks at immediate testing, and then slowly decreases after time, there are also instances where children show better memory at delayed testing. In a word-learning experiment, 3-to-5-year-olds showed higher than chance looking to a target image after a 1-week delay, but not at immediate testing or after a 5-minute delay (Sakhon et al., 2018). Similarly, when asking children about new information they learned, 5-year-olds showed better memory (higher accuracy) after a 2-to-3-day delay compared to immediate testing, whereas 4-year-olds showed similar performance at both immediate testing and after a 2-to-3-day delay (Bemis & Leichtman, 2019).

The likelihood of remembering something can be increased by engaging with the item either at the moment of encoding or afterwards. For example, the inclusion of memory cues (like mnemonic devices) or support techniques (like multiple exposures to items) can increase 3-year-olds' retention of newly learnt words. Children who studied new words with support techniques showed better retention after initial exposure and were less likely to forget the words over time than children who learned without support techniques (Vlach & Sandhofer, 2012). Distinctive processing has been shown to play an important role from very early stages of development, regardless of cognitive limitations, as it helps guide attention, perception and encoding by making certain items stand out compared to others (Howe, 2006; Howe et al., 2000). Regarding the Production Effect, limitations in children's working memory and cognitive skills could affect the encoding process of the distinctive item, disturbing the Production Effect. For example, while Icht and Mama's study with 5-year-olds found that increased distinctiveness (more processing levels) helped memorization, it is possible that this memory benefit is disrupted in time, thus resulting in the Production Effect being present at immediate test, but not after a delay.

The current project extends investigation of the Production Effect in children by looking at whether the effect persists after a 1-week delay. That is, assuming children can create distinctive encodings for produced words, is this information still available after 1 week? Children aged 5- and 6-years old were trained on familiar words paired with pictures in two conditions: half of the words were said aloud by participants during training (*Say*) while for the other half, participants heard an audio recording of the word (*Listen*). Their memory of these words was tested using a free recall task immediately after training, and after a 1-week delay. Testing after the delay also included a recognition task and a standardized vocabulary test.

Based on previous research with the same-aged children using familiar words (Icht & Mama, 2015; López Assef et al., 2021), we expected the Production Effect to be present at immediate testing. In other words, at immediate testing, there would be a memory advantage for items that were said aloud during training over items that were heard during training. We also had two possible hypotheses for how the delay could affect the Production Effect. The first hypothesis was that the Production Effect creates distinct encodings for items that are being memorized, and these are not affected by children's less developed cognitive skills. Thus, we expected that the memory trace for produced aloud items would be stronger than for heard-only items at both immediate and delayed testing. This follows from previous patterns observed in the literature from the Production Effect with delayed testing with adults (Grohe & Weber, 2018; Icht & Mama, 2019; Kaushanskaya & Yoo, 2011; Ozubko et al., 2012). This also follows from previous memory literature showing that children can retain information in long-term memory after a 1-week-delay (Holland et al., 2015; Scarf et al., 2013). Our second hypothesis was that children are able to create distinct encodings, but these are sensitive to children's less developed cognitive skills and less likely to maintain the distinct encoding at a delay. Thus, we expected a change in performance for *Say* items at delayed testing, resulting in *Say* items showing equal or worse performance at test than *Listen* items at delayed testing. This follows from previous studies using mnemonic tasks with children suggesting that these do not benefit long-term memory (e.g. Krinsky & Krinsky, 1994) and studies on the Production Effect with children

(López Assef et al., 2021; Zamuner et al., 2018) in which developmental stage was identified as a factor mediating the presence of the Production Effect.

Our study also included an exploratory analysis (preregistered) aimed to investigate whether vocabulary score would predict the Production Effect in our participants. Two of the factors proposed to alter the Production Effect are linguistic and experience-related factors (see Zamuner et al., 2017 for a review). Furthermore, studies have found long-term memory and learning benefits in children with larger vocabularies (Daidone & Darcy, 2021; Gathercole et al., 1997; Munro et al., 2012). Vocabulary size in children has been shown to support phonological awareness (Gorman, 2012), phonological development (Edwards, et al., 2004), contribute to new word learning (Gathercole et al., 1997), and predict performance in spoken word recognition (Law et al., 2017; Munson, 2001). Additionally, 2- and 3-year-olds' vocabulary has also been linked with their ability to retrieve words at different delays (Munro et al., 2012). Thus, we explored a potential difference in performance related to children's vocabulary sizes. By exploring the effect of vocabulary size, we hoped to gain insights into the potential relation between children's linguistic abilities and the presence and/or strength of the Production Effect. We predicted that higher standardized vocabulary scores would be associated with a more robust Production Effect, and thus a higher proportion during recall and higher accuracy during recognition for words produced than words heard. The rationale is that larger vocabularies, and thus more robust lexical representations due to higher numbers of connections in the lexicon, could make the task less cognitively demanding for children, thus allowing them to benefit even more from production.

3.2 Method

3.2.1 Participants

Participants were 60 English-speaking children aged 5 years ($n = 30$, 13 males, 17 females) and 6 years ($n = 30$, 20 males, 10 females). This age group was chosen as they have consistently shown the Production Effect with familiar words in previous studies (Icht & Mama, 2015; López Assef et al., 2021). The target sample size of 60 was determined by conducting a power analysis prior to data collection, to obtain 80% power to detect an effect size (d) of 0.4 at the standard .05 alpha error probability. The effect size was derived from prior work and calculated doing ANOVAs on results from participants of similar age from the study by López Assef et al. (2021) on the Production Effect. Our power analysis was done in two ways: using PANGEA (v0.2; Westfall, 2016), and mixed models using the SimR package (v1.0.7, Green & MacLeod, 2016). See pre-registration materials for more details. Data collection stopped once we reached 30 participants with usable data for each age group. Participants were recruited through online advertisements on social media, childrenhelpingscience.com (website for online studies), and a participant database. Participants were given the option to enter their name in a draw for a gift card once they completed the experiment. The experiment was conducted online through Zoom. All our participants resided in Canada at the time of testing. Most of our participants lived in the provinces of Ontario and Alberta, Canada. While demographics varied across participants, most were of middle- to high-socioeconomic background. Following previous studies, such as Zamuner et al. (2018) and López Assef et al. (2021), participants were required to have a minimum lifetime average of 70% exposure to English (M = 89%, SD = 12.8, range = 70-100), to have learned English from birth, and to have not more than two consecutive years of 30+% exposure to another language as estimated from a language background questionnaire completed by parents. Participants were also required to have normal hearing, normal-to-corrected vision, and no history of language impairments or developmental disabilities as determined by parental questionnaire. Lastly, children had to have standardized vocabulary scores on the Expressive One Word Picture Vocabulary Tests (EOWPVT-4, Martin & Brownell, 2011) that were not lower than 85 (1 standard deviations below the mean). Other than the 60 participants included in the analysis, three additional participants were tested but not included in the analysis for having vocabulary scores below 85. Eleven additional participants were tested but excluded for speaking during all heard trials (*n* = 1), not completing the experiment (*n* = 1), and not participating in both test sessions (*n* = 9).

3.2.2 Stimuli

Test stimuli consisted of 20 monosyllabic English words paired with images (*dog, bed, tree, boat, pig, cow, car, train, duck, frog, truck, chair, shoe, fish, cat, door, horse, book, bee, sock*), with an additional 4 disyllabic words used for practice trials (*apple, dolphin, cookie, flower*). Test words were chosen to have a high percentage (96%) of children that produced the words at 30-months-old based on norms available in Wordbank (Frank et al., 2016). 10 words were assigned for each training condition. There were an additional 20 English monosyllabic words (audio only) for the Old/New recognition task. Stimuli were pre-recorded by two female native speakers of English (one speaker recorded items for training, another speaker recorded items for the recognition task) and normalized for amplitude (70 dB). Visual stimuli were colored clipart. Images and sound files can be found in the dedicated OSF repository https://osf.io/gxphj/?view_only=9f08953a1d724882a96549ef1130b696.

3.2.3 Design

There were two training conditions: *Listen* where participants saw the picture and heard an audio recording of the label of the image and *Say* where participants saw a picture on the screen and name the picture aloud. These were presented in a Block design such that a given child was, for example, presented all the *Say* items in Block 1, and all the *Listen* items in Block 2. Participants were randomly assigned to one of four orders which differed in the order of the training conditions (*Listen, Say*) and on the order of the items within each Block. Orders 1 and 2 had the *Say* training condition in Block 1 and *Listen* in Block 2 whereas order 3 and 4 had the reverse. A single item could appear in the *Say* condition in one Order and in the *Listen* condition in another Order. Tables with information for each order can be found on OSF repository. During the recognition task participants heard 20 new items (items which were not part of training) and 20 old items from training. Recognition items were not blocked, but pseudorandomized (see OSF for table with descriptions for each list).

Due to the COVID-19 pandemic, this study was conducted on Zoom. Participants were asked to join the experiment in a room without external distractions. Once participants joined the Zoom call and the experimenter checked that the audio and video were working properly, experimenters used the screen share feature to conduct the experiment and vocabulary test using PowerPoint. The experiment was divided in two sessions: Immediate and Delayed testing (1-week-delay). Sessions were audio and video recorded for off-line coding.

Before training started, participants were told that they would be playing a game. Children saw a clipart image of a boy and were told that they would be helping him put all his toys in a box. Participants were told that there was a challenge in the game: in one part they would hear the child's mom name his toys (*Listen* condition) and for the other part they would have to name the toys that appeared on the screen (*Say* condition). The training portion of the experiment was divided into two Blocks, one for each training condition. There were two practice trials for each training condition, followed by the 10 training trials. If participants made a mistake during practice (for example, producing *Listen* items), these trials would be repeated until the correct response was provided (this was done for 2 participants, both cases for *Listen* practice trials). During the presentation of the items, children saw a picture of a toy box at the bottom of the screen, and then the toy appeared above it, on the center of the screen. Once the child had performed the required action (listening or speaking aloud), the toy was animated to slide into the toy box. Once the training portion was finished (both blocks completed), participants completed a free recall test, where they were asked to name the toys that went inside the box. After each recalled word (regardless of accuracy), a star animation appeared on the screen to keep the child engaged. Script and examples of stimuli can be found on OSF.

Delayed testing was scheduled 1 week after Immediate testing (range = 6-9 days). The session began with a free recall task. Children were told that the boy they helped last week wanted to know if they (the participating child) remembered which toys had gone inside the box the week before. Once the child indicated that they were done and did not remember any more toys, they moved to the Old/New recognition task. This task was added during the second session (compared to López Assef et al. 2021, who only used free recall) to address the possibility that the delay might increase task difficulty, causing children's recall to be at floor level. Previous studies have found different results between free recall tasks and recognition tests (MacLeod & Bodner, 2017; Zamuner et al., 2018). During the Old/New recognition task, children were told that they would hear the boy's sister name some of his toys, and that they had to say "Yes" (or indicate with a 'thumbs up' gesture) if the toy they heard corresponded to one of the toys that had gone into the box the week before, and to say "No" (or do a thumbs down gesture) if it corresponded to a new toy that had not gone inside the box the week before. Participants were rewarded with stars after every five trials. Once the recognition task was over, participants completed the standardized vocabulary test (EOWPVT).

3.3 Results

Data and code have been made publicly available on OSF at the following link: https://osf.io/gxphj/?view_only=9f08953a1d724882a96549ef1130b696

3.3.1 Coding

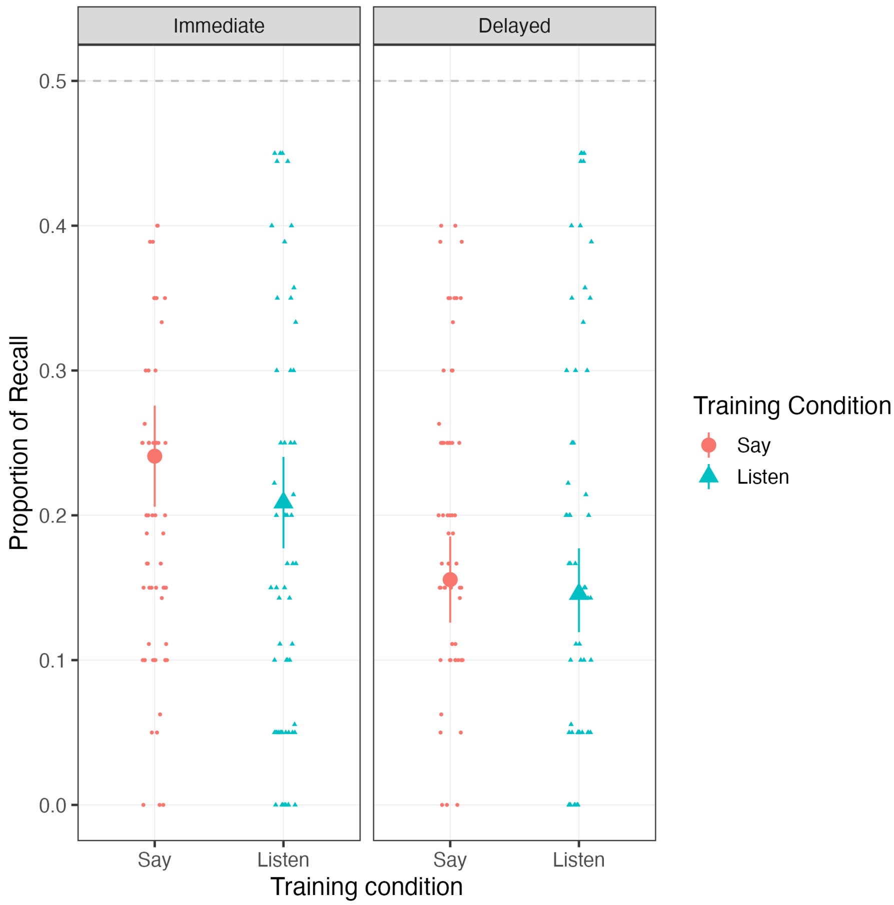
Testing was conducted by the first or second authors, who also did a first pass at coding responses. The main coding was done based on audio-video recordings, by a research assistant who was blind to the initial coding done by the two experimenters. Discrepancies between both codings were resolved by an additional research assistant. This additional research assistant only had access to the recordings of trials with discrepancies between the experimenters and main coder; thus, the additional RA was blind to the rest of the trials. There were 10 recalled words and 13 old/new recognition trials that had discrepancies between the initial two codings. Coding that matched the additional research assistant's coding was marked as correct and resolved. The main coding with resolved discrepancies was the data used for the analyses.

Training trials were coded to ensure that participants completed training properly. Items with training errors were excluded from the analysis (n = 57): these were excluded for producing *Listen* items (n = 13), elaborating during *Say* items (for example, saying "I like socks", instead of just "socks") (n = 33), trials where children were looking away from the screen or loud noises or distractions were happing during the trial (n = 2), technical issues (n = 1), producing the word during training more than once (n = 1), and other errors (n = 7). Training errors represented 5% of our total trials. Exclusions from the recall task included practice items from training (*apple, dolphin, cookie, flower*) and extra recalled words (uttered by a child during recall but not part of training). Words that were recalled more than once were counted as a single recall. Some variation was allowed during training and the free recall task, for example, saying *kitty* for *cat*, although the trials corresponding to these words were excluded from the recognition analysis as the word produced during training did not match the auditory stimuli (*cat*) heard during recognition (*n* = 2).

3.3.2 Free Recall task

We used a generalized linear model framework to examine whether training condition and test session significantly predicted recall rates. This analysis followed our preregistration. Our dependent variable was recall (yes, no). Our fixed effects were Training Condition (*Listen, Say*), Test Session (*Immediate, Delayed*), and Training Order (i.e., order of Training Conditions in Blocks: *Listen in Block 2, Say in Block 2*). Training Order was included as a control variable to account for potential recency effects. Results from the recall task are shown in Figure 5.

Figure 5 *Proportion of Recall by Training Condition (Listen, Say) and Test Session (Immediate, Delayed).*



Note. Points are the condition means by participant with error bars indicating 95% confidence intervals. Dotted line indicates 50% chance level.

All analyses were performed in R (v4.2.0, R Core Team) using the logistic mixed effect model function from the lme4 package (v1.1-27.1, Bates et al., 2015). We started from the most complex model structure with a random structure containing random by-participant intercepts, by-participant slopes for Training Condition and Test Session and their interaction, random by-item intercepts, and by-item slopes for Training Condition, Test Session, Training Order, and their interactions. Models were simplified to account for convergence issues and singularity errors, following our preregistered analysis, until a model converged. Table 6 provides results from the model, along with the final model structure.

Table 6 *Results from Linear Mixed Model Estimating Free Recall by Training Condition (Listen, Say), Test Session (Immediate, Delayed), And Training Order (Listen In Block 2, Say In Block 2).*

Fixed Effects	Estimate	*SE*	*z value*	p-value
Training Condition	0.34	0.18	1.83	0.06
Test Session	-0.49	0.11	-4.25	< 0.001
Training Order	-0.05	0.15	-0.33	0.73
TrainingCondition*TestSession	-0.05	0.22	-0.22	0.82
TrainingCondition*TrainingOrder	-0.86	0.36	-2.38	< 0.05
TestSession*TrainingOrder	-0.06	0.22	-0.27	0.79
TrainingCondition*TestSession*TrainingOrder	1.7	0.45	3.77	< 0.001

Note. The final model had the following syntax specified in the lme4 package: recall_yes_na ~ traincondition_sum*session_sum*trainorder_sum + (traincondition_sum|participant) + (1|target). The proportion of variance accounted for by the final model (pseudo- R^2) was calculated using the r.squaredGLMM function: fixed effects (marginal theoretical $R^2_m = 0.04$); fixed and random effects (conditional theoretical $R^2_c = 0.20$).

There was a significant main effect of Test Session (β = -0.49, *SE* = 0.11, p = <.001), a significant interaction between Training Condition and Training Order (β = -0.86, *SE* = 0.36, p = 0.016), and a 3-way interaction between Training Condition, Test Session, and Training Order (β = 1.7, *SE* = 0.45, p = <.001). No main effect of Training Condition was found (β = -0.34, *SE* = 0.19, p = .07). Post-hoc comparisons were done on the final model with the emmeans package, using asymptotic estimations for degrees of freedom (v1.7.4-1; Lenth, 2020). Estimated marginal means back-transformed into proportions are reported. Post-hoc tests for the significant effect of Test Session revealed a higher probability of recall for words at Immediate Testing (*M* = 0.19*,* *SE* = 0.02, *95% CI* [0.15-0.24]) than at Delayed Testing (*M* = 0.13, *SE* = 0.02, *95% CI* [0.10-0.16]). The interaction between Training Condition and Training Order (Table 6) showed a significant difference between recall across Training Conditions when *Say* was in Block 2 (*Say M* = 0.22, *SE* = 0.03, *95% CI* [0.17-0.27]; *Listen M* = 0.11, *SE* = 0.02, *95% CI* [0.07-0.17]), but not when *Listen* was in Block 2 (*Say M* = 0.15, *SE* = 0.21, *95% CI* [0.11-0.19]; *Listen M* = 0.16, *SE* = 0.03, *95% CI* [0.11-0.23]).

Table 7. *Post-Hoc Tests of the Estimated Marginal Means for Significant 2-Way Interaction from the Model Predicting Proportion of Recall by Training Condition (Listen, Say) and Training Order (Listen in Block 2, Say in Block 2).*

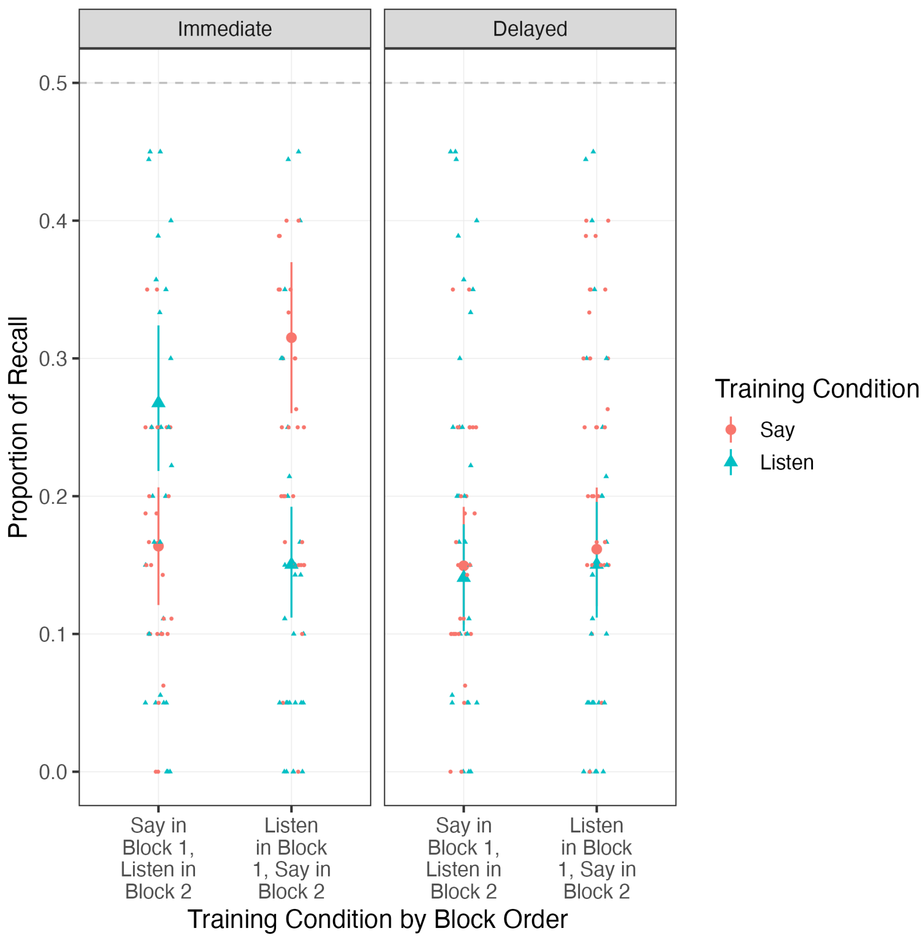
	Odds.ratio	*SE*	*df*	z.Ratio	p-value
Listen in Block 2	0.91	0.24	Inf	-0.38	0.70
Say in Block 2	2.18	0.57	Inf	2.97	0.003

Note. Results are averaged over the levels of Test Session; tests were performed on the log odds ratio scale. Odds ratio represents a comparison of performance on Say to Listen training conditions (Say/Listen).

Lastly, post-hoc tests for the 3-way interaction between Training Condition, Test Session, and Training Order are plotted in Figure 6 and presented in Table 8. There was an overall recency effect at Immediate Testing, in which the Training Condition presented in Block 2 showed higher recall. This recency effect was not present at Delayed Testing. At Immediate Testing in Block 1, there was no significant difference in recall proportion for items presented across Training Conditions. Average recall proportion for the Block 1 *Say* training condition was

0.16 (*SD* = 0.37) and for the Block 1 *Listen* training condition was 0.15 (*SD* = 0.36). On the other hand, for items that appeared in Block 2, *Say* items had an average recall proportion of 0.32 (*SD* = 0.47), while the average for *Listen* items in Block 2 was 0.27 (*SD* = 0.44), which was a significant difference. This suggests that *Say* items in Block 2 received a bigger memory boost than *Listen* items in Block 2, showing a Production Effect on top of the recency effect.

Figure 6 *Mean Recall by Training Condition, Test Session, and Training Order.*



Note. Points are the recall means by participant with error bars indicating 95% confidence intervals. Dotted line indicates 50% chance level.

Table 8 *Post-Hoc Tests of the Estimated Marginal Means for Significant 3-Way Interaction from the Model Predicting Proportion of Recall by Training Condition (Listen, Say), Test Session (Immediate, Delayed), and Training Order (Listen in Block 2, Say in Block 2).*

	odds.ratio	*SE*	*df*	z.Ratio	p-value
Immediate Testing					
Listen in Block 2	0.61	0.18	Inf	-1.69	0.09
Say in Block 2	3.43	1.03	Inf	4.14	<.0001
Delayed Testing					
Listen in Block 2	1.35	0.43	Inf	0.95	0.34
Say in Block 2	1.38	0.43	Inf	1.02	0.31

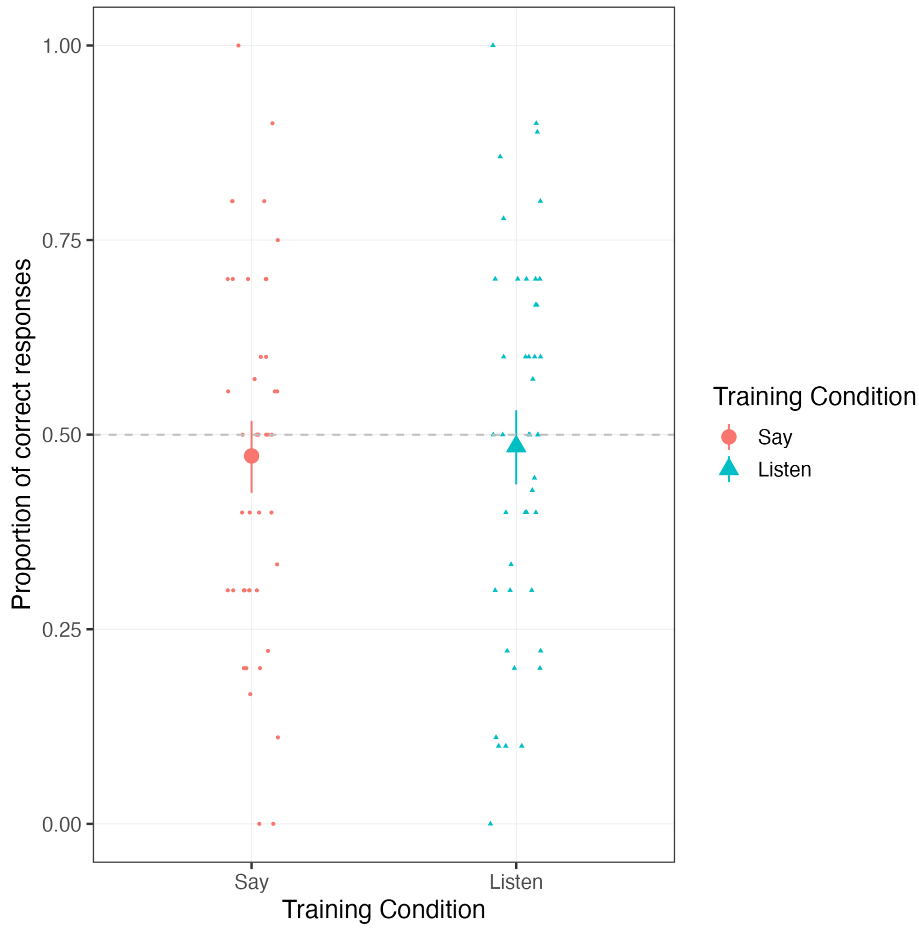
	M	*SE*	*df*	*95% CI*
Immediate Testing				
Listen in Block 2				
Listen	0.23	0.04	Inf	[0.16-0.32]
Say	0.15	0.03	Inf	[0.11-0.21]
Say in Block 2				
Listen	0.11	0.03	Inf	[0.07-0.18]
Say	0.30	0.04	Inf	[0.24-0.39]
Delayed Testing				
Listen in Block 2				
Listen	0.11	0.03	Inf	[0.07-0.17]
Say	0.14	0.03	Inf	[0.10-0.20]
Say in Block 2				
Listen	0.11	0.03	Inf	[0.07-0.18]
Say	0.15	0.03	Inf	[0.10-0.21]

Note. Tests were performed on the log odds ratio scale. Odds ratio represents a comparison of performance on Say to Listen training conditions (Say/Listen).

3.3.3 Old/New Recognition Task

During the Old/New recognition task children heard old test words from the week prior, along with new words in a randomized order. They were instructed to answer "Yes" if the word they heard corresponded to one of the toys from the week before, and to answer "No" if it was a new word. Following our pre-registration, we excluded a subset of participants (*n* = 15) who had an equal or higher proportion of "Yes" answers for new items over old items (Bernard & Onishi, 2023). These children were excluded because it was not clear how well they understood the recognition task given that they were equally or more likely to answer “Yes" (i.e., “I saw this during training last week”) for New items that were not part of training, than for items they actually heard in training. After exclusions, recognition analysis included data from 45 participants. Statistical analysis was restricted to old items only (items from training) since we are interested in the comparison between previously trained *Say* and *Listen* items. The recognition analysis followed the same procedure as for our recall task, except for the exclusion of Test Session in the model, because the Old/New recognition task was only done at Delayed Testing. Accuracy was the dependent variable (correct, incorrect). Results from the Old/New recognition task are shown in Figure 7. Average accuracy for *Say* items was 0.46 (SD = 0.50), and for *Listen* items was 0.48 (SD = 0.50). Due to results for this task hovering around chance, we used d-prime scores (found on OSF repository) to further analyze participants’ performance in this task, and to make sure they had completed the task correctly (i.e., identified *Old* items over *New* items). All of our participants had d-prime scores above 0 (M = 0.98, SD = 0.42), suggesting that their performance is above chance, and thus, their low recognition rates across Training Conditions were not due to task-related issues, since they were able to accurately distinguish between *Old* and *New* items, but because of low memory for items in both Training Conditions. The initial model included Training Condition as a fixed effect, with Training Order as a control variable, along with random by-participant intercepts, by-participant slopes for Training Condition, by-item intercepts, and by-item slopes for Training Condition, Training Order, and the interaction between them. The model was simplified following the same procedure as the previous recall analysis. There were no significant main effects or interactions (Table 8).

Figure 7 *Proportion of Correct Responses During Old/New Recognition Task by Training Condition (Listen, Say).*



Note. Recall was only tested after 1-week delay. Points with error bars indicating 95% confidence intervals are the condition means. Dots are proportion of correct responses by participant. Dotted line indicates 50% chance level.

Table 9 *Results from Linear Mixed Model Estimating Recognition Accuracy by Training Condition (Listen, Say) and Training Order (Listen in Block 2, Say in Block 2).*

Fixed Effects	Estimate	*SE*	*z value*	p-value
Training Condition	-0.07	0.17	-0.41	0.68
Training Order	0.52	0.29	1.80	0.07

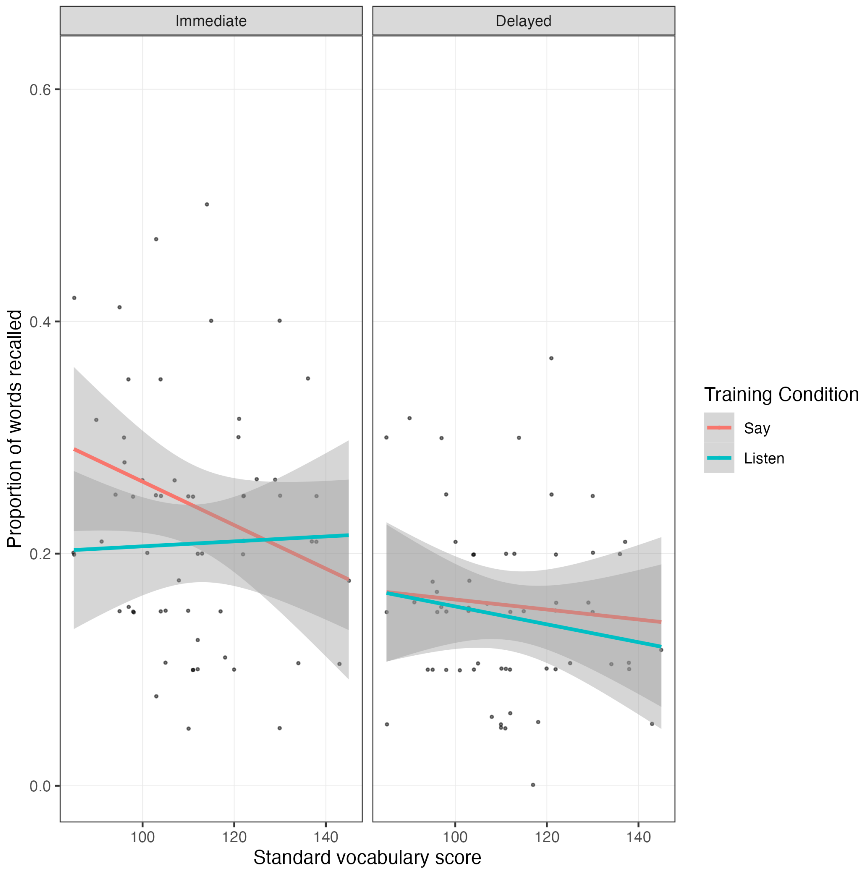
TrainingCondition*TrainingOrder	-0.30	0.31	-0.96	0.34

Note. The final model had the following syntax specified in the lme4 package: recognition~ traincondition_sum*trainorder_sum + (1|participant) + (traincondition_sum|target). The proportion of variance accounted for by the final model (pseudo- R2) was calculated using the r.squaredGLMM function: fixed effects (marginal theoretical R2m = 0.01); fixed and random effects (conditional theoretical R2c = 0.21).

3.3.4 Exploratory analysis: Vocabulary Score

Statistical analyses for incorporating Vocabulary Scores followed the same procedure as the recall analysis. The analysis was conducted on recall data from Immediate Testing only, as this was the only instance where we found a Production Effect. Additionally, visualizations of the vocabulary data (found on OSF) suggested similar performance across Vocabulary Scores at Delayed Testing, but possible differences at Immediate Testing. The variable for standardized Vocabulary Score was centered and used as a fixed effect, with proportion of recall as the dependent variable. We used the standardized Vocabulary Score from the EOWPVT as opposed to the raw scores because the standardized score controls for age. The initial model included Training Condition, Training Order and Vocabulary Score as fixed effects; Training Order as a control variable; random by-participant intercepts, by-participant slopes for Training Condition, by-item intercepts, and by-item slopes for Training Condition, Training Order, Vocabulary size, and the interaction between them. Results for our exploratory analysis are shown in Figure 8. There was a significant interaction between Training Condition and Training Order (β = -1.74, *SE* = 0.40, *p* = <0.001), as seen in the main analyses, but no effects of Vocabulary Score (the corresponding model and results can be found in OSF repository).

Figure 8 *Proportion of Recall by Training Condition (Listen, Say) and Standardized Vocabulary Score.*



Note. Points are the condition means by participant with error bars indicating 95% confidence intervals.

As stated in our preregistration, we also analyzed the effect of vocabulary using a Production Effect Score to see whether vocabulary size affected the strength of the Production Effect. This was calculated from the difference between proportions of recall across both training condition, by participant. A negative score would indicate a child with a Reverse Production Effect (i.e., higher recall for *Listen* than *Say*), whereas a positive score shows the classic Production Effect (i.e., higher recall for *Say* than for *Listen*). Our analysis was done using the lm() function from the stats package (v4.2.0, R Core Team). The Production Effect Score was our dependent variable, with standardized Vocabulary Score and Training Order as our fixed effects (Training Condition is not present here since each participant only has one score). This model did not include item-level effects because each participant only had one score. Results showed only a significant main effect of Training Order (β = -2.59, *SE* = 0.61, *p* = <.001). When *Listen* was in Block 2, participants were more likely to show a negative Production Effect score, whereas when *Say* was in Block 2, they were more likely to show a positive score. These results pattern with those of our recall task. Importantly for our exploratory analyses and the factors that we were examining, vocabulary and the interaction between Vocabulary Score and Training Order were not significant (full results in OSF repository).

Based on a suggestion from a reviewer, we also explored the effect of amount of exposure to English on the Production Effect, and we also included Age in this exploratory analysis to tease apart any confounding effects between amount of exposure to English and age. Exposure to English was calculated as the lifetime percentage of English (based on hours), as measured in our Language Background Questionnaire. Age was calculated in months. This exploratory analysis is available on OSF. The correlation between Exposure to English and Age was significant (*p* = < .001); however, not highly correlated (*r* = -.056). Models that included both Exposure to English and Age did not converge; therefore, we created separate models. We created a simple model with Training Condition and Exposure to English and compared this to another model with Training Condition and Age to explore the effect of each separately. These analyses were based only on the data from the Immediate Testing session. Results showed a significant effect of Exposure to English on recall (β = -0.017, *SE* = 0.006, *p* = < .001), with overall recall decreasing as the percentage of exposure to English increased. There was no significant interaction between Exposure to English and Training Condition. It is not clear why children with less language exposure recalled more words, but interpretation of this effect must be made cautiously, as most participants clustered at the higher end of language exposure (70% to 100%). Only six participants had less than 75% exposure to English, with a high degree of variability on recall rates. When we exclude those six participants, the effect of language exposure on overall recall is no longer significant, but the model shows a significant interaction between Training Condition and Language Exposure (β = -0.73, *SE* = 0.30, *p* = 0.012). Based on these results, if we tested children along the full continuum of language exposure, one would expect a negative correlation between the amount of language exposure and overall recall.

The model for Age had a significant interaction between Training Condition and Age (β = -0.07, SE = 0.018, p = <.0001), with children between 60 to 72 months showing a recall advantage for *Say* items, whereas for older participants (73 to 83 months) this advantage was not present. This pattern is comparable to a previous study with familiar words using free recall by López Assef et al. (2021). In that study, there was an advantage for *Listen* items for 3 and 4-year-olds (36 to 59 months), which shifted to an advantage for *Say* items for 5 and 6-year-olds. The production advantage shift began at 61 months; however, it was only statistically significant for participants aged 75.4 months and older. Thus, the timing of appearance of a significant Production Effect is slightly different across the two studies, showing variability across 5- and 6-year-olds which arguably varies depending on the stimuli and the task. While this is an initial exploration, this gives further indication of possible age-related factors on the Production Effect in children. However, as many variables are often related, more research is needed. These studies would need to continue to carefully control for factors such as stimuli and experiment design; moreover, to adequately include participants across a range of language exposure, age, vocabulary size, and training order (due to the significant recency effect).

3.4 Discussion

In the current study we investigated whether the Production Effect would remain after a 1-week delay in children aged 5 and 6 years old. Results showed a recency effect at Immediate Testing, in which the Training Condition that appeared in Block 2 showed higher recall. When comparing recalled items at Immediate Testing in Block 2, we found a recall advantage for produced words (*Say* condition) over heard words (*Listen* condition), however, no advantage for any Training Condition was found after a 1-week delay. An Old/New recognition task done after a 1-week delay also showed no Training Condition advantage, with similar recognition accuracy in both training conditions.

Thus, our results are in contrast with previous studies with adults which found the Production Effect held at delayed testing (Grohe & Weber, 2018; Icht & Mama, 2019; Kaushanskaya & Yoo, 2022; Ozubko et al., 2012). Our results suggest saying the items aloud can create distinct encodings for produced words at immediate testing, although the success in creating these encodings can also be affected by recency. However, this memory trace appears to decay over time, possibly due to the cognitive limitations stemming from developmental characteristics of children, resulting in difficulty maintaining or retrieving the distinct encoding from memory. While the lack of the Production Effect at Delayed testing was somewhat surprising, previous studies have shown that for children, memory advantages stemming from a mnemonic task could be restricted to immediate testing, and do not seem to aid long-term memory (Krinsky & Krinsky, 1994). Krinsky and Krinsky (1994) investigated the effect of using mnemonic training on children (ages 10 to 12) in two experiments. Children were trained on how to use a mnemonic device, and were tested on recall for familiar, high-frequency nouns. Results showed an increase in recall after being trained on how to use the mnemonic, suggesting that children could benefit from using mnemonics. Interestingly, this benefit was restricted to immediate testing and was not found at delayed testing. At delayed testing, children showed similar recall to what they showed before training, suggesting that mnemonic task benefits could be restricted to immediate testing for children.

To date, studies with children paint a complex picture for the Production Effect. While studies support the theory that production can create distinctive encodings, whether this occurs and if it is possible to maintain in time seem to be affected by multiple factors. In the current study we explored one possible factor that could affect the Production Effect: vocabulary size. Although previous proposals have hypothesized that language-related factors may influence the Production Effect (Zamuner et al., 2017; see Vihman, 2022 for a discussion on the effect of vocal production on word learning in infants and children), the current results fail to support this hypothesis. For our participants, Vocabulary Score was not a predictor of the Production Effect. While we did not find any significant effects for Vocabulary Score, Figure 8, showing proportion of recall by Training Condition across different Vocabulary Scores, suggests that participants with lower vocabulary scores benefit from production more than those on the higher end of the scale, and thus show a Production Effect. Further research could look at different types of measures for vocabulary and other language skill indicators to explore if any other measure may be a clearer indicator. Furthermore, since we excluded participants who fell below 1 standard deviation from the mean Vocabulary Score, it is possible that vocabulary size could predict the Production Effect for children with lower scores.

Another approach to address how language experience might relate to the Production Effect would be to test bilingual and/or second language learners. Comparing monolingual children to other groups would also tap into a different type of experience which is not measured by vocabulary tests. Knowing another language and/or having less experience or proficiency with a language could impact the likelihood of the Production Effect. This may be similar to what has been found in previous studies with adults: the Production Effect has been shown to be impacted by language experience in the form of proficiency (Baese-Berk & Samuel, 2016) and degree of familiarity with a speaker's accent (Grohe & Weber, 2018). Differences in language experience with children could affect the activation of information during the learning process and interact with the Production Effect. For example, Grohe and Weber found that, overall, new words presented in familiar accents were recalled more than new words presented in unfamiliar accents. They argue that for new words presented in familiar accents, adult learners activate accent-related information, and this facilitates processing. This information is not available for words in unfamiliar accents. Applying this to the Production Effect with children, differences in language experience might also lead to differences in how information is activated or processed, possibly causing different effects for production on language learning or memorization. Moreover, L2 learners, may not have accurate vocal productions or maybe not be able to create distinctive representations in their L2 in the same way that monolingual speakers would.

Although it was not part of our original hypotheses, we also found a main effect of recency in our analyses since we included the order of the blocked conditions as a control. We found a significant interaction between Training Condition and Block Order for the recall test after Immediate Testing. When the *Listen* condition was in Block 1 and *Say* in Block 2, more items were recalled from the *Say* condition. Conversely, when the *Say* condition was in Block 1 and *Listen* in Block 2, more items were recalled from the *Listen* condition. However, when looking at recall rates with just Block 1, there were similar recall rates for both training conditions (see Figure 6). In contrast, for just Block 2, recall rates for the training conditions were statistically different from each other: produced words from the second Block showed a higher recall than heard words from the second Block. Thus, *Say* items assigned to the second Block of the study received a bigger memory boost for the free recall task compared to *Listen* items assigned to the second Block. This interaction between order and the production has not been previously reported in the studies with children; however, of the four existing studies with children, most have used a mixed-condition design rather than a block design (mixed-condition design: Icht & Mama, 2015; Pritchard et al., 2019; Experiment 1 in Zamuner et al., 2018; block-condition design: López Assef et al., 2021; Pritchard et al., 2019; Experiment 2 in Zamuner et al., 2018). Focusing on just the studies with a block design which analyzed block order effects, Pritchard et al., (2019) found no main effect of block order using mixed ANOVAs. It is possible that if children create distinctive encodings for produced items in Block 1, these could be more difficult to maintain in memory during the second half of the experiment due to cognitive load effects. Children in this scenario would have to retain distinctive representations for words from Block 1 in memory, while also having to attend to the new information from Block 2. This could lead to forgetting some of the information from Block 1 to prioritize completing the task in Block 2. In this case, children would more easily recall *Listen* items from Block 2 (recency effect), over the *Say* items from Block 1. While unexpected, recent studies have also found a recency advantage in Production Effect designs with adults, where the last few items show better recall (Cyr et al., 2021; Gionet et al., 2022; Saint-Aubin et al., 2021). Most relevant to our study, in adults, produced items showed worse recall when they appeared earlier in the study compared to silently studied items (similar to our results), and showed better recall at later positions. This is relevant for future research looking into the Production Effect, as it is possible that reversal effects or disadvantages for produced items in blocked within-subject designs are due to order effects, and not specific to speech production. However, we did still find within Block 2, a significant Production Effect in the predicted direction. More research on the role of order effects and blocked vs. mixed designs is needed to get a better understanding of their interaction with the Production Effect. Additionally, this opens the question of other types of non-linguistic factors that interact with the Production Effect.

One thing to consider is that our experiment was conducted online, instead of in-person, like previous child and adult studies, which could have caused participants to be less engaged or more distracted during the tasks. We addressed this concern by framing our experiment as a game for children and creating a storyline that would help them become more interested in the task. Our participants showed overall a higher percentage of recall for both training conditions at Immediate testing (24% for *Say* words and 20% for *Listen* words), than the 5- and 6-year-olds in the in-person López Assef et al. (2021) study (18% for *Say* items, 13% for *Listen* items), suggesting that the switch from in-person testing to online testing did not render the task more difficult. Our results are also in line with other studies on the Production Effect with similar aged children: Icht and Mama's (2015) participants (5-year-olds) recalled 29% of *Say* words and 21% of *Listen* items in Experiment 1 and recognized 54.2% of *Say* items and 40.4% of *Listen* items in Experiment 2. Participants in Zamuner et al. (2018), aged 4.5-to-6-years-old, recalled an average of 1.7 out of 4 words (42.5%) for *Say* items, and 0.21 out of 4 words (5.25%) for *Listen* items.

While immediate recall was not affected by the change to online testing, it is possible that because there were more distractions during our experiment, from children not being in a controlled environment, that were impediments to creating long-lasting distinctive encodings. Furthermore, the location in which encoding took place could affect recall. We did not control continuity in the environment where Immediate and Delayed testing took place (e.g., changing from the living room to an office between sessions, although this was not common). Replicating the experiment, and having participants come into a lab twice might provide insights into whether continuity in the training and testing environment affected encoding and/or recall.

Furthermore, while our recall task was below chance (although showing similar rates to previous studies), participants were above chance in our recognition task. It is possible that free recall, retrieving a word from memory with no additional aid, might be too difficult of a task for children, and thus not be fully appropriate to investigate the Production Effect. Other tasks, such as our recognition task, might be easier for children and thus allow us to tap into the Production Effect without large task-related confounding effects.

To conclude, our findings contrast with adult studies which showed the Production Effect after a delay (Grohe & Weber, 2018; Icht & Mama, 2019; Kaushanskaya & Yoo, 2022; Ozubko et al., 2012), and add to the growing literature on the Production Effect in younger participants. So far studies with children have shown that the Production Effect arises with familiar words (Icht & Mama, 2015; older participants in López Assef et al., 2021), and written stimuli with elementary school age children (Pritchard et al., 2019); but a Reversed Production Effect is found with pre-school children and novel words (Zamuner et al., 2018) and for known words for 2-and 3-year-olds (López Assef et al., 2021). The general pattern seen in child studies is that when cognitive demands are high, whether this is due to developmental, language, cognitive issues (or a combination of them), the memory for produced items cannot be predicted solely by relying on distinctiveness (otherwise child and adult studies would consistently show a Production Effect). Our results support the hypothesis that children can create distinctive encodings for produced items, while also adding that it is possible that these encodings are not robust enough to last over long delays. This could be caused by either issues during encoding, or at later stages (e.g., when transferring information from short-term memory to long-term memory). Additionally, our design only included one exposure to each training item, and it is possible that increasing the number of repetitions could provide children with more information or opportunities to create a robust memory trace. Future research looking at cognitive skills and additional measures is needed to better understand the Production Effect in both children and adults. Work on the Production Effect not only leads to a better understanding of the role of speech production on language learning, but also provides further insight into the type of information that is encoded in lexical representations or in memory, and that can later be used to retrieve items. This is especially relevant for psycholinguistic models of lexical representations, since if language experience or production can be activated and influence retrieval, then it is possible they are part of a word's representation. Additionally, the Production Effect has practical uses in real-life situations, as it can be used for pedagogical purposes as a mnemonic device.

References

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

Baese-Berk, M. M., & Samuel, A. G. (2022). Just give it time: Differential effects of disruption and delay on perceptual learning. *Attention, Perception, & Psychophysics, 84*(3), 960-980. https://doi.org/10.3758/s13414-022-02463-w

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. http://dx.doi.org/10.18637/jss.v067.i01

Bemis, & Leichtman, M. D. (2019). That was last time! The effect of a delay on children's episodic memories of learning new facts. *Infant and Child Development, 28*(1), e2113–n/a. https://doi.org/10.1002/icd.2113

Bernard, A., & Onishi, K.H. (2023). Novel phonotactic learning by children and infants: Generalizing syllable-position but not co-occurrence regularities. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2022.105493

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

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

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

Cyr, V., Poirier, M., Yearsley, J. M., Guitard, D., Harrigan, I., & Saint-Aubin, J. (2021). The Production Effect over the long term: Modeling distinctiveness using serial positions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(12), 1797-1820. https://psycnet.apa.org/doi/10.1037/xlm0001093

Daidone, D., & Darcy, I. (2021). Vocabulary size is a key factor in predicting second-language lexical encoding accuracy. *Frontiers in Psychology*, *12*, 2769. https://doi.org/10.3389/fpsyg.2021.688356

Edwards, J; Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, *47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language 44*(3), 677-694. doi: 10.1017/S0305000916000209.

Forrin, N. D., & MacLeod, C. M. (2016). Auditory presentation at test does not diminish the Production Effect in recognition. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, 70*(2), 116-124. doi: https://doi.org/10.1037/cep0000092

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

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition, 16*, 110–119. https://doi.org/10.3758/BF03213478

Gathercole, S.E, Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological Short-Term Memory and New Word Learning in Children. *Developmental Psychology, 33*(6), 966–979. https://doi.org/10.1037/0012-1649.33.6.966

Gionet, S., Guitard, D., & Saint-Aubin, J. (2022). The Production Effect interacts with serial positions: Further evidence from a between-subjects manipulation. *Experimental Psychology , 69*(1), 12–22. https://psycnet.apa.org/doi/10.1027/1618-3169/a000540

Gorman, B. K. (2012). Relationships between vocabulary size, working memory, and phonological awareness in Spanish-speaking English language learners. *American Journal Of Speech-Language Pathology, 21*(2), 109-123.https://doi.org/10.1044/1058-0360(2011/10-0063)

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Grohe, A. K., & Weber, A. (2018). Memory advantage for produced words and familiar native accents. *Journal of Cognitive Psychology, 30*(5-6), 570-587. https://doi.org/10.1080/20445911.2018.1499659

Holland, A., Simpson, A., and Riggs, K. J. (2015). Young children retain fast mapped object labels better than shape, color, and texture words. *Journal of Experimental Child Psychology, 134*, 1–11. doi: 10.1016/j.jecp.2015.01.014

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

Howe, M. L. (2006). Distinctiveness effects in children's memory. In Hunt, R. R., Worthen, J. B. (Eds.), *Distinctiveness and Memory* (pp. 237–257). New York, NY: Oxford University Press.

Howe, M. L., Courage, M. L., Vernescu, R., & Hunt, M. (2000). Distinctiveness effects in children's long-term retention. *Developmental Psychology*, *36*(6), 778-792. https://psycnet.apa.org/doi/10.1037/0012-1649.36.6.778

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. (2019). The effect of vocal production on vocabulary learning in a second language. *Language Teaching Research, 26*(1), 79-98. 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*(6), 715-739. https://doi.org/10.1177/0145445520906583

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

Kan, P. F. (2014). Novel word retention in sequential bilingual children. *Journal of Child Language*, *41*(2), 416-438. https://doi.org/10.1017/S0305000912000761

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), 394–415. https://doi.org/10.1037/xlm0001129

Kapnoula, E.C., & Samuel, A.G. (2023). Wait long and prosper! Delaying production alleviates its detrimental effect on word learning. *Language, Cognition, and Neuroscience, 38*, 724-744. https://doi.org/10.1080/23273798.2022.2144917

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

Krinsky, R., & Krinsky, S. G. (1994). The peg-word mnemonic facilitates immediate but not long-term memory in fifth-grade children. *Contemporary Educational Psychology*, *19*(2), 217-229. https://doi.org/10.1006/ceps.1994.1018

Law, F., Mahr, T., Schneeberg, A., & Edwards, J. (2017). Vocabulary size and auditory word recognition in preschool children. *Applied Psycholinguistics*, *38*(1), 89-125. doi:10.1017/S0142716416000126

Lawson, & London, K. (2015). Tell Me Everything You Discussed: Children's Memory for Dyadic Conversations after a 1-Week or a 3-Week Delay. *Behavioral Sciences & the Law*, *33*(4), 429–445. https://doi.org/10.1002/bsl.2184

Lenth, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.4.4. https://CRAN.R-project.org/package = emmeans

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). doi: https://doi.org/10.16995/glossa.8582

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., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The Production Effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 385*(6619), 813–815. https://doi.org/10.1038/385813a0

Martin, N. A., & Brownell, R. (2011). *Expressive One-Word Picture Vocabulary Test* (4th ed.). Academic Therapy Publications Assessments.

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology*, *25*(4), 530-535. https://doi.org/10.1016/j.cub.2014.11.068

Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, *3*, 41. https://doi.org/10.3389/fpsyg.2012.00041

Munson, B. (2001). Relationships Between Vocabulary Size and Spoken Word Recognition in Children Aged 3 to 7. *Contemporary Issues in Communication Science and Disorders*, *28*, 20-29. https://doi.org/10.1044/cicsd_28_s_20

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The Production Effect endures and improves memory for text. *Memory, 20*(7), 717-727. https://doi.org/10.1080/09658211.2012.699070

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

Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2019). 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

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the Production Effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language*, *118*, 104219–. https://doi.org/10.1016/j.jml.2021.104219

Sakhon, Edwards, K. K., Luongo, A. A., Murphy, M. M., & Edgin, J. J. (2018). Small sets of novel words are fully retained after 1-week in children with and without Down syndrome: A fast mapping study. *Journal of the International Neuropsychological Society, 24(*9), 955–965. https://doi.org/10.1017/S1355617718000450

Scarf, Gross, J., Colombo, M., & Hayne, H. (2013). To have and to hold: Episodic memory in 3- and 4-year-old children. *Developmental Psychobiology, 55*(2), 125–132. https://doi.org/10.1002/dev.21004

Vihman, M. M. (2022). The developmental origins of phonological memory. *Psychological Review, 129*(6), 1495–1508. https://doi.org/10.1037/rev0000354

Vlach, & Sandhofer, C. M. (2012). Fast Mapping Across Time: Memory Processes Support Children's Retention of Learned Words. *Frontiers in Psychology, 3*, 46–46. https://doi.org/10.3389/fpsyg.2012.00046

Wang, Bui, V.-K., & Song, Q. (2015). Narrative organisation at encoding facilitated children's long-term episodic memory. *Memory, 23*(4), 602–611. https://doi.org/10.1080/09658211.2014.914229

Wang, Weber, F. D., Zinke, K., Inostroza, M., & Born, J. (2018). More Effective Consolidation of Episodic Long-Term Memory in Children Than Adults—Unrelated to Sleep. *Child Development, 89(*5), 1720–1734. https://doi.org/10.1111/cdev.12839

Westfall, J. (2016). PANGEA (v0.2): Power analysis for general ANOVA designs [Unpublished manuscript]. http://jakewestfall.org/publications/pangea.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. (2018). 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

Zamuner, T. S., Yeung, H. H., & Ducos, M. (2017). The many facets of speech production and its complex effects on phonological processing. *British Journal of Psychology*, *108*(1), 37–39. https://doi.org/10.1111/bjop.12220

4 Task effects in children's word recall: Expanding the reverse Production Effect [3]

Abstract

Words said aloud are typically recalled more than words studied under other techniques. In certain circumstances, production does not lead to this memory advantage. We investigated the nature of this effect by varying the task during learning. Children aged five to six years were trained on novel words which required no action (Heard) compared to Verbal-Speech (production), Non-Verbal-Speech (stick out tongue), and Non-Verbal-Non-Speech (touch nose). Eyetracking showed successful learning of novel words in all training conditions, but no differences between conditions. Both non-verbal tasks disrupted recall, demonstrating that encoding can be disrupted when children perform different types of concurrent actions.

Keywords: speech production, word learning, word recognition, recall, child development, task difficulty, eyetracking

[3] López Assef, B. & Zamuner, T. S. (Accepted). Task effects in children's word recall: Expanding the reverse Production Effect. *Journal of Child Language.*

4.1 Introduction

The Production Effect is a phenomenon where items that are produced (e.g., read aloud) during learning or memorization are better remembered than under other conditions (e.g., read silently) (MacLeod et al., 2010). The advantage for produced items seen in the Production Effect has been explained by distinctiveness. Produced items are encoded with additional information, thereby creating a distinctive representation which is subsequently more easily retrieved and remembered (MacLeod et al. 2010). Other actions, such as writing, typing, whispering, gesturing, acting, and drawing also benefit word recognition and recall (Fernandes et al., 2018; Forrin et al., 2012; Mama & Icht, 2016; Mathias et al., 2021; Mayer et al., 2015; Wammes et al., 2018). Conversely, the Reverse Production Effect (Zamuner et al., 2018) is when production disrupts learning or memorization, such that items learned in tasks like listening show better memory than items produced aloud. Researchers have explored various aspects of the Production Effect and the Reverse Production Effect, including its underlying mechanisms (Icht & Mama, 2015; MacLeod et al., 2010; Ozubko et al., 2014; Zamuner et al., 2016), factors influencing memory performance such as stimuli characteristics (e.g. Baese-Berk & Samuel, 2016; López Assef et al., 2023); study design (e.g. Jones & Pyc., 2014), timing of actions performed during the experiment (e.g. Dauphinee et al., 2023), among other manipulations. Research has also examined how the effects interact with participant-related characteristics, such as cognitive and attentional factors and the age of the learner (López Assef et al., 2021; Zamuner et al., 2018). However, compared to adult research, far less has been done with children, and with fewer variables examined. We investigated the nature of the Reverse Production Effect in children by comparing different types of actions to identify if the Reverse Production Effect is specific to speech production or also triggered by other actions.

While results with adults are fairly consistent in finding a Production Effect (although there are exceptions, see overview in López Assef et al. 2024), findings with children show a more complex picture. Some studies with children have shown a Production Effect when tested on a memorization task with real words (five-year-olds in Icht & Mama; 2015) and in a reading task with familiar and novel words (seven-to-ten-year-olds in Pritchard et al., 2019). These contrast with López Assef et al. (2021), which extended Icht and Mama's study with a developmental approach and tested children ages two-to-six years old. Interestingly, López Assef et al. found a typical Production Effect for older children (five-to-six-year-olds), but the opposite with two-to-three-year-olds, who showed a memory advantage for items that had been heard compared to items that were produced aloud. This switch from a recall advantage for heard items to produced items between two-to-six years was explained in terms of task difficulty effects, which depend on the age of the learner, and cognitive and linguistic development. For younger children, producing while memorizing entails a more difficult task than for older children, whose cognitive skills are more developed and have more practice with speech production. As children age, their memory skills improve and they gain more experience with speech production and language, this allows them to take advantage of the distinctiveness for produced items.

In another study with word-learning, Zamuner et al., (2018) also appealed to task complexity and development-related factors to explain their results in which four-and-a-half-to-six-year-olds also showed a Reverse Production Effect. They used eye-tracking to measure participants' recognition of newly learned words in two experiments: Experiment 1 had a mixed design (heard and produced trials combined) with 2 novel words learned under each training condition, and Experiment 2 used a blocked design to lower the task difficulty (separate blocks for the different training conditions) with 4 novel words learned under each training condition.

Children showed a Reverse Production Effect in both experiments, suggesting that speech production disrupted learning. It was argued that the reversal stemmed from task and developmental factors, given that adults tested in a similar design showed a Production Effect (Zamuner et al., 2016). For children, the act of producing novel words during learning was argued to be cognitively demanding. Thus, more processing resources were spent producing items during training rather than learning the mapping between the novel word and referent, compared to just listening. While work with children has investigated familiar words, novel words, reading aloud versus reading silently and naming images versus hearing the label for images, no studies have manipulated the overt task during the learning phase. This would identify if disruption effects are specifically linked to speech production or could extend to performing other tasks while learning. While this is a gap in the Production Effect literature pertaining to children, the effect of different types of actions on learning has been investigated with children, including multisensory learning and gesture. We now shift our focus to a discussion of this work.

Research on multisensory learning/multisensory enrichment, has found that individuals learn more effectively when engaging multiple senses compared to unisensory processes (Andrä et al., 2020; Mayer et al., 2015; Shams & Seitz, 2008; von Kriegstein & Giraud, 2006). For example, infants' learning benefits from having both visual and auditory sensory information (Gogate & Hollich, 2016; Samuelson et al., 2011; Seidl et al., 2023); including sensory information such as touch (Nomikou & Rohlfing, 2011; Seidl et al., 2023; Tincoff et al., 2019). Similar to the contrasting results found for the Production Effect (i.e., the Reverse Production Effect), while multisensory learning can be helpful, the effect depends on multiple factors: it can either enrichen the encoding and retention of new words or it can impede learning with younger children due to their limited processing capacity (see Lewkowicz, 2014).

Research on the effect of gestures on learning follows a similar pattern. Gestures have been shown to support learning across various domains. For instance, gestures can be beneficial during mathematical learning (Cook et al., 2012) and facilitate word learning in both first language (Goodrich & Kam, 2009; McGregor et al., 2009) and second language (Tellier, 2008). Incorporating gestures in educational settings has been shown to promote learning. For example, students benefit more from lessons that include gestures compared to those without (Alibali et al., 2013), students learn more effectively when teachers use speech with gestures (Congdon et al., 2017; Singer & Goldin-Meadow, 2005; Wakefield et al., 2018), and when students are encouraged to gesture during lessons (Broaders et al., 2007; Brooks & Goldin-Meadow, 2016; Goldin-Meadow et al., 2009). However, the efficacy of gestures depends on several factors, including children's cognitive development and the context in which gestures are used. For example, children with high phonological competence (sounding out novel words) benefit more from speech and gesture instruction compared to those with low phonological competence (Wakefield & James, 2015). Thus, while gestures can be a powerful instructional tool, their efficacy may depend on children’s cognitive development and familiarity with the subject. Similar results with gesture and task (stimuli familiarity) has been found with adults (e.g., Kelly & Lee, 2011).

Overall, the effectiveness of multisensory learning and gestures for memory and learning depends on factors related to characteristics of the learner (cognitive skills and development), and characteristics of the material being learned (familiarity, complexity), similar to the Production Effect. However, to date no research on children has looked at whether the disruption in learning associated with speaking extends to other learning situations. As such, the current study investigates whether the Reverse Production Effect is triggered primarily by linguistic factors, such as producing speech during learning, or whether it can be a consequence of overall task difficulty, caused by performing concurrent actions during learning. Children aged five-to-six years old were trained on novel words paired with nonce animals. During training, half of the novel words were *Heard*, while the other half belonged to one of three *Action* conditions. The *Action* conditions were designed to create a continuum for language specificity, to compare the effects of linguistic and non-linguistic actions on word learning. Participants were assigned randomly to one of Heard-Action Condition Pairs. One *Action* group was instructed to repeat the novel word aloud (*Verbal/Speech* condition), the second group to stick out their tongue (*Verbal/Non-Speech* condition, articulators are engaged without speech production) and the third group was told to touch their nose with their finger (*Non-Verbal/Non-Speech* condition, articulators are not engaged, no speech production).

Based on previous literature showing that the direction of the Production Effect is sensitive to linguistic and cognitive/task difficulty effects, one could expect that novel words learned under all *Action* training conditions would show lower recognition and lower recall rates than the *Heard* condition. However, within the *Action* training conditions, more demanding actions may lead to greater disruptions in learning. Under this scenario, we hypothesized that *Verbal/Speech* could cause the biggest disruption, due to the linguistic processes involved in speech production that are missing from the other two *Action* conditions: e.g., motor-related processes for articulation, activation and retrieval of linguistic representations of the words to be spoken. On the other hand, *Verbal/Non-Speech* and *Non-Verbal/Non-Speech* may also be more demanding, as children are less familiar with the actions compared to *Verbal/Speech*. Children have daily experience with *Verbal/Speech,* also within the context of learning during school-based activities, and children will often spontaneously imitate (repeat) novel words (see review in Zamuner & Thiessen, 2018). Lastly, another possibility is that if the Reverse Production Effect is caused by verbal factors (motor planning of speech articulators), then both the *Verbal/Speech* and *Verbal/Non-Speech* conditions should show lower recognition and recall rates compared to the *Heard* condition.

4.2 Methods

4.2.1 Participants

English-speaking children aged five-to-six years (*N*=48) were randomly assigned to one of the three Condition Pairs: *Verbal/Speech-Heard* (*n* = 16, 8 males, 8 females, *MAge* = 71.3 months, *SD* = 8.4, *range* = 60 – 83), *Verbal/Non-Speech-Heard* (*n* = 16, 4 males, 12 females, *MAge* = 71.8 months, *SD* = 8.7, *range* = 60 – 83) and *Non-Verbal/Non-Speech-Heard* (*n* = 16, 6 males, 10 females, *MAge* = 71.8 months, *SD* = 5.4, *range* = 65 – 80). Participants had a minimum 70% lifetime exposure to English (*M* = 91.5%, *SD* = 8.3*, range* = 70 – 100). All children had learned English from birth, with no more than two consecutive years of 30+% exposure to another language, as estimated from a language background questionnaire completed by parents. All children were also reported to have normal hearing, normal-to-corrected vision, and no history of language impairment. Families were recruited on the museum floor at a Living Lab and children received a sticker for participating. Eighteen additional children were tested but not included in the analyses for having no video for off-line coding due to technical issues (*n* = 5), did not perform the task properly, e.g., produced all words in the *Heard* condition (*n* = 11), and not being able to code the actions because children were wearing a mask (*n* = 2).

4.2.2 Stimuli

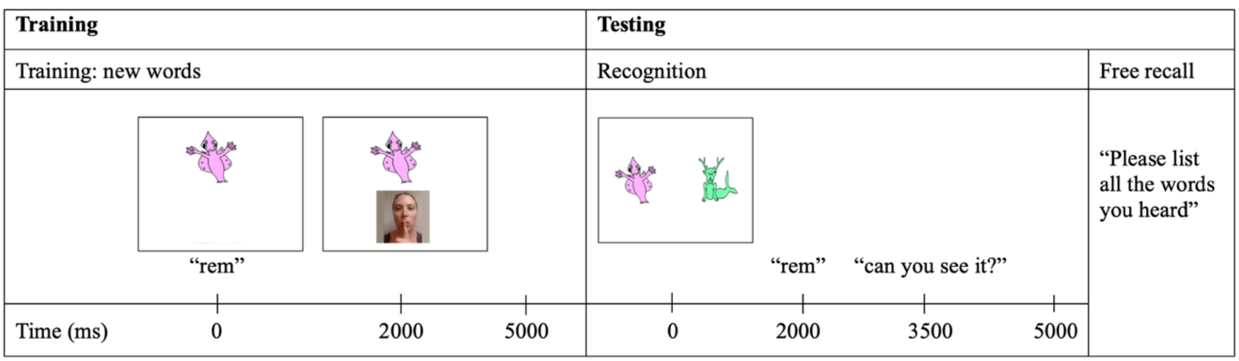
There were 16 novel words paired with nonce images, divided into two lists for counterbalancing (List 1: wis, zel, vup, bos, gub, mig, das, rem; List 2: nis, kel, tup, los, fub, jig, has, pem). Each list had 4 items *Heard* and 4 items *Action* condition. Audio stimuli were recorded by a female native English speaker and controlled for pitch, length and amplitude. The visual stimuli consisted of nonce animals developed by Ohala (1996, p. 60). Audio and visual stimuli were the same as in Experiment 2 from Zamuner et al. (2018). Audio stimuli were recorded by a native speaker of English, normalized for amplitude (70 dB). Stimuli were not controlled for their phonotactic probabilities and neighbourhood densities. However, the stimuli did not contain low frequency sound patterns, calculated using Storkel's (2013) child directed speech materials (see Zamuner et al., 2018, pp. 4-5).

4.2.3 Design

The design was adapted from Experiment 2 from Zamuner et al. (2018). The experiment was blocked: the first block included practice training (4 trials), practice testing (4 trials), experiment training (16 trials) and experiment testing (16 trials), free recall for the Heard condition, which was then followed by the second block which included practice training, practice testing, experiment training and experiment testing, free recall for one of the Action conditions. This was counterbalanced by participants for whether the Heard or Action condition was the in the first block. Practice had the same procedure as the experimental sections and was used to familiarize participants with the learning conditions, using real words paired with real images (fruits).

The design is illustrated in Figure 9. During training, participants saw a fruit (practice) or nonce animal (experiment) and heard the corresponding audio stimuli. After 1000 ms, the learning condition appeared below as an image. Participants were trained on the learning condition-image correspondence during the practice trials. With the picture of a woman gesturing "shh", participants were asked to remain silent (*Heard*). With a picture of a finger pointing outwards, participants were instructed to repeat the word (*Verbal/Speech*). Two learning conditions were added to the original Zamuner et al. (2018) study: a picture of a woman sticking out her tongue, (*Verbal/Non-Speech*) and a picture of a woman touching her nose (*Non-Verbal/Non-Speech*), in both cases participants were instructed to imitate the woman in the picture. Following Zamuner et al., (2018)'s design, participants heard the audio stimuli a second time after the presentation of the learning condition for all conditions except for *Verbal/Speech*, in order to have an equal number of exposures to the stimuli across all learning conditions.

Figure 9 *Example of training and testing trials for an item under the Heard training condition.*



Note. All blocks for all training conditions followed the same procedure, with the exception of the Verbal/Speech condition in which novel words were not repeated during training, but instead were produced by children.

Following training, participants were tested on their recognition of novel words. Two nonce animals (target and distractor) were shown on the screen for 2000 ms, after which participants heard a novel word corresponding to one of the pictures on the screen. To keep participants' attention on the images we also included phrases such as "Do you see it?" and "Can you find it?". Each novel word appeared twice during training and twice as a target during testing. Afterwards participants performed a free recall task in which they were asked to name all the novel words they remembered. For the second block, participants repeated the same procedure with the other learning condition. In Zamuner et al. (2018), free recall was only done once, at the end of the entire experiment. This was modified so that participants did the free recall task twice - after the *Heard* and *Action* blocks.

4.2.4 Procedure

Children were tested in a sound-attenuated room at a museum-based lab. Parents remotely observed their child on an iPad from an adjoining waiting area. Eye movements were recorded using an Eyelink 1000 Plus eyetracker in monocular remote mode (SR Research, Ottawa). Children were audio and video recorded using a GoPro Hero7 camera.

4.3 Results

4.3.1 Coding

Each session was coded off-line by two researchers for children's responses during training and recall. This ensured that children gave the correct response for the training condition (e.g., remaining silent during *Heard* items, saying the word aloud for *Verbal/Speech* trials, touching their nose for *Non-Verbal/Non-Speech* trials, and sticking out their tongue for *Verbal/Non-Speech* trials). Children's pronunciation of words during training and recall were transcribed and coded for accuracy. To be counted as a correct pronunciation, participants had to correctly produce both consonants for each word. For example, for the item *zel,* [zʌl] was counted as a correct pronunciation, regardless of the vowel being pronounced incorrectly. Items with training errors were excluded from both recall and recognition analysis (n = 39, 5% of total trials): producing *Heard* items (n = 6), talking or noises during trials (n = 3), looking away from the screen (n = 1), producing the word more than once (n = 7), not performing the correct action (n = 8), mouthing words (n = 10) and mispronunciations during training (n = 4). Items incorrectly produced were excluded from recall analyses (but not recognition) if they had incorrect pronunciations during the recall task, following the same criteria as for training mispronunciations (n = 65 items). Words that were recalled more than once were counted as a single recall.

4.3.2 Free Recall task

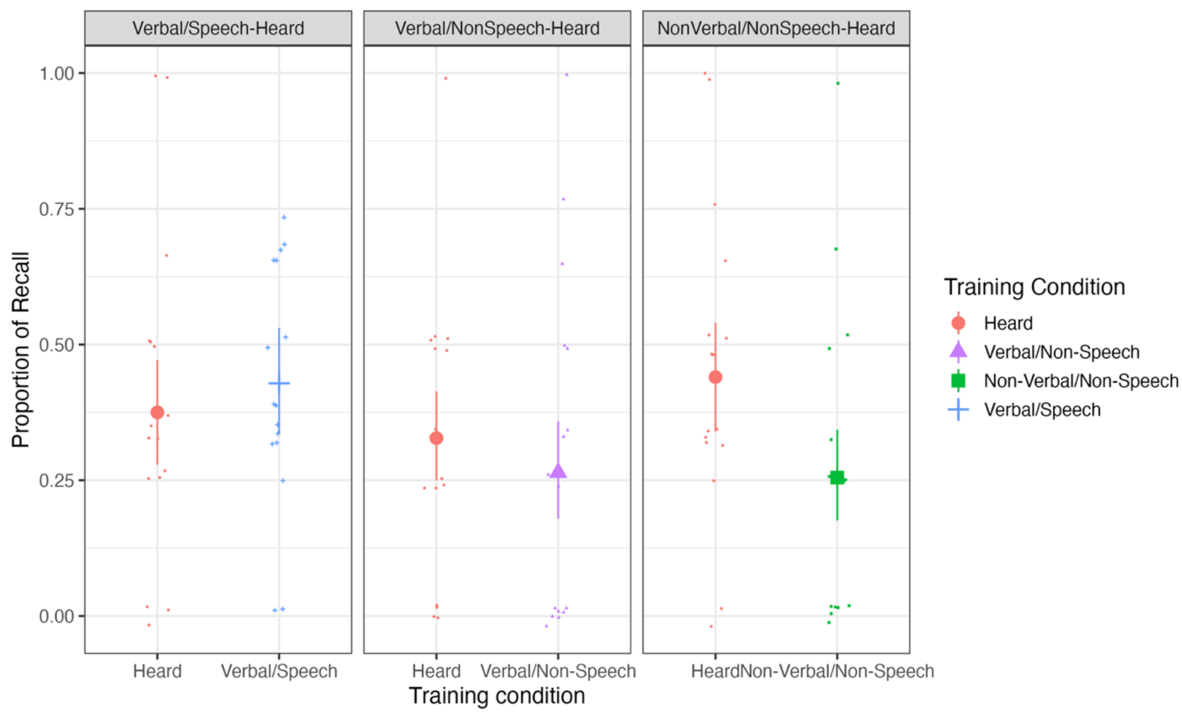
Our first analysis used generalized linear models (glmer()) to examine the effect of training condition on recall. Our dependent variable was recall (yes, no). Analyses were conducted separately for each *Heard-Action* pair. Our fixed effects were Training Condition (*Heard,* and the corresponding Action condition: *Verbal/Speech, Verbal/Non-Speech, Non-Verbal/Non-Speech*) and Training Order (i.e., order of Training Conditions in Blocks: *Heard in Block 1, Action in Block 1*). Training Order was included as a control variable to account for potential recency effects. Figure 10 illustrates the results for the recall task and Table 10 provides the average recall rates for each Training Condition Pair.

Table 10 *Recalls and proportion of target fixations average and standard deviation by Training Condition (Verbal/Speech– Heard, Verbal/Non-Speech– Heard, Non-Verbal/Non-Speech– Heard).*

Condition Pair	Training Condition	**Recalls**		**Proportion of Target Fixations**	
		M	*SD*	*M*	*SD*
Verbal/Speech-Heard	Verbal/Speech	0.43	0.50	0.65	0.30
	Heard	0.38	0.49	0.66	0.32
Verbal/NonSpeech-Heard	Verbal/Non-Speech	0.26	0.44	0.61	0.30
	Heard	0.33	0.47	0.66	0.33
	Non-Verbal/Non-Speech	0.25	0.44	0.66	0.31

NonVerbal/NonSpeech-Heard	Heard	0.44	0.50	0.69	0.31

Figure 10 Proportion of Recall by Training Condition (Verbal/Speech – Heard, Verbal/Non-Speech – Heard, Non-Verbal/Non-Speech – Heard).



Note. Points are the condition means by participant with error bars indicating 95% confidence intervals.

Statistical analyses were performed in R (v4.2.0, R Core Team) using the glmer() function from the lme4 package (v1.1-27.1, Bates et al., 2015). We started from the most complex model structure, containing random by-participant intercepts, by-participant slopes for Training Condition, random by-item intercepts, and by-item slopes for Training Condition and

Training Order, and their interactions. Models were simplified to account for convergence issues and singularity errors until a model converged. Table 11 provides results from the three models (corresponding to each condition pair), along with the final model structure for each *Heard-Action* pair.

Table 11 *Results from model estimating free recall by Training Condition (Verbal/Speech– Heard, Verbal/Non-Speech– Heard, Non-Verbal/Non-Speech– Heard) and Training Order (Heard in Block 1, Action in Block 1).*

Fixed Effects	Estimate	*SE*	*z value*	p-value
Verbal/Speech-Heard				
Training Condition	-0.39	0.76	-0.51	0.61
Training Order	0.62	1.02	0.60	0.55
TrainingCondition*TrainingOrder	2.73	1.92	1.42	0.16
Verbal/Non-Speech-Heard				
Training Condition	8.61	3.88	2.22	0.027*
Training Order	8.10	5.74	1.41	0.16
TrainingCondition*TrainingOrder	1.44	10.16	0.15	0.89
Non-Verbal/Non-Speech-Heard				
Training Condition	1.72	0.74	2.31	0.021*
Training Order	-0.35	0.92	-0.38	0.70
TrainingCondition*TrainingOrder	1.00	2.17	0.46	0.64

Note. The final models had the following syntax specified in the lme4 package:

Verbal/Speech-Heard: recalled ~ traincondition_sum*trainorder_sum + (1|participants) + (1+traincondition_sum |target)

Verbal/Non-Speech-Heard: recalled ~ traincondition_sum*trainorder_sum + (1+traincondition_sum|participants) + (1+trainorder_sum |target)

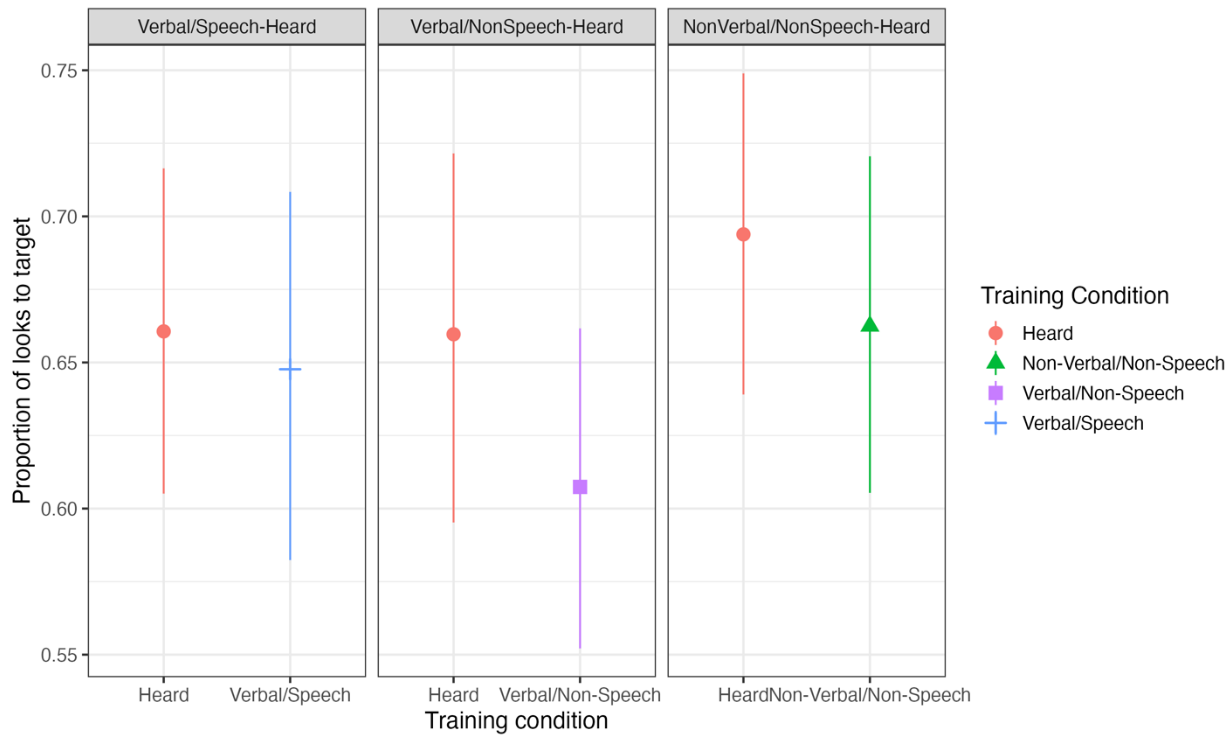
Non-Verbal/Non-Speech-Heard: recalled ~ traincondition_sum*trainorder_sum + (1+traincondition_sum|participants) + (1+trainorder_sum |target)

Statistical analyses did not show any significant effects in the *Verbal/Speech-Heard* group. No main effect of Training Condition was found (β = -0.39, *SE* = 0.76, *p* = .61), indicating no statistical difference in recall between *Heard* and *Verbal/Speech* items. For the *Verbal/Non-Speech-Heard* group, there was a significant effect of Training Condition (β = 8.6, *SE* = 3.88, *p* = <.05). For the *Non-Verbal/Non-Speech-Heard* group, there was also a significant effect of Training Condition (β = 1.72, *SE* = 0.74, *p* = <.05). Post-hoc comparisons were done on the final models with the emmeans package, using asymptotic estimations for degrees of freedom (v1.7.4-1; Lenth, 2020). Estimated marginal means backtransformed into proportions are reported. Post-hoc tests revealed a higher probability of recall for *Heard* words (*M* = 0.31, *SE* = 0.01, *95% CI* [0-0.95]) than *Verbal/Non-Speech* words (*M* = 0, *SE* = 0, *95% CI* [0-0.02]), and a higher probability of recall for *Heard* words (*M* = 0.46, *SE* = 0.16, *95% CI* [0.19-0.75]) than *Non-Verbal/Non-Speech* words (*M* = 0.13, *SE* = 0.09, *95% CI* [0.03-0.42]).

4.3.3 Recognition task

Recognition analyses were based on the proportion of looks to target, from 200ms after word onset to allow time to initiate eye movements, and the window ended at 1500ms, emulating the window of analyses used in Zamuner et al., (2018). Figure 11 illustrates the results for the proportion of fixations to targets and Table 12 provides the averages for each Training Condition Pair.

Figure 11 *Proportion of Looks to Target by Training Condition (Verbal/Speech – Heard, Verbal/Non-Speech – Heard, Non-Verbal/Non-Speech – Heard).*



Note. Error bars indicate 95% confidence intervals

Proportion of looks to the target was the dependent variable for the linear mixed-effects models performed in R (R Core Team) using the lmer() function from the lme4 package (version 1.1-26; Bates et al., 2015). The methodology was the same as with recall analyses. Table 12 provides results from the three models (corresponding to each condition pair), along with the final model structure for each *Heard-Action* pair. Statistical models did not show significant effects for any of the Training Conditions and groups. Although there were no differences in the proportion of looking to targets in the different conditions, as seen in Table 12 and Figure 11, children looked above chance (50%) to the correct images.

Table 12 *Results from model estimating proportion of target fixations by Training Condition (Heard – Verbal/Speech, Verbal/Non-Speech, Non-Verbal/Non-Speech) and Training Order (Heard in Block 1, Action in Block 1).*

Fixed Effects	*F*-value	*df*	*df.Res*	*p*-value
Verbal/Speech-Heard				
Training Condition	*0.06*	*1*	*117.37*	*0.80*
Training Order	*0.48*	*1*	*13.46*	*0.50*
TrainingCondition*TrainingOrder	*0.30*	*1*	*13.99*	*0.60*
Verbal/Non-Speech-Heard				
Training Condition	0.74	1	13.14	0.40
Training Order	0.23	1	13.67	0.64
TrainingCondition*TrainingOrder	0.0006	1	12.75	0.99
Non-Verbal/Non-Speech-Heard				
Training Condition	0.35	1	13.17	0.56
Training Order	0.59	1	13.63	0.46
TrainingCondition*TrainingOrder	0.001	1	10.80	0.97

Note. Wald F-tests with Kenward-Roger estimates for df. The final models had the following syntax specified in the lme4 package:

Verbal/Speech-Heard: Target_Proportion ~ traincondition_sum * trainorder_sum + (1 | participants) + (1 | target)

Verbal/Non-Speech-Heard: Target_Proportion ~ traincondition_sum * trainorder_sum + (1 + traincondition_sum | participants) + (1 | target)

Non-Verbal/Non-Speech-Heard: Target_Proportion ~ traincondition_sum * trainorder_sum + (1 + traincondition_sum | participants) + (1 | target)

4.4 Discussion

The present study investigated whether the Reverse Production Effect (López Assef et al., 2021; Zamuner et al., 2018) was caused by children's speech productions or whether it could also be triggered by performing different types of concurrent tasks during word learning. Our free recall task shows an interesting pattern, which did not fully fit into any of our predictions. The *Verbal/Non-Speech-Heard* and *Non-Verbal/Non-Speech-Heard* condition pairs showed a significant effect of Training Condition: in both cases children recalled more newly learned words from the *Heard* condition than the two *Non-Speech Action* conditions. No significant effects were found for the *Verbal/Speech-Heard* pair: repeating items aloud during the training task did not result in a learning disruption, nor a memory advantage.

Eye-tracking results show no differences across all three *Heard-Action* pairings (*Verbal/Speech-Heard; Verbal/Non-Speech-Heard; Non-Verbal/Non-Speech-Heard*). Participants were above chance for target looking in all Training Conditions (Figure 3), indicating successful learning in all Training Conditions. The contrast between our free recall results, which showed significant differences, and our eye-tracking results, which did not, may be attributed to the distinct nature of the tasks. Eye-tracking captures online processing, reflecting the ongoing activation of a word, whereas free recall is an offline task that assesses the outcome of this activation and processing. It is possible that similar patterns emerge during the activation and processing stages across all actions, but differences between training conditions emerge only at the final stage of retrieval. Further research is needed to better understand the differences between online and offline tasks and the Production Effect.

Our eye-tracking results are in line with Zamuner et al. (2018) given that both did not find a Production effect; however, while the current study had null eyetracking results, Zamuner et al. found a significant Reverse Production Effect in eyetracking, with greater looking to targets in the *Heard* condition versus the *Verbal/Speech* condition. While the ages of our participants match those in Zamuner et al., data from the two studies were collected using different language background questionnaires, and after the roll-out of full-day kindergarten for four- and five-year-old children. Thus, it is possible that the same aged children had slightly different language backgrounds and different experiences that map to the experimental task (e.g., learning and repeating novel words). This is relevant because the direction of the Production Effect depends on both developmental and linguistic factors, where familiarity and exposure have been identified as important variables. For instance, greater proficiency or amount of language exposure may increase the likelihood of a child exhibiting the Production Effect. Similarly, experience with structured learning tasks, such as those encountered in daycare or school, could reduce the overall difficulty of our experimental task, increasing the chances of observing the Production Effect. Additional factors that might contribute to a transition from a Reversal Effect to a Production Effect include improvements in language and cognitive skills. Thus, it is possible that children in the current study are at a transition stage of moving from a Reverse Production Effect to a Production Effect (recall that adults tested on the same materials showed the Production Effect). Regardless, the learning disruption caused by speech production was less pronounced in our participants compared to those in Zamuner et al. (2018). Future work looking at longitudinal and language experience factors are needed to shed light on the direction of the production effect across development.

Going back to our research question, we asked whether speech production makes the word learning task more difficult, or whether performing *any* action during learning disrupts encoding. The results from recall align most closely with the prediction that the Reverse Production Effect is influenced by the overall demands of the task; more challenging tasks lead to a greater disruption in learning. Similar results have also been found with adults: performing any action, in this case both producing the experimental stimuli or producing random letters during training, can disrupt learning and lead to a Reverse Production Effect (Baese-Berk & Samuel, 2016). Both the *Verbal/Non-Speech-Heard* and *Non-Verbal/Non-Speech-Heard* groups showed better recall for novel words learned under the *Heard* condition compared to the *Action* conditions, suggesting that listening alone is less demanding compared to listening while sticking out your tongue or touching your nose. This disruption was not found in the *Verbal/Speech-Heard* condition pair, as children's recall was the same across conditions. We postulated that *Verbal/Speech* may pattern differently as children have more experience with repeating words compared to the other actions. *Verbal/Speech* words may have received a Production-Effect-like boost, leading to them being more distinctive and easier to retrieve than words learned under the other *Action* conditions. Therefore, while there is still an overall difficulty effect, *Verbal/Speech* words received a memory boost from speech production, resulting in the memory advantage we see when comparing *Verbal/Speech* versus the other *Action* conditions (Figure 2). Thus, we hypothesize that children are at a transition stage, between a Reverse Production Effect and a Production Effect. See Richtsmeier et al. (2024) for similar flat results in a novel word-learning task with three-and four-year-old children which manipulated both training condition (*Heard* versus *Verbal/Speech*) and the presence or absence of semantic cues. Children's mere above chance performance on an identification task suggests that Richtsmeier et al.'s task and stimuli were more complex. They found no effect of *Verbal/Speech* in either direction. As noted by Richtsmeier et al., the field on the Production Effect and Reverse Production Effect with children is relatively new, and future work is needed to establish how the effects may or may not generalize across variables, such as the task and stimuli.

Our results are consistent with the multisensory learning literature and gesture literature, where the possible beneficial effect from these actions on learning and memorization will depend on cognitive-related factors, such that these actions can either aid in the encoding process and retention, or on the other hand, hinder the process by overwhelming learners due to limitations in their cognitive skills. Gaining a better understanding of the Production Effect and Reverse Production Effect and its intricacies has both theoretical and practical implications. This can inform psycholinguistic theories on the relationships between perception and production; and it can inform educational practices, as it has implications for how engaging in generating information during studying may and may not enhance learning outcomes.

References

Alibali, M. W., Young, A. G., Crooks, N. M., Yeo, A., Wolfgram, M. S., Ledesma, I. M., ... & Knuth, E. J. (2013). Students learn more when their teacher has learned to gesture effectively. *Gesture*, *13*(2), 210-233. https://doi.org/10.1075/gest.13.2.05ali

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von 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

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

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

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136,* 539–550. http://dx.doi.org/10.1037/0096-3445.136.4.539

Brooks, N., & Goldin-Meadow, S. (2016). Moving to learn: How guiding the hands can set the stage for learning. *Cognitive Science*, *40*(7), 1831–1849. https://doi.org/10.1111/cogs.12292

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, *50*, 65–74. https://doi.org/10.1016/j.learninstruc.2017.03.005

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language & Cognitive Processes*, *27*, 594–610. https://doi.org/10.1080/01690965.2011.567074

Dauphinee, I., Roy, M., Guitard, D., Yearsley, J. M., Poirier, M., & Saint-Aubin, J. (2024). Give me enough time to rehearse: Presentation rate modulates the Production Effect. *Psychonomic Bulletin & Review*, 1-12. https://doi.org/10.3758/s13423-023-02437-5

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

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science*, *27*(5), 302-308. https://doi.org/10.1177/0963721418755385

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

Gogate, L., & Hollich, G. (2016). Early verb-action and noun-object mapping across sensory modalities: A neuro-developmental view. *Developmental Neuropsychology*, *41*, 293–307. https://doi.org/10.1080/87565641. 2016.1243112

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20,* 267– 272. http://dx.doi.org/10.1111/j.1467-9280.2009.02297.x

Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science*, *12*, 81–87. https:// doi.org/10.1111/j.1467-7687.2008.00735.x

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

Kelly, S. D., & Lee, A. L. (2011). Language and cognitive processes when actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high demands are high. *Language and Cognitive Processes, 27,* 37–41. https://doi.org/10.1080/01690965.2011.581125

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology*, *56*(2), 292–315. https://doi.org/10. 1002/dev.21197

López Assef, B., McDonald, M., Bernard, A., & Zamuner, T. S. (2024). Effects of speech production training on memory across short and long delays in 5-and 6-year-olds: A pre-registered study. *Applied Psycholinguistics*, 1-24. https://doi.org/10.1017/S0142716424000109

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). https://doi.org/10.16995/glossa.8582

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The Production Effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

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

Mathias, B., Waibel, A., Hartwigsen, G., Sureth, L., Macedonia, M., Mayer, K. M., & von Kriegstein, K. (2021). Motor cortex causally contributes to vocabulary translation following sensorimotor-enriched training. *The Journal of Neuroscience, 41*(41), 8618–8631. https://doi.org/10.1523/JNEUROSCI.2249-20.2021

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology*, *25*(4), 530-535. https://doi.org/10.1016/j.cub.2014.11.068

McGregor, K. K., Rohlfing, K. J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language, 36*, 807–828. doi:10.1017/S0305000908009173

Nomikou, I., Koke, M., & Rohlfing, K. J. (2017). Verbs in mothers' input to six-month-olds: Synchrony between presentation, meaning, and actions is related to later verb acquisition. *Brain Sciences*, *7*(5), 52. https://doi.org/ 10.3390/brainsci7050052

Ohala, D.K. (1996). Cluster reduction and constraints in acquisition. *Doctoral dissertation,* The University of Arizona.

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory, 22*(5), 509–524. https://doi.org/10.1080/09658211.2013.800554

Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2019). 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

Richtsmeier, P. T., Gladfelter, A., & Moore, M. W. (2024). Contributions of speaking, listening, and semantic depth to word learning in typical 3-and 4-year-olds. *Language, Speech, and Hearing Services in Schools*, 1-14. https://doi.org/10.1044/2024_LSHSS-23-00105

Samuelson, L., Smith, L., Perry, L., & Spencer, J. (2011). Grounding word learning in space. *PLoS ONE*, *6*(12), e28095. https://doi.org/10.1371/ journal.pone.0028095

Seidl, A. H., Indarjit, M., & Borovsky, A. (2024). Touch to learn: Multisensory input supports word learning and processing. *Developmental Science*, *27*(1), e13419. https://doi.org/10.1111/desc.13419

Shams, S., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*, 85–89. https://doi.org/10.1111/j.0956-7976.2005.00786.x

Storkel, H.L. (2013). A corpus of consonant-vowel- consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods, 45,* 1159–1167. https://doi.org/10.3758/s13428-012-0309-7

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219-235. https://doi.org/10.1075/gest.8.2.06tel

Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2019). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development*, *15*(2), 103– 125. https://doi.org/10.1080/15475441.2018.1533472

von Kriegstein, K., & Giraud, A. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, *4*(10). https://doi.org/10.1371/journal.pbio.0040326

Wakefield, E. M., & James, K. H. (2015). Effects of learning with gesture on children's understanding of a new language concept. *Developmental Psychology*, *51*, 1105–1114. https://doi.org/10.1037/a0039471

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science*, *21*, e12664. https://doi.org/10.1111/desc.12664

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2018). Creating a recollection-based memory through drawing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(5), 734–751. https://doi.org/10.1037/xlm0000445

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. (2018). Reverse Production Effect: Children recognize novel words better when they are heard rather than produced. *Developmental Science, 21*(4). https://doi.org/10.1111/desc.12636

Zamuner, T. S., & Thiessen, A. (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics*, *63*(4), 609-632. https://doi.org/10.1017/cnj.2018.10

5 Word learning in a storybook setting: Recognition is better for younger children when they simply listen rather than produce aloud[4]

Abstract

Studies on the Production Effect with children have shown mixed results. While some studies have shown the expected memory advantage for items spoken aloud during learning, other studies have found an advantage for items only heard or no difference between learning conditions. The lack of a Production Effect has been argued to partly stem from cognitive factors. The present pre-registered study investigated whether adapting a more familiar learning task would lower task difficulty effects and result in a positive effect of speech production on learning. Children aged 4-(*n*=20) and 6-(*n*=20) years-old were taught novel words which were either *Heard* or produced aloud (*Say*) in a shared storybook task, while also manipulating the number of exposures (*1 Presentation*, *4 Presentations*). Learning was measured by a combined eye-tracking and pointing task. Results did not show a Production Effect across any of the manipulations. For the eye-tracking data, there was a significant three-way interaction between Training Condition, Number of Exposures And Age, such that 4-year-olds showed better recognition for items that were *Heard* and had *4 Presentations*, while 6-year-olds showed no difference in learning between the conditions. For the pointing data, items with *4 Presentations* also showed higher accuracy. No memory advantage for produced items was found. This highlights the complex effect of production on memory and the importance of different approaches to understanding the relationship between speech production and learning.

---

[4] López Assef, B. & Zamuner, T. S. (Submitted). Word learning in a storybook setting: Recognition is better for younger children when they simply listen rather than produce aloud *Developmental Psychology*.

Keywords: Production Effect, Reverse Production Effect, word learning, speech production, repeated exposures, word recognition, storybook reading

5.1 Introduction

To learn a novel word, learners must acquire information about the word's referent (what the word stands for), the label (the word itself), and the connection between the two. The ability to map referents and labels is present from infancy (Waxman & Gelman, 2009; Werker et al., 1998; Zamuner et al., 2014) and further develops as children grow older, so that by the time they reach adulthood they are estimated to understand approximately 60,000 words (McMurray, 2007). Speech production has been suggested as a key factor in language acquisition from the earliest stages. For instance, research shows that infants' speech production influences speech processing even before they start producing adult-like speech (Bruderer et al., 2015; DePaolis et al., 2011, 2013; Yeung & Werker, 2013). The impact of speech production extends beyond language perception and processing; it also influences our ability to memorize and learn words, a phenomenon known as the Production Effect.

The Production Effect refers to a memory advantage for words that are said aloud during learning or memorization compared to words that have been studied under other actions, such as listening or reading silently (MacLeod et al., 2010). The Production Effect has been replicated under different manipulations, some of the most relevant for the present study: different age groups (e.g., 5-year-old children, Icht & Mama, 2015; adults, MacLeod et al., 2010), testing methodologies (e.g., free recall task, Cho & Feldman, 2016; recognition task, Zamuner et al., 2016; translation task, Kaushanskaya & Yoo, 2011), and stimuli characteristics (e.g., images, Icht et al., 2020; auditory stimuli, Forrin & MacLeod, 2016). While some studies with children have found a Production Effect (Icht & Mama, 2015; López Assef et al., 2021; Pritchard et al., 2019), other studies with children have found contrary patterns, finding a memory advantage for items that were heard during training over those produced aloud, labeled Reverse Production Effect (Zamuner et al., 2018). Thus, while the Production Effect is generally believed to benefit our memory of words, studies suggest that it can be affected by multiple factors, which can be specific to the learner (developmental and experience-related factors) and the task (types of production and learning tasks), see Baese-Berk et al., 2024 and Zamuner et al., 2017.

The current study investigates whether framing a Production Effect task within a more familiar context for 4-and-6-year-olds, will lead to a positive effect of production during word learning. This was done by using a storybook setting and including repeated exposures to the novel items. The study's methodology, hypotheses and analysis plan were pre-registered on OSF (https://osf.io/swmgz/?view_only=b001a93b12944368ad938e2ca657e438). Exploring the contexts in which speech production results in a memory advantage for children can shed light on the various factors influencing the Production Effect, helping to determine when speaking or listening is a more effective learning strategy.

5.1.1 Factors in the Production Effect and Reverse Production Effect

The Production Effect has been argued to stem from distinctive encoding that occurs when learners actively engage with items via speech production, making the memory trace more memorable compared to other items which lack this distinctiveness. The Production Effect fits within the same literature showing that actively engaging with information, like acting out instructions (Allen et al., 2020) or drawing images (Fernandes et al., 2018), can improve our memory for items compared studying them more passively, such as only listening to labels or reading silently.

While most research examining the Production Effect has found a memory advantage for items produced aloud, more recent research suggests that the Production Effect depends on multiple factors. For example, one of the factors that has resulted in mixed results is language familiarity, with linguistically familiar stimuli showing a Production Effect (Bodner & Taikh, 2012; López Assef et al., 2023), and linguistically unfamiliar stimuli showing a Reverse

Production Effect or no effect either way (Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Kaushanskaya & Yoo, 2011; López Assef et al., 2023). Furthermore, mixed results have also been found depending on the methodology. For example, while many different types of tasks have successfully measured the Production Effect, such as free recall task (Cho & Feldman, 2016, López et al., 2021), eye-tracking recognition task (Zamuner et al., 2016), and translation task (Kaushanskaya & Yoo, 2011), some tasks appear to be better at detecting the effect than others. For example, Bodner and Taikh (2012) compared different tasks while keeping all other experimental conditions constant. Their results revealed a Reverse Production Effect during a list discrimination task, but a Production Effect when using an old/new recognition task. They suggested that the specific characteristics and inherent biases of the list discrimination task may have blocked the Production Effect from being detected.

5.1.2 The Production Effect And Reverse Production Effect With Children

While the previous studies were done with adults, studies with children also show mixed results: some studies show the Production Effect (Icht & Mama, 2015; Pritchard et al., 2019), while other studies show an attenuation or reversal of the effect (López Assef et al., 2021, 2024; Richtsmeier et al., 2024; Zamuner et al., 2018). Icht and Mama's (2015) experiment tested recall for familiar and unfamiliar words with visual referents across different training conditions. Five-year-olds were shown pictures of familiar and unfamiliar objects which could correspond to one of the three learning conditions, chosen based on the different number of encoding processes they each entail: *Look* (only used for familiar items), *Heard* ('Look and Listen' in the original article, terms changed for consistency across studies) and *Say* ('Look and Say' in the original study). Children's recall rates patterned as a Production Effect, with highest recall for *Say,* items, followed by *Heard*, and lastly, followed by *Look*. Pritchard et al. (2019) also found a Production Effect in children aged 7-to-10-years. Children were trained to read words and non-words aloud or silently. They were then tested on a recognition task, where they were asked to indicate whether or not they remembered the item. Pritchard et al.'s experiments with real words and novel words found a memory advantage for *Read-Aloud* vs. *Read-Silently*.

The above studies are in contrast to the results found in López Assef et al. (2021, 2024), Richtsmeier et al. (2024) and Zamuner et al. (2018). López Assef et al. (2021) investigated the direction of the Production Effect across development in children aged 2-to-6-years old, using a similar paradigm as Itch and Mama (2015). Children were trained on familiar words paired with images across three different conditions: *Look*, *Heard* and *Say*, and were later tested on a free recall task. There was a change in the direction of the Production Effect across development: a Reverse Production Effect for children aged 2-to-3-years, that shifted to a Production Effect for the 6-year-olds. Similar results were found by Richtsmeier et al. (2024) who trained 3- and 4-year-olds on a novel word learning task of aliens' names where they manipulated semantic depth (semantic cues in the form of characteristics of the visual stimuli) and training condition (*Say* and *Heard)*. Using an identification and naming task to measure learning, they found that while children were better at identifying alien names when semantic depth was involved, no advantage was found for either training condition. In other studies, direct comparisons between children and adults using the same novel word learning paradigm showed contrasting results: adults showed a Production Effect (Zamuner et al., 2016), whereas children showed a Reverse Production Effect (Zamuner et al., 2018), indicating that children were not able to take advantage of the distinctive representation from produced items in the same manner as adults. Drawing on developmental frameworks that model task difficulty (Curtin et al., 2011; Vihman et al., 2014; Werker & Curtin, 2005), Zamuner et al. (2018) proposed that the differences in performance between children and adults stemmed from the added challenge children face in novel word learning tasks. The task was more difficult for children, as not only do children have less experience with word learning, but they also have less experience producing speech. Thus, during the novel-word learning task, children needed additional cognitive resources to produce the unfamiliar stimuli aloud, rather than concentrating solely on learning.

5.1.3 Storybook Methodology

Studies on the Production Effect with children so far have used similar methodologies: children are trained on items in isolation, either on a computer or using flash cards. While children complete these studies successfully, they have some limitations. First, it is possible that the encoding of new words is fairly shallow, due to the lack of contextual cues during the word learning task. For example, in both Icht and Mama (2015) and López Assef et al. (2021), items during training were presented in isolation (i.e. image ‘dog’) as bare nouns (i.e. word ‘dog’). The present study addresses some of these limitations by framing the Production Effect (comparing produced and heard words) in a shared storybook design, to investigate whether a more familiar, engaging, and richer learning environment could lead to a Production Effect, where previous studies found a Reverse Production Effect.

Context plays an important role in word learning. When children encounter novel words in meaningful contexts, such as storybooks or everyday conversations, they are able to use contextual clues to infer the meanings of these words (Horst & Samuelson, 2008; Sénéchal et al., 1995), such that words embedded in semantically rich contexts are processed more easily and retained more effectively (Buchana et al., 2001; Siakaluk et al., 2008). Similarly, words embedded in sentence frames rather than in isolation are processed faster than words in isolation (Fernald & Hurtado, 2006). Contextual cues, such as illustrations or narrative cues in storybooks, provide support for learning new vocabulary (Sénéchal et al., 1995). Context not only helps children create initial hypotheses about word meaning but also facilitates the integration of novel words within previous knowledge, making them easier to later retrieve (Collins & Loftus, 1978; Singleton, 2012). While many studies highlight the importance of contextual cues in word learning, research on the effects of context on word learning has shown mixed results. Some studies show better retention of words learned in context (Baleghizadeh & Shahry, 2011), while others report no benefit or even an advantage for learning words without context (Choi et al., 2014; Prince, 1996; Webb, 2007). One possible explanation is that while context aids in the inference of word meaning, it may lead to insufficient processing of the word form itself (Lawson & Hogben, 1996). Thus, although context facilitates the initial understanding of novel words, its effects on learning remain less clear.

Shared storybook reading is a common activity in many Western households (Simcock & DeLoache, 2006), and its influence on vocabulary growth and comprehension has been extensively studied. Shared storybook reading has multiple beneficial factors for children in areas such as literacy skills (Pillinger & Wood, 2014) and parent-child bonding (Schwartz, 2004). Joint attention also benefits from shared storybook reading (Lawson, 2012), which is known to be related to later language development. It also contributes to word comprehension (Justice et al., 2005) and vocabulary development (Elley, 1989; Robbins & Ehri, 1994; Sénéchal & Cornell, 1993). These benefits are argued to stem from the richer linguistic input provided to the child in the context of shared storybook reading. For example, storybooks have been found to contain more unique word types than child-directed speech (Montag et al., 2015). Shared book reading is particularly effective for vocabulary development, with the added benefit of extratextual engagement strategies such as asking questions or making comments (Blewitt & Langan, 2016). These strategies enhance children's attention and engagement, which are crucial for learning novel words.

The cognitive processes involved in word learning during shared storybook reading, such as attention to word form and meaning, are influenced by how words are presented. In particular, novel words can be highlighted using techniques such as dramatic pauses, repeated exposure, and providing definitions or explanations (Read et al., 2019; Elley, 1989). The frequency of word exposure and the use of interactive techniques, such as asking questions about novel words, can further enhance children's learning of these words (Blewitt et al., 2009; Justice, 2002). Some studies have shown positive effects from accompanying the novel words with definitions or explanations to facilitate understanding (Beck & McKeown, 2007; Houston-Price et al., 2014), while others have shown positive effects from simple repetition without additional context (Horst et al., 2011; McLeod & McDade, 2011). For instance, O'Fallon et al. (2020) found that repeated exposure to novel words, even without elaboration, led to successful word learning in preschoolers.

While multiple factors have been proposed to influence word learning during shared storybook reading, such as who reads the book (e.g., Hindman et al., 2008), reading style (e.g., Reese & Cox, 1999), and age of the listener (e.g., Hargrave & Sénéchal, 2000), among others, a meta-analysis of 38 studies done by Flack et al. (2018) found that reading style, tokens and the number of words being learned moderated word learning effects. On the other hand, the child's age, the person who read the story (experimenter or teacher) and the time between training and test did not mediate learning (Flack et al. 2018). This is relevant as our study investigates the Production Effect across different ages (4-and-6-year-olds), thus the task needs to be appropriate for both groups. Shared storybook reading provides a plausible methodology to test whether the Production Effect is found in a more familiar and richer design, as it introduces some of the important factors in our design.

5.1.4 Repeated exposure during word-learning

Another limitation of the tasks used in previous studies is that they were relatively short, with participants typically exposed to items only once. For example, in both Icht and Mama (2015) and López Assef et al. (2021), items were presented only once during training. Although initial word learning can occur through fast mapping, where children quickly link a novel word to its referent after minimal exposure (Carey, 2010), repeated exposures are necessary for creating robust word representations (Horst & Samuelson, 2008; Mather & Plunkett, 2009). Increasing the number of times learners are exposed to a target word can provide the additional exposure needed to create a robust representation (Elley, 1989). Repeated exposure to stimuli facilitates faster and more accurate processing (Jacoby & Dallas, 1981; Scarborough et al., 1979). Multiple exposures not only strengthen memory, but also provides additional opportunities to consolidate information, allowing children to build richer mental representations of words over time (McMurray et al., 2012; Yu & Smith, 2007). For example, Horst et al. (2011) taught 3-year-old children novel words using a shared storybook methodology. Children were either exposed to the same story repeatedly or to a variety of different stories containing the same novel words. Results showed that children who heard the same story multiple times learned more new words than those who listened to a variety of stories

Contextual repetition, such as repeating words within a structured environment like a storybook, also aids in word learning. Repetition increases predictability, reduces cognitive load and allows children to focus their attention on new information (Chun, 2000; Meints et al., 2004). This is particularly beneficial for young children, who tend to thrive in highly structured and predictable learning environments, such as familiar games and repetitive songs (Seuss, 1960). Active participation from learners can further enhance word learning through repetition. Ewers and Brownson (1999) showed that asking children to produce repetitions of novel words during shared reading sessions led to better word retention compared to passive exposure. The authors proposed that this active engagement likely reinforced the connection between the word and its meaning.

As for the Production Effect, in adults, the impact of speech production is influenced by the amount of training. For example, increasing the number of training trials can diminish or even reverse the memory benefit of speech production when learning new words: while speech production can enhance memory in early stages of learning, further practice may lead to a Reverse Production Effect (Kapnoula & Samuel, 2022). Additionally, different types of engagement with the material produce varied effects. Baese-Berk et al., (2024) suggest that simple repetition is not as helpful during the creation of a novel item's representation. However, retrieval practice seems to offer more substantial benefits (Kang et al., 2013; Rose et al., 2023). This study is a starting point to investigate the effect of repeated exposures to items and the Production Effect in children.

The present study aims to answer the following research questions:

1. Does word-learning in a richer, more familiar context lead to a Production Effect rather than Reverse Production Effect?
2. Does increasing the number of exposures to items during training affect learning?
3. Does the effect of context and number of exposures depend on the age of the participants?

These questions are investigated in children aged 4-and 6-years-old. This age group was chosen as both have shown a Reverse Production Effect in a word learning task in a previous study (Zamuner et al., 2018). In our design, the child and experimenter read a book together, resembling what parents do with their children when reading books. The expectation was that this more engaging task would lead to improved word learning compared to previous designs. Specifically, this contrasts with Zamuner et al. (2018) study with children, which used a computer to present stimuli, with nonce animals displayed against a white background and presented as pre-recorded tokens of isolated bare nouns (e.g. 'ged'). Instructions for their training conditions were shown as images on the screen and the experimenter did not interact with participants during learning.

We hypothesized that the Reverse Production Effect observed by Zamuner et al. (2018) is partly attributed to task demands, more specifically, that the methodology and design heightened the overall difficulty of the task. Using a shared storybook reading task could be more engaging and easier to follow for children, possibly leading to more successful learning and creating a more favorable scenario for children to show the Production Effect. Furthermore, we also expected to see a difference between the two age groups, with older children (6-year-olds) being more likely to demonstrate a Production Effect, as the task would likely be easier for them compared to 4-year-olds. Regarding the Number of Exposures, we expected that items presented multiple times during training would have stronger representations and thus perform better than those presented only once.

5.2 Method

5.2.1 Participants

Participants were 40 English-speaking children aged 4 years ($n = 20$, 7 males, 13 females, *Mage* = 53.5 months, *SD* = 4.0, *range* = 47 – 59) and 6 years ($n = 20$, 5 males, 14 females, 1 non-binary, *Mage* = 76.5 months, *SD* = 4.2, *range* = 71 – 84). The target sample size of 40 was determined by conducting a power analysis prior to data collection using PANGEA (v0.2; Westfall, 2016), to obtain 80% power to detect an effect size (d) of 0.4 at the standard .05 alpha error probability. See pre-registration materials for more details.

Participants had a minimum 70% lifetime exposure to English (4-year-olds*:* *M* = 92.8%, *SD* = 5.3*, range* = 84.4 – 100; 6-year-olds*:* *M* = 92.4%, *SD* = 7.1*, range* = 71.5 – 100). All children had learned English from birth, with no more than two consecutive years of 30+% exposure to another language, as estimated from a language background questionnaire completed by parents. All children were also reported to have normal hearing, normal-to-corrected vision, and no history of language impairment. Families were recruited on the museum floor at a Living Lab and children received a sticker for participating. Fifteen additional children were tested but not included in the analyses for having no video or audio for off-line coding (*n* = 2), did not perform the task properly, i.e., produced all words in the *Heard* condition (*n* = 2), errors resulting in no data in a condition (*n* = 3) and did not finish the study (*n* = 8).

5.2.2 Stimuli

There were a total of 16 novel words (*dax, sud, krat, jun, neis, hux, leip, tunk, ret, gop, zel, fos, blick, maip, whis, poud*) paired with nonce images, 8 for each Training Condition (8 *Say*, 8 *Heard*), divided into four lists for counterbalancing for Training Condition (*Say, Heard*) and Number of Exposures (*1 Presentation, 4 Presentations*). The novel words and visual stimuli were taken from the Novel Object and Unusual Name (NOUN) Database (Horst & Hout, 2016). The novel objects from the NOUN database were edited to ensure that the quality and drawing style was consistent with the background images. Some of the novel words were adapted so that each started with a unique consonant. Stimuli were controlled for their phonotactic and neighbourhood density (using the calculator from Storkel & Hoover, 2010). All novel words fell within the mid-high to highest phonotactic probability quartiles and ranged from low to highest neighbourhood density quartiles as reported for the stimuli in Storkel et al. (2013). Novel word stimuli were pre-recorded by a female native speaker of English and normalized for amplitude

(70 dB). Details about phonotactic and neighbourhood density for the items can be found on OSF.

5.2.3 Design

The experiment started with a practice phase to introduce the activity, followed by the experiment phase. The practice phase used pictures of 4 fruits, with 2 items corresponding to each Training Condition. For each Training Condition, 1 of the items was presented once during training, and one was presented 4 times, with a total of 10 practice training trials (5 for each Training Condition). For the experiment phase, which used the novel stimuli, there were 8 items for each Training Condition, with 4 items were presented once during training, and 4 items were presented 4 times during training, with a total of 40 training trials (20 for each Training Condition). Items were presented in a mixed design, with no more than 2 items from the same Training Condition presented consecutively. For items that were repeated during training, the same label did not appear consecutively during testing.

5.2.4 Procedure

Children were tested in a sound-attenuated room at a museum-based lab, and were audio and video recorded using an iPad. The same experimenter collected all of the data and was a native speaker of English. The storybook was designed and printed for this study. The story followed the narrative of an alien who brought gifts from their planet, but lost them while landing on Earth, and needed to collect them. A similar design, using the same stimuli database, was used in Brouillard et al. (2022). Children were informed that for some words in the story they should repeat the word aloud (*Say*), and for the rest they should only listen (*Heard*). The experimenter told children which words to repeat aloud by asking them to do so. To ensure consistency in the presentation of items across Training Conditions, each item was repeated twice by the experimenter: for an item corresponding to the *Say* Training Condition, the experimenter read “We found a *kern*, can you say *kern*?”, whereas for *Heard* items the experimenter said, “We found a *kern*, look at the *kern*”. This meant that participants had three exposures to *Say* items in total (twice from the experimenter, once from themselves producing the label) and two exposures for *Heard* items (twice from the experimenter). If multiple exposures are beneficial, then this contrast in the number of exposures should bias participants to show better recognition to *Say* items compared to *Heard* items.

The pages’ background had illustrations of nature (trees, flowers) that changed throughout to give the sense that the alien was walking across a field and collecting the gifts. Backgrounds were kept simple and did not include a lot of detail so as to not distract from the experiment stimuli. Every 10 trials there was a page with no test items on the illustration, where the alien encouraged the participant by saying “Let’s keep searching!”, “You’re doing so well!”. These pages served as a small break for participants and to direct their attention to the task.

Following training, learning was measured using eye-tracking using the Visual World Paradigm (VWP). Participants were asked to look at the item on the screen corresponding to the label they heard. Eye movements were recorded using an Eyelink 1000 Plus eye-tracker in monocular remote mode (SR Research, Ottawa). On each trial, two novel items (target and distractor) were shown on the screen for 2000 ms in silence, after which participants heard a novel word corresponding to one of the images. Each novel word was used as the target twice, with a total of 32 trials. Distractors corresponded to items from the same Training Condition as the target (i.e., two *Heard*, *4 Presentations* items, one as target, one as distractor). After 2000ms of hearing the target novel word, participants heard “Which one is it?” and had to point at the item on the screen. Pointing was added to the eye-tracking measurement to have an additional overt recognition response.

5.3 Results

5.3.1 Transparency and openness

This study was pre-registered prior to beginning data collection and the pre-registration is available at https://osf.io/swmgz/?view_only=b001a93b12944368ad938e2ca657e438. Within the pre-registration, explanations of hypotheses, design, sample size calculations, and analysis are available for the primary analyses. Note that the pre-registered primary analyses are different from the analyses presented in the paper as the dependent variable for the eye-tracking recognition analyses was changed from the proportion of looking to targets, to the amount of looking to targets to better capture the pattern in Figure 1. Code and data for this study are openly available in the OSF repository, including results from the original pre-registered analyses and description of changes which are provided in the transparency document.

5.3.2 Coding

Coding was done to ensure that children gave the correct response for the Training Condition (e.g., remaining silent during *Heard* items, correctly repeating the word aloud for *Say* items). Children's responses during training were coded off-line by two researchers (the experimenter and first author, both masked to each other's coding). Disagreements in coding were settled by a third coder, who was masked to the first two codings. Trials with items with training errors appearing as target or distractor were excluded from both the eye-tracking recognition and pointing analysis. Items were excluded for the following reasons: producing *Heard* items (n = 14), talking or noises during trials (n = 2), producing an item more than once (n = 7), mouthing words (n = 6), mispronunciations during training (n = 11) and tech error (n = 1).

5.3.3 Recognition task

The time window of analysis started 200ms after word onset to allow time to initiate eye movements and ended after 1000ms. This time window was selected as fixations to the target picture peaked at 1000ms. The data was filtered by participants' accuracy during the pointing task, so that only trials with correct pointing, that thus indicate successful learning, were included in the analysis. Total fixations in milliseconds to the target was the dependent variable for the linear mixed-effects models performed in R (R Core Team) using the lmer() function from the lme4 package (version 1.1-26; Bates et al., 2015). Our fixed effects were Training Condition (*Heard, Say*), Number of Exposures (*1 Presentation, 4 Presentations*) and Age (*4yrs, 6yrs*). List was also included as a control variable to account for potential list effects. As indicated in our pre-registration, we started from the most complex model structure with a random structure containing random by-participant intercepts, by-participant slopes for Training Condition and Number of Exposures and their interaction, random by-item intercepts, and by-item slopes for Training Condition, Age, List and their interactions. Models were simplified to account for convergence issues and singularity errors, until a model converged.

Table 12 provides results from this analysis along with the final model structure, and Figure 12 plots the model predictions. There was a significant three-way interaction between Training Condition, Number of Exposures, and Age (β = -93.6, *SE* = 45.53, *p* = <.05). No other effects were found. Post-hoc comparisons were done on the final model with the emmeans package, using Bonferroni estimations for degrees of freedom (v1.7.4-1; Lenth, 2020). Post-hoc tests for the significant interaction between Training Condition, Number of Exposures, And Age (Table 13) showed a significant difference between in looking to the target in 4-year-olds across Training Conditions for *4 Presentations* items (*Say M* = 181, *SE* = 21.2, *95% CI* [138-223]; *Heard M* = 254, *SE* = 20.8, *95% CI* [213-295]), but not for *1 Presentation* items (*Say M* = 208, *SE* = 20.1, *95% CI* [168-248]; *Heard M* = 217, *SE 22.2*, *95% CI* [172-261]). In this case, 4-year-olds showed a Reverse Production Effect, advantage for *Heard* items, for items that had multiple exposures during training, but no difference in looks between Training Conditions for items that only had one exposure during training.

Table 13 *Model estimating number of fixations to target by Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations) and Age (4yrs, 6yrs).*

Fixed Effects	*F*-value	*df*	*df.Res*	*p*-value
Training Condition	1.02	1	14.23	0.33
Number of Exposures	0.98	1	30.39	0.33
Age	0.08	1	32.91	0.79
List	1.04	3	30.58	0.39
Training Condition*Number of Exposures	2.32	1	546.55	0.13
Training Condition*Age	2.74	1	546.58	0.10
Number of Exposures*Age	1.94	1	31.04	0.17
Training Condition*Number of Exposures*Age	4.13	1	549.32	<0.05

Note. Wald F-tests with Kenward-Roger estimates for df. The final model: had the following syntax specified in the lme4 package: *total_target ~ condition_sum * repetition_sum * age_sum + list + (1 + repetition_sum | participant) + (1 + condition_sum | sound_file)*

Figure 12 *Predictions from model estimating number of fixations to target by Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations), and Age (4yrs, 6yrs), showing the three-way interaction.*

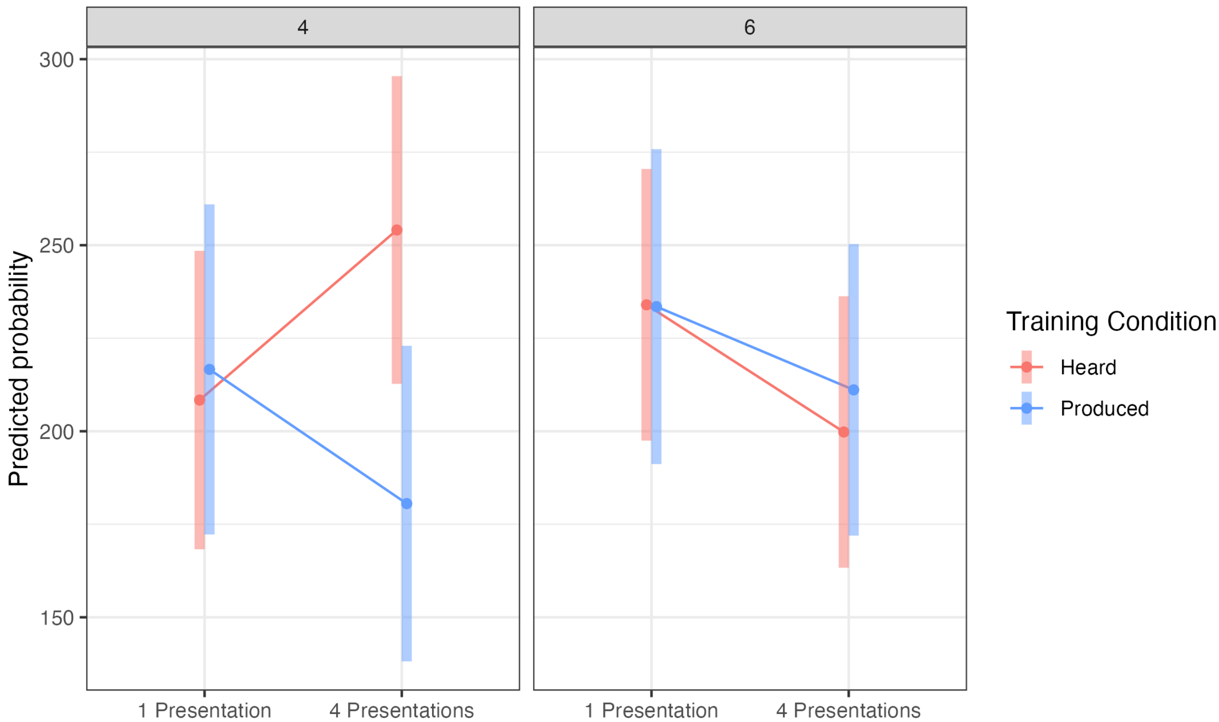


Table 14 *Post-hoc tests of the estimated marginal means for significant three-way interaction from the model predicting Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations) and Age (4yrs, 6yrs).*

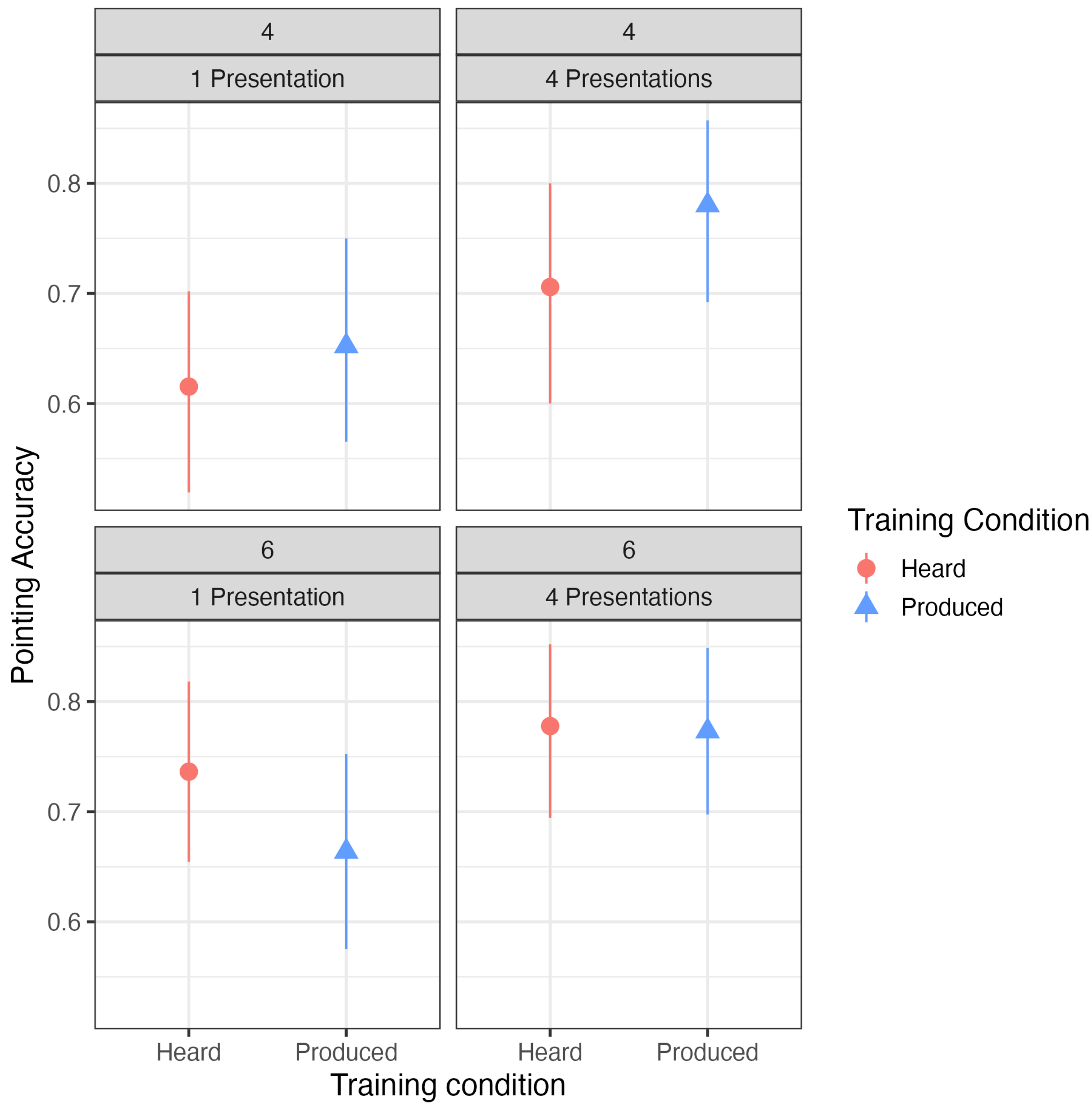
Age	Number of Exposures, Training Condition	Estimate	SE	df	t.Ratio	p-value
4yrs	*4 Presentations*					
	Heard-Say	73.531	25.1	144.9	2.93	<.005
	1 Presentation					
	Heard-Say	-8.24	25.8	152.7	-0.32	0.75
6yrs	*4 Presentations*					
	Heard-Say	-11.326	21.6	83.2	-0.52	0.60
	1 Presentation					
	Heard-Say	0.49	23.3	105.8	0.02	0.99

		M	*SE*	*df*	*95% CI*
4yrs	*4 Presentations*				
	Heard	254	20.8	78.6	[213-295]
	Say	181	21.2	60.5	[138-223]
	1 Presentation				
	Heard	208	20.1	65.3	[167-248]
	Say	217	22.2	69.8	[172-261]
6yrs	*4 Presentations*				
	Heard	200	18.2	53.6	[163-236]
	Say	211	19.5	47.1	[172-250]
	1 Presentation				
	Heard	234	18.2	53.1	[195-270]
	Say	233	21.2	53.1	[191-275]

5.3.4 Pointing task

As mentioned previously, we included a pointing task with eyetracking, in order to also have an overt and explicit measure of recognition. Children were asked to look at an image, and then asked to point to the picture that corresponded to the audio stimuli. Responses were coded as correct (1) or incorrect (0). At times, children changed their responses; however, only the first pointing response was analyzed to maintain consistency. The analyses followed the same procedure as with the eyetracking analyses, except that Pointing Accuracy was the dependent variable, and the analyses were based on both correct and incorrect trials (recall that eye-tracking analyses were done on correct pointing data). As can be seen in Figure 13, pointing accuracy was above change (50% since there were two options) across all ages and variables, showing that children successfully learned the items during training across all the conditions.

Figure 13 *Mean pointing Accuracy By Training Condition (Heard, Say), Number of Exposures (1 Presentation, 4 Presentations) and Age (4yrs, 6yrs).*



Note. Error bars indicate 95% confidence intervals.

Statistical analyses did not find any significant effects from Training Condition (β = -0.12, *SE* = 0.17, *p* = .47), nor from Age (β = -0.21, *SE* = 0.28, *p* = 0.45); however, there was a significant effect of Number of Exposures (β = -0.47, *SE* = 0.22, *p* = <.05). Results from this analysis are shown in Table 15. The same post-hoc comparisons procedure was done as in the previous analysis. Post-hoc tests revealed a higher probability of accuracy in pointing for *1 Presentation* items (*M* = 0.79, *SE* = 0.03, *95% CI* [0.71-0.84]) than for *1 Presentation* items (*M* = 0.70, *SE* = 0.04, *95% CI* [0.60-0.77]). Therefore, pointing accuracy was only predicted by whether items had multiple exposures during training, and not by Training Condition or Age.

Table 15 *Model estimating pointing accuracy by Training Condition (Heard, Say), Number of Exposures (4 Presentations, 1 Presentation) and Age (4yrs, 6yrs).*

Fixed Effects	Estimate	*Std.Err*	*z value*	*p*-value
Training Condition	-0.12	0.17	-0.73	0.47
Number of Exposures	-0.47	0.22	-2.13	<.05
Age	-0.21	0.28	-0.75	0.45
Training Condition* Number of Exposures	0.30	0.34	0.90	0.37
Training Condition*Age	-0.38	0.34	-1.12	0.26
Number of Exposures *Age	-0.18	0.43	-0.42	0.67
Training Condition* Number of Exposures *Age	-0.09	0.68	-0.13	0.90

Note. Wald F-tests with Kenward-Roger estimates for df. The final model: had the following syntax specified in the lme4 package: *Accuracy ~ condition_sum * repetition_sum * age_sum + list + (1 + repetition_sum | Participant) + (1 | sound_file)*

5.4 Discussion

The present study investigated whether the Production Effect would surface with children using an enrichened learning environment of shared storybook learning. Additionally, we manipulated the number of exposures children had to the novel words during training, to determine if increased exposure could also lead to a Production Effect. Results from the eyetracking analyses showed a significant interaction between Training Condition, Number of Exposures and Age, such that 4-year-olds showed a Reversed Production Effect for *4 Presentations* items (better recognition for novel words *Heard* four times during training), but no difference across Training Conditions in the *1 Presentation* condition. No differences in Training Condition or Number of Exposures were found for 6-year-olds. Results from the pointing task showed an effect of Number of Exposures, such that overall, children were more accurate in correctly pointing to target images that had *4 Presentations* during training, compared to items with only one exposure. Our results pattern with previous findings with children which found an attenuation (lack of advantage for any condition) or reversal (advantage for *Heard* items) of the Production Effect (López Assef et al., 2021; 2024; Richtsmeier et al., 2024; Zamuner et al., 2018).

Although the current design embedded the word learning task in a richer and more child-friendly design than previous studies, unexpectedly, we did not observe a Production Effect. While the 6-year-olds showed similar recognition rates across Training Condition, the results from the 4-year-olds patterned in the opposite direction as predicted. Thus, our modified task, with higher familiarity and lower task demands, did not lead to a Production Effect. This suggests that components of our task were still challenging for 4-year-olds. That is, for the younger participants, even in a storybook setting, learning the mapping between the novel word form and referent, while simultaneously producing the novel word was cognitively demanding. This could be due to several factors. Our shared storybook included a narrative and multiple presentations of the novel words, both elements that have been found to aid in word learning in previous storybook studies. However, shared storybook reading usually also involves interactions between the reader and child, in the form of comments from the readers about the story, questions to direct children's attention to elements from the book, and explanations about the novel words/items (Blewitt et al., 2009; Elley, 1989; Justice, 2002; Sénéchal, 1997). These elements were excluded from our design to simplify the task and ensure that all participants received the same instructions and experience during training.

We found support for our prediction that increased exposure leads to better encoding and retention (Elley, 1989). In the eyetracking recognition data, the effect of Number of Exposures was dependent on Training Condition and Age; and in the pointing task, children were more accurate on *4 Presentations* trials, regardless of Training Condition and Age. The difference between eye-tracking and pointing could stem from the stage of retrieval that the tasks are tapping into. Eye-tracking allows us to measure the ongoing recognition of stimuli as participants are activating and retrieving the newly learned words (McDonald et al., 2023), whereas pointing refers to the final step in this process, once the participant has resolved the mapping between the audio stimuli and images on the screen.

Age differences were seen in the eye-tracking recognition task: 4-year-olds showed a Reverse Production Effect for *4 Presentations* items, but no effect for *1 Presentation* items, and 6-year-olds had no differences between Training Conditions. This aligns with the cross-sectional study by López Assef et al. (2021), which showed that the direction of the Production Effect changed across ages: for younger children there was detrimental effect from speech production, which shifted to a Production Effect with older children. Our current data from 4-year-olds also mirrors findings from adults, where increasing the amount of speech production training results in a Reverse Production Effect (Kapnoula & Samuel, 2022). This may stem from how speech production differentially affects word learning at different learning stages (Kapnoula & Samuel, 2022; Leach & Samuel, 2007). At beginning stages (lexical configuration), learners are accumulating information about the novel item. Speech production is beneficial at this stage for adults as the overall difficulty of the task is not high, and speech production increases attention to the word-form, which is likely to have a positive effect on learning (Baese-Berk et al., 2024). After word representations are formed, the next step is lexical engagement which involves establishing connections between the new representation and previous knowledge, a task argued to be more cognitively demanding (Baese-Berk et al., 2024).

The current experiment was not designed to distinguish between lexical configuration and lexical engagement; however, across both age groups, producing items once (*1 Presentation*) and multiple times (*4 Presentations*) had no beneficial effect on children's learning compared to listening. As such, our storybook design did not appear to improve lexical configuration in the *Say* condition. Further work would be needed to determine whether children's difficulty stemmed from lexical configuration and/or lexical engagement. As mentioned above, one element of storybook reading which we did not include in our design, was interactions between the reader and child, in the form of comments/questions. These elements also increase retrieval practice, an element of lexical engagement (Baese-Berk et al., 2024). Moreover, these interactions also increase the timing between the target and response, where longer delays show less disruption from speech production (Baese-Berk & Samuel, 2022; Kapnoula & Samuel, 2023).

Even with the implementation of a storybook training methodology and with increased training exposures, we observed a Reverse Production Effect with young children. One avenue for future research on the Production Effect is to further investigate tasks that promote lexical engagement and/or configuration, to pinpoint the stages of learning where speech production is beneficial and to further understand the underlying processes that are behind the memory advantage for produced items. Furthermore, recent studies indicate that the learning disruption may not be limited to speech production and could stem from other factors. For instance, López Assef et al. (submitted) suggest that engaging in any task during the learning process may negatively impact children's learning outcomes. Future research should also explore factors beyond speech production that may disrupt the learning process. Studies on the Production Effect are not only are interesting at the theory-academia level, for example, on the relation between perception and production, but also, as mentioned in the beginning, is of great interest to different scholars as this could inform theories on memory devices to be used with different populations.

References

Allen, R. J., Hill, L. J. B., Eddy, L. H., & Waterman, A. H. (2020). Exploring the effects of demonstration and enactment in facilitating recall of instructions in working memory. *Memory & Cognition*, *48*(3), 400–410. https://doi.org/10.3758/s13421-019-00978-6

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

Baleghizadeh, S., & Shahry, M. N. N. (2011). The Effect of Three Consecutive Context Sentences on EFL Vocabulary-Learning. *TESL Canada Journal*, *28*(2), 74. https://doi.org/10.18806/tesl.v28i2.1073

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, *107*(3), 251-271. https://doi.org/10.1086/511706

Blewitt, P., & Langan, R. (2016). Learning words during shared book reading: The role of extratextual talk designed to increase child engagement. *Journal of Experimental Child Psychology*, *150*, 404-410. https://doi.org/10.1016/j.jecp.2016.06.009

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, *101*(2), 294. http://dx.doi.org/10.1037/a0013844

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences*, *2*(2), 67-73. https://doi.org/10.1016/S1364-6613(98)01121-8

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://psycnet.apa.org/doi/10.1037/a0028466

Brouillard, M., Dubé, D., & Byers-Heinlein, K. (2022). Reading to bilingual preschoolers: An experimental study of two book formats. *Infant and Child Development*, *31*(3), e2294. https://doi.org/10.1002/icd.2294

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), 13531–13536. https://doi.org/10.1073/pnas.1508631112

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*(3), 531-544. https://doi.org/10.3758/BF03196189

Carey, S. (2010). Beyond fast mapping. *Language Learning And Development*, *6*(3), 184-205. https://doi.org/10.1080/15475441.2010.484379

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

Choi, S., Kim, J., & Ryu, K. (2014). Effects of context on implicit and explicit lexical knowledge: An event-related potential study. *Neuropsychologia*, *63*, 226–234. https://doi.org/10.1016/j.neuropsychologia.2014.09.003

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407. https://doi.org/10.1037/0033-295X.82.6.407

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

Curtin, S., Byers-Heinlein, K., & Werker, J.F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*, 492–504. https://doi.org/10.1016/j.wocn.2010.12.002

DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, *34*(4), 590–601. https://doi.org/10.1016/j.infbeh.2011.06.005

DePaolis, R. A., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*(4), 642–649. https://doi.org/10.1016/j.infbeh.2013.06.007

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24,* 174–187. https://doi.org/10.2307/747863

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental science*, *9*(3), F33-F40. https://doi.org/10.1111/j.1467-7687.2006.00482.x

Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology*, *54*(7), 1334-1346. https://doi.org/10.1037/dev0000512

Forrin, N. D., & MacLeod, C. M. (2016). Auditory presentation at test does not diminish the production effect in recognition. *Canadian Journal of Experimental Psychology*, *70*(2), 116–124. https://doi.org/10.1037/cep0000092

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

Grohe, A. K., & Weber, A. (2018). Memory advantage for produced words and familiar native accents. *Journal of Cognitive Psychology, 30*(5-6), 570-587. https://doi.org/10.1080/20445911.2018.1499659

Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*(1)*,* 75–90. https://doi.org/10.1016/S0885-2006(99)00038-1

Hindman, A. H., Connor, C. M., Jewkes, A. M., & Morrison, F. J. (2008). Untangling the effects of shared book reading: Multiple factors and their associations with preschool literacy outcomes. *Early Childhood Research Quarterly, 23*(3)*,* 330–350. https://doi.org/10.1016/j.ecresq.2008.01.005

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393-1409. https://doi.org/10.3758/s13428-015-0647-3

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*(2), 128-157. https://doi.org/10.1080/15250000701795598

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. Frontiers in Psychology, 2, 17. https://doi.org/10.3389/ fpsyg.2011.00017

Houston-Price, C., Howe, J. A., & Lintern, N. J. (2014). Once upon a time, there was a fabulous funambulist…: What children learn about the "high-level" vocabulary they encounter while listening to stories. *Frontiers in psychology*, *5*, 75. https://doi.org/10.3389/fpsyg.2014.00075

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. (2019). The effect of vocal production on vocabulary learning in a second language. *Language Teaching Research, 26*(1), 79-98. 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

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*(3), 306. https://doi.org/10.1037//0096-3445.110.3.306

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://psycnet.apa.org/doi/10.1037/a0033337

Justice, L. M. (2002). Word exposure conditions and preschoolers' novel word learning during shared storybook reading. *Reading Psychology*, *23*(2), 87–106. https://doi.org/10.1080/027027102760351016

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1)*,* 17–32. https://doi.org/10.1044/0161-1461(2005/003)

Kang, S. H., Gollan, T. H., & Pashler, H. (2013). Don't just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin & Review*, *20*, 1259-1265. https://doi.org/10.3758/s13423-013-0450-z

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

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*(28),1-14 . https://doi.org/10.1186/s40359-017-0198-8

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning*, *46*, 101–135. https://doi.org/ 10.1111/j.1467-1770.1996.tb00642.x

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

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., McDonald, M., Bernard, A., & Zamuner, T. S. (2024). Effects of speech production training on memory across short and long delays in 5-and 6-year-olds: A pre-registered study. *Applied Psycholinguistics*, 1-24. https://doi.org/10.1017/S0142716424000109

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). https://doi.org/10.16995/glossa.8582

López Assef, B. & Zamuner, T. S. (Accepted). Task effects in children's word recall: Expanding the reverse production effect. *Journal of Child Language.*

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The Production Effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

McLeod, A. N., & McDade, H. L. (2011). Preschoolers' incidental learning of novel words during storybook reading. *Communication Disorders Quarterly*, *32*(4), 256-266. https://doi.org/10.1177/1525740109354777

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), 1-16. https://doi.org/10.1080/09658211.2014.986135

Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, *14*(1), 60-76. https://doi.org/10.1080/15250000802569702

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631-631. https://doi.org/10.1126/science.1144073

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119,* 831–877. https://doi.org/10.1037/a0029872

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science, 26,* 1489 –1496. https://doi.org/10.1177/0956797615594361

O'Fallon, M., Von Holzen, K., & Newman, R. S. (2020). Preschoolers' word-learning during storybook reading interactions: Comparing repeated and elaborated input. *Journal of Speech, Language, and Hearing Research*, *63*(3), 814-826. https://doi.org/10.1044/2019_JSLHR-19-00189

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., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory, 20*(7), 717-727. https://doi.org/10.1080/09658211.2012.699070

Pillinger, C., & Wood, C. (2014). Pilot study evaluating the impact of dialogic reading and shared reading at transition to primary school: Early literacy skills and parental attitudes. *Literacy, 48,* 155–163. https://doi.org/10.1111/lit.12018

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, *80*, 478–493. https://doi.org/10.1111/j.1540-4781.1996.tb05468.x

Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2019). 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

Reese, E., & Cox, A. (1999). Quality of adult book reading affects children's emergent literacy. *Developmental Psychology, 35,* 20–28. https://doi.org/10.1037//0012-1649.35.1.20

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86,* 54–64. https://doi.org/10.1037/0022-0663.86.1.54

Rose, M. C., Brooks, P. J., Lodhi, A. K., & Cortez, A. (2024). Benefits of testing and production for learning Turkish as a new language. *Language Learning*, *74*(2), 365-401. https://doi.org/10.1111/lang.12602

Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory & Cognition*, *7*(1), 3-12. https://doi.org/10.3758/BF03196928

Schwartz, J. I. (2004). An observational study of mother/child and father/ child interactions in story reading. *Journal of Research in Childhood Education, 19,* 105–114. https://doi.org/10.1080/02568540409595058

Sénéchal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28*(4)*,* 360–374. https://doi.org/10.2307/747933

Sénéchal, M., Thomas, E., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal Of Educational Psychology*, *87*(2), 218. https://doi.org/10.1037/0022-0663.87.2.218

Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., & Owen, W. J. (2008). The benefits of sensorimotor knowledge: Body–object interaction facilitates semantic processing. *Cognitive Science*, *32*(3), 591-605. https://doi.org/10.1080/03640210802035399

Simcock, G., & DeLoache, J. (2006). Get the picture? The effects of iconicity on toddlers' reenactment from picture books. *Developmental Psychology*, *42*(6), 1352-1357. https://doi.org/10.1037/0012-1649.42.6.1352

Singleton N. C. (2012). Can semantic enrichment lead to naming in a word extension task?. *American Journal of Speech-Language Pathology*, *21*(4), 279–292. https://doi.org/10.1044/1058-0360(2012/11-0019)

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, *42*(2), 497–506. https://doi.org/10.3758/BRM.42.2.497

Storkel, H. L. (2013). A corpus of consonant–vowel–consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods*, *45*(4), 1159-1167. https://doi.org/10.3758/s13428-012-0309-7

Vihman, M.M., DePaolis, R.A., & Keren-Portnoy, T. (2014). The role of pro- duction in infant word learning. *Language Learning*, *64*, 121–140. https://doi.org/10.1111/lang.12058

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258-263. https://doi.org/10.1016/j.tics.2009.03.006

Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, *11*, 63–81. https://doi.org/10.1177/1362168806072463

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289-1309. https://doi.org/10.1037/0012-1649.34.6.1289

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development*, *1*(2), 197-234. https://doi.org/10.1080/15475441.2005.9684216

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *24*(5), 603–612. https://doi.org/10.1177/0956797612458802

Zamuner, T. S., Fais, L., & Werker, J. F. (2014). Infants track word forms in early word–object associations. *Developmental Science*, *17*(4), 481-491. https://doi.org/10.1111/desc.12149

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. (2018). 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

Zamuner, T. S., Yeung, H. H., & Ducos, M. (2017). The many facets of speech production and its complex effects on phonological processing. *British Journal of Psychology*, *108*(1), 37–39. https://doi.org/10.1111/bjop.12220

6 General Discussion and Future Directions

This general overview summarizes the findings of this dissertation and how the discoveries contribute to a better understanding of the Production Effect. This chapter will also present a model for framing outcomes with children and will discuss future directions for research on the Production Effect.

6.1 Summary of the Findings

Chapter 2 examined the Production Effect in children aged 2–6 years with familiar words. It found an exciting new pattern: a developmental shift in the Production Effect. Younger children recalled more *Heard* words, whereas older children showed the typical Production Effect, recalling more words they *Produced*. This suggests that age and cognitive factors influence how speech production influences memory across development. Results from Chapter 2 with older children align with other studies that show a Production Effect with children (Icht & Mama, 2015; Chapter 3 - López Assef et al., 2023; Pritchard et al., 2019).

Chapter 3 presented another new finding: while children aged 5–6 years initially benefited from producing known words, this advantage disappeared after a one-week delay, indicating that the Production Effect in children may be less durable over time. This finding contrasts with adult studies, where the Production Effect often persists even after delays (e.g., Forrin & MacLeod, 2016). The lack of sustained benefit in children might reflect their still-developing cognitive and linguistic systems, which affect how well they encode and retrieve information over time.

Chapter 4 explored how different types of actions during word learning impacted recall and recognition. Children aged 5–6 years were trained on novel words using different tasks: Listening compared to Verbal-Speech (production), Non-Verbal-Speech (sticking out tongue), and Non-Verbal-Non-Speech (touching nose). While all groups successfully learned the words, the new addition to our understanding of The Production Effect is the demonstration that non-verbal tasks also disrupted learning. These disruptions were likely due to divided attention and task switching, which are especially demanding for children whose long-term memory, working memory, cognitive control, and attentional capacities develop throughout childhood (though note that in some instances, young children's learning sometimes outperforms adults, Gualtieri & Finn, 2022).

Chapter 5 was the last study, which extended our knowledge of the Production Effect by expanding the methodology and testing the effects of multiple exposures. No Production Effect was found in the more familiar and richer shared storybook setting. Instead, 4-year-olds showed better recognition for *Heard* words when they were repeated multiple times. This aligns with findings by Zamuner et al. (2018), who also observed a Reverse Production Effect in children aged 4-6 years, and who suggested that producing novel words while learning may be too cognitively demanding, especially as adults on the same methodology showed a Production Effect.

6.2 Modelling the Production Effect

In the studies presented in this dissertation, characteristics of the learner in the form of cognitive, linguistic and developmental factors have been argued to play a critical role in shaping the effectiveness of production. Specifically, the Reverse Production Effect has been argued to stem from the demands of coordinating speech production while learning novel words, a task that requires linguistic and cognitive abilities, which are developing throughout childhood.

The initial phase of learning a new word typically involves mapping the auditory word form to its referent (e.g., Carey & Bartlett, 1978). Research shows that even brief exposure allows individuals to establish and link the phonological and semantic representations of a newly learned word (Kan & Kohnert, 2008, 2012). This exposure can provide different types of information that aid in encoding the word into memory. For instance, hearing a word enables us to store auditory information, while reading a word aloud allows us not only to encode its production-related properties but also to retain its visual-orthographic features. This multi-level encoding process (gathering information through multiple sensory modalities during learning) is known as multisensory enrichment (Andrä et al., 2020; Mayer et al., 2015; Perez Repetto et al., 2017). Both pedagogical and cognitive studies show that multisensory input enhances learning outcomes (Shams & Seitz, 2008; von Kriegstein & Giraud, 2006). For example, pairing gestures with heard words improves later recognition (Mayer et al., 2015). The act of speech production offers a richer multisensory experience compared to other conditions such as listening or reading alone. This is why the Production Effect—a memory advantage for *Produced* items—fits within the framework of multisensory enrichment, as it involves multiple levels of encoding.

As a word is learned or memorized, its representation can deteriorate without continued exposure or practice (Gathercole, 2006; Kan et al., 2014; Vlach & Sandhofer, 2012). Rehearsal, however, can help maintain and strengthen the word's representation (Gathercole, 2006; Kaushanskaya & Yoo, 2011; Thorn & Gathercole, 2001). Since the Production Effect relates to memory, it is important to consider its role within the broader memory literature. For this, I will appeal to the Baddeley-Hitch Model (Baddeley & Hitch, 1974; Baddeley, 1986, 2002), a prominent model of working memory that has been applied to language learning and the Production Effect (e.g., Kaushanskaya & Yoo, 2011).

Working memory, as described by Baddeley and Hitch (1974), is the system responsible for short-term memory retention, allowing information to be available temporarily, as opposed to long-term memory, which stores information indefinitely. For example, memorizing a phone number relies on working memory. The Baddeley-Hitch model outlines three components of working memory: the central executive, the visuospatial sketchpad, and the phonological loop. The central executive oversees and manages the information in working memory, allocating cognitive resources. This management is critical because our ability to multitask is limited, and increased task demands reduces processing efficiency. The visuospatial sketchpad handles short-term storage of visual and spatial information, while the phonological loop manages verbally and auditory encoded information, with the phonological store retaining phonological representations for short periods and the articulatory rehearsal system aiding in rehearsal and storage.

The phonological loop is crucial for language learning, specifically word learning, as it maintains unfamiliar sounds in working memory while long-term representations are constructed (Baddeley et al., 1998). If there is a lack of engagement with what is being stored in short term memory, the information is likely to be impacted negatively, leading to decay from memory (Camos et al., 2009; Camos & Barrouillet, 2014; Lewandowsky & Oberauer, 2015). Developed working memory skills, especially related to the phonological loop, are associated with better word learning (Kohonen & Service, 1995; Michas & Henry, 1994; Papagno & Vallar, 1995; Service, 1992). As mentioned previously, both adults and children have limited cognitive resources, restricting the number of tasks they can perform simultaneously. In multitasking situations, cognitive strain forces us to prioritize tasks (Courage et al., 2015). This strain is even greater for infants and preschoolers, who typically struggle with multitasking (Courage et al., 2015), although this capacity improves as children develop. By the age of 4, children can perform tasks requiring inhibition, and by age 6, their working memory is developed enough to handle more complex tasks (Best & Miller, 2010). As children age, this cognitive strain lessens as their cognitive and language skills mature.

The development of cognitive control is particularly relevant when discussing the Production Effect in children. Cognitive abilities improve with age, leading to adult-like performance over time. For example, Davidson et al. (2006) studied memory, inhibition, and task-switching in children aged 4-to-13 years and young adults. Their results showed that younger children could perform tasks as long as difficulty was minimized. However, inhibition and task-switching demands affected younger children more strongly, and even the oldest children had lower performance on task-switching, indicating that adult-like performance develops later. Developmental changes in cognitive control likely contributes to differences in the Production Effect across age groups and tasks. Under certain conditions, children's limited cognitive resources, and their still-developing working memory can make it more difficult for them to process the additional encoding layers provided by producing words aloud. This explains why, in some cases, younger children experience a Reverse Production Effect: the cognitive load of producing a word may interfere with their ability to encode and later recall words. In contrast, older children and adults might be more likely to benefit from producing words because their more mature cognitive systems can handle the task's demands, making the *Produced* words more distinctive and easier to remember.

To date, the Production Effect has been primarily explained through distinctiveness: reading a word aloud creates additional encoding layers (e.g., articulatory and auditory cues) compared to silent reading (MacLeod et al., 2010). These additional cues are thought to make the memory trace of the *Produced* item more distinct, facilitating its retrieval during recall or recognition. However, the results from this dissertation challenge the universality of this approach, especially in children (also see Baes-Berk et al for a discussion of adult findings). While the Production Effect is robust in adults, children exhibit more variable outcomes. Table

16 provides an overview of the studies done on The Production Effect with children. Training Condition labels have been adjusted for consistency across studies, such that *Produced* refers to training conditions in which participants heard the word and then repeated it aloud, and *Heard*, training conditions in which participants only listened to the word. A number of patterns and observations can be made.

Table 16. ***Summary of studies with children on the Production Effect (PE) and Reverse Production Effect (RPE)***

Authors	Training conditions	Stimuli	Age	Test	**PE** /*RPE*
López Assef et al., 2021 (Ch. 2)	Look, Hear, Produced	Familiar words + images	2 yrs	Free recall after each block	*RPE*
López Assef et al., 2021 (Ch. 2)	Look, Hear, Produced	Familiar words + images	3-6 yrs	Free recall	*RPE* shift to **PE**
Richtsmeier et al. (2024)[c]	Heard, Produced With and without Semantic Cues	Novel words + nonce aliens	3-4 yrs	3-AFC	No effect
				Naming task	No effect
Zamuner et al., 2018	Heard, Produced, Mixed Design	Novel words + nonce animals	4-6 yrs	Recognition (eyetracking)	*RPE*
Zamuner et al., 2018	Heard, Produced, Block design	Novel words + nonce animals	4-6 yrs	Recognition (eyetracking)	*RPE*
López Assef et al. (Chapter 5)	Storybook Heard, Produced 4 Presentations, 1 Presentation	Novel words + nonce images	4 & 6 yrs	Recognition (eye-tracking)	*RPE* (4yrs, 4 Presentations)
				2-AFC	No effect
Icht & Mama, 2015	Look, Heard, Produced	Familiar words + images	5 yrs	Free recall	**PE**
Icht & Mama, 2015	Heard, Produced	Unfamiliar words + images	5 yrs	4-AFC	**PE**
López Assef et al., 2023 (Ch. 3)[b]	Heard, Produced Immediate vs. Delayed Testing	Familiar words + images	5-6 yrs	Free recall	**PE**
				Free recall Delayed	No effect
				Old/New Recognition Delayed	No effect
López Assef et al. (Ch. 4)	Verbal/Speech (Produced)-Heard	Novel words + nonce animals	5-6 yrs	Recognition (eyetracking)	No effect
				Free recall	No effect
López Assef et al. (Ch. 4)	Verbal/Non-Speech-Heard	Novel words + nonce animals	5-6 yrs	Recognition (eyetracking)	No effect
				Free recall	*RPE*
López Assef et al. (Ch. 4)	Non-Verbal/Non-Speech-Heard	Novel words + nonce animals	5-6 yrs	Recognition (eyetracking)	No effect
				Free recall	*RPE*
Junttila & Ylinen, 2020[a]	Intentional task with (Produced) and without production (Heard), Finnish monolinguals	English words, pseudowords + animal images	5-8 yrs	Word form test	No effect
				Word meaning test	No effect
Pritchard et al., 2019	Read-Silent, Read-Aloud	Familiar written words.	7-10 yrs	Old/New Recognition	**PE**
Pritchard et al., 2019	Read-Silent, Read-Aloud	Novel written words	7-10 yrs	Old/New Recognition	**PE**

[a] Included Digit Span (memory), NEPSY-II Memory for Names (recall) NEPSY-II Phonological Processing (phonological skills), WISC-IV Block Design (perceptual reasoning skills) measures. None of the measures were found to predict performance on PE related conditions.
[b] Included a vocabulary size measure (EOWPVT-4), which did not predict the PE.
[c] Included Test of Early Grammatical Impairment (TEGI) language screener, Primary Test of Non-verbal Intelligence (PTONI), CUBED Narrative Language Screener, name writing ability (literacy skill development) and flanker task (inhibition) scores, and a non-word repetition task (phonological short-term memory) measure. No relation to PE found.

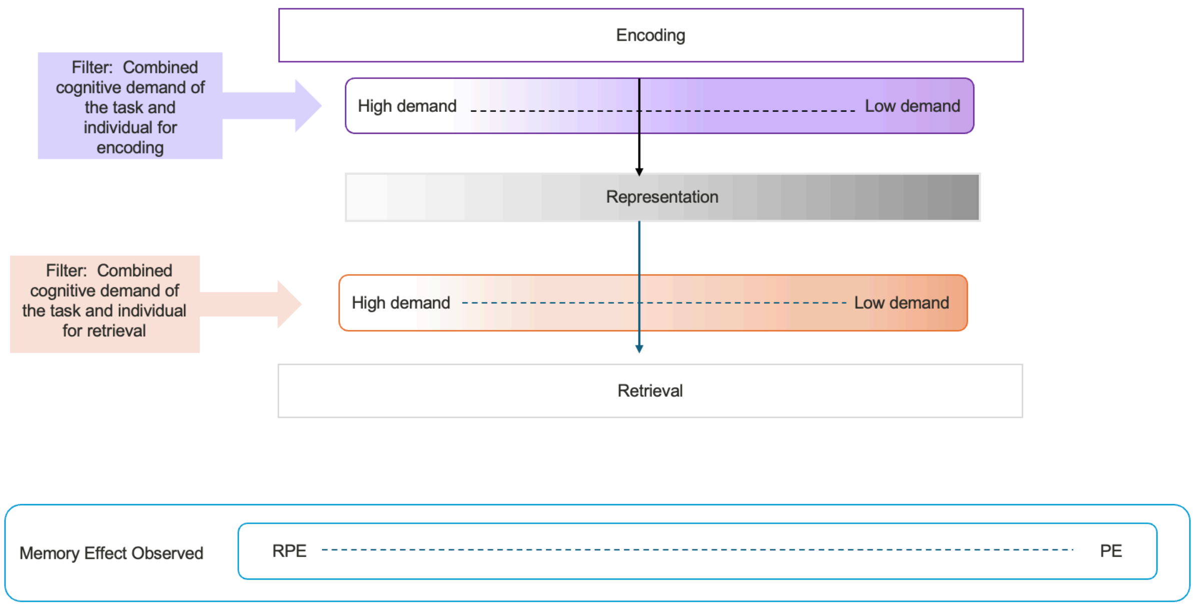
First, while adults typically show a Production Effect with familiar stimuli, young children also show a reversal effect. Chapter 2 - López Assef et al. (2021) found a developmental change: a Reverse Production Effect seen with younger children using familiar stimuli, shifting to a Production Effect with older children that is observed across a number of studies (Icht & Mama, 2015; Chapter 3 - López Assef et al. 2023; Pritchard et al., 2019). However, this distinctive encoding is still not as robust as the one created by adults: after a 1-week delay, the Production Effect with children disappeared (Chapter 3 - López Assef et al. 2023). In word learning studies, children tend to show a Reverse Production Effect, with better learning for items *Heard* during training, though this also depends on how learning is assessed. This suggests that listening is less cognitive demanding during encoding, leading to better memory performance.

In Table 16, we can also see that as children get older, the range of stimuli in which they show a Production Effect increases. For example, while 2–4-year-olds do not show a Production Effect with familiar words (Chapter 2 - López Assef et al., 2021) and novel words (Richtsmeier et al., 2024), studies with older children show a Production Effect with familiar words (Icht & Mama, 2015; Chapter 2 - 3 - López Assef et al., 2021, 2023), unfamiliar words (Icht & Mama, 2015), and also in a reading study with familiar and novel words (Pritchard et al., 2019, though note that their conditions were *Read-Aloud* and *Read-Silent*, the latter which maps more closely to the *Look* condition from Icht & Mama, 2015 and Chapter 2 - López Assef et al., 2021).

Lastly, another observation that can be made from Table 16 is that very few studies have been done with children under 5 years of age and with children over 6 years of age. Further research is needed with younger age groups to establish whether varying tasks and methodologies would replicate previous results. Additionally, studies targeting older age groups are essential for understanding when and how the developmental shift to adult-like-processing takes place.

I propose a model (shown in Figure 14) that aims to capture the influence from speech production on encoding and retrieval as shaped by the combined effects of the cognitive demand imposed by the task and the characteristics of the individual.

Figure 14. *Framework for the Production Effect.*



The first level of the model, depicted in purple, represents the encoding stage. At this stage, participants are actively processing and encoding information about the items being learned (for novel items) or memorized (for familiar items). The robustness encoding depends on cognitive demands during the encoding process, represented as a “filter” between encoding and the resulting memory representation in Figure 14. This filter is represented as a continuum from lower to higher cognitive demands, which reflects the combined cognitive load triggered by the task and trigger by individual factors. Tasks that imply a lower cognitive demand allow participants to process and then encode more information, resulting in stronger memory representations, shown by a darker color in Figure 14 . On the other hand, high-demand tasks require participants to allocate more cognitive resources to meet the tasks' demands, leaving fewer resources for processing and encoding, which subsequently leads to weaker representations. This is represented by a lighter color. For example, cognitive difficulty varies depending on whether the task involves memorizing known words or learning new ones. If the stimuli consist of novel words, which need to be repeated aloud, this is a multi-tasking situation, with a high cognitive load. Participants have to learn and establish new label-object associations, while they also have to rehearse aloud and maintain the item for later retrieval. On the other hand, if the stimuli are highly familiar words, participants can focus solely on rehearsing and maintaining the item in memory, which makes completing the task easier than in the novel word scenario.

The proposed filter not only includes task related effects, but also incorporates individual factors that influence the task demands, such as age, language background, and executive functioning skills. For example, the familiarity of a word may vary based on a participant's age and language background, which can subsequently affect the robustness of the resulting representation. A word like *padel* might be less familiar to a 2-year-old, making the task more similar to a novel word learning task, which would result in less robust encoding. In contrast, an older child familiar with sports would know the word *padel*, which would reduce the cognitive demands of the task and allow them to a create robust encoding, increasing the likelihood of a Production Effect. For older children, as in Pritchard et al. (2019)'s study with 7–to-10-year-olds, tasks involving reading novel words demonstrated a Production Effect. Thus, even though the hypothesis is that tasks with novel words are more demanding than tasks with familiar words, a novel word reading task is not necessarily excessively challenging with older children. Thus, novel word learning tasks fall along the continuum of cognitive demand but are not positioned at either extreme.

The outcome of the encoding stage is the representation of the word. As mentioned before, in order for the Production Effect to emerge participants have to create distinctive representations of the words, than include additional levels of information compared to other items. This lexical representation is shown in gray in Figure 14, on one end the darker gray suggesting a more robust (complete) representation, whereas the other end (in white) suggests a weaker representation as a result of the encoding stage. It is important to mention that participants can create robust representations for both *Produced* and *Heard* items. In a low cognitive demand task, such as memorizing familiar words, participants create robust representations for both training conditions, with *Produced* representations also being distinctive, and thus leading to a Production Effect. In more demanding tasks, such as a novel word task for children, *Produced* items involve another level of multi-tasking: not only do they have to learn and rehearse new material, but they have to do this while also producing the word aloud. While participants can create representations for these items, the overall cognitive demand of the task does not allow them to create representations as robust, and distinctive, as *Produced* items in a familiar word task. *Heard* items lack this additional level of multi-tasking, resulting in similar strength-representations for *Produced* and *Heard* items, or an advantage for *Heard* items.

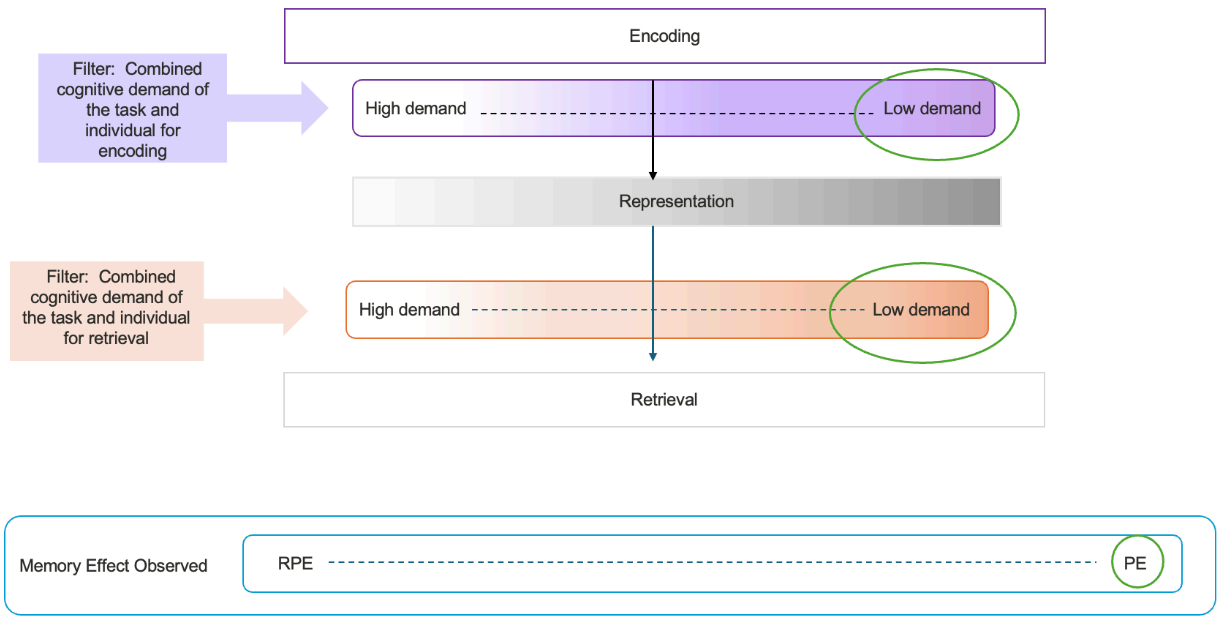
The next stage in the model is related to activating and retrieving the items to evaluate memory or learning. The retrieval stage, represented in orange, parallels the components of the encoding stage. The ease of retrieval is also influenced by the cognitive demands of the task and individual factors. Note that there is a distinction being made here between the cognitive demand of the training phase, and the cognitive demand of the testing/assessment phase. First, we have the effect of the cognitive demand from the training phase on retrieval: for less demanding encoding tasks, participants can allocate more cognitive resources on creating and then maintaining distinctive representations for Produced items, resulting in easier retrieval and demonstrating a strong Production Effect. In contrast, more demanding tasks may result in either an attenuation of the effect (no advantage for any condition) or a Reverse Production Effect, as the resulting representation from the encoding stage is less robust, and thus activation and retrieval are less likely. In addition to this, another factor needs to be considered: the type of retrieval task. This factor is hypothesized to influence the likelihood of observing the Production Effect. More challenging retrieval tasks, which place higher cognitive demands on participants, may be less sensitive to detecting the Production Effect, as such tasks can limit participants' ability to access the distinctive representations formed during encoding. For example, free recall, which requires participants to retrieve information without external cues, is a high-demand task that may fail to detect the Production Effect due to the greater cognitive load it imposes. In contrast, tasks like eye-tracking recognition, which provide visual cues (e.g., target images) and auditory cues (e.g., target sounds), reducing cognitive demands, and thus more likely to reveal a Production Effect.

A key point emphasized in this dissertation, supported by recent research (Baese-Berk et al., 2024), is that methodologies targeting different stages of processing and learning reveal varying impacts of speech production. In some cases, differences between *Produced* and *Heard* items may arise during tasks that measure word activation rather than final lexical retrieval. For instance, recognition accuracy might be comparable across conditions, but one training condition may exhibit faster or stronger activation. This is illustrated in Chapter 5 (López Assef et al., submitted), where eyetracking recognition (an online task) and pointing accuracy (an offline task) yielded differing results. These findings align with prior studies showing that the effects of speech production on word learning vary depending on the learning stage (Baese-Berk et al., 2024; Kapnoula & Samuel, 2022; Leach & Samuel, 2007). Specifically, assessments focusing on the earlier stages of learning, such as lexical configuration, may yield different outcomes than those targeting later stages, such as lexical engagement. This distinction highlights the dynamic nature of the Production Effect across different phases of memory and learning.

The output of the model, the memory effect observed, addresses the likelihood of observing a Production Effect considering the filters in the previous levels. This likelihood operates along a continuum, ranging from a Reverse Production Effect to a Production Effect. Previous studies have shown cases where the effect is diminished (with no advantage for either group), suggesting that the Production Effect is not binary.

Figure 15 below represents the classic Production Effect with a 5-year-old child memorizing familiar words. In this scenario, there are no factors that imply a higher cognitive demand given that the linguistic stimuli are familiar, and the task is relatively easy given the child's age. Therefore, speech production during training, and creating robust, distinctive representations is completed without difficulty. The mapping from stimuli to word representation, rehearsal, and maintenance of words in memory for later retrieval is successfully performed, allowing the distinctiveness for *Produced* items to benefit children's memory retrieval, leading to the Production Effect. This scenario applies to child studies such as Icht and Mama (2015), Chapter 2 - López Assef et al. (2021), Chapter 3 - López Assef et al. (2023) and Pritchard et al. (2019).

Figure 15 *Representation of Production Effect for a child memorizing familiar words*

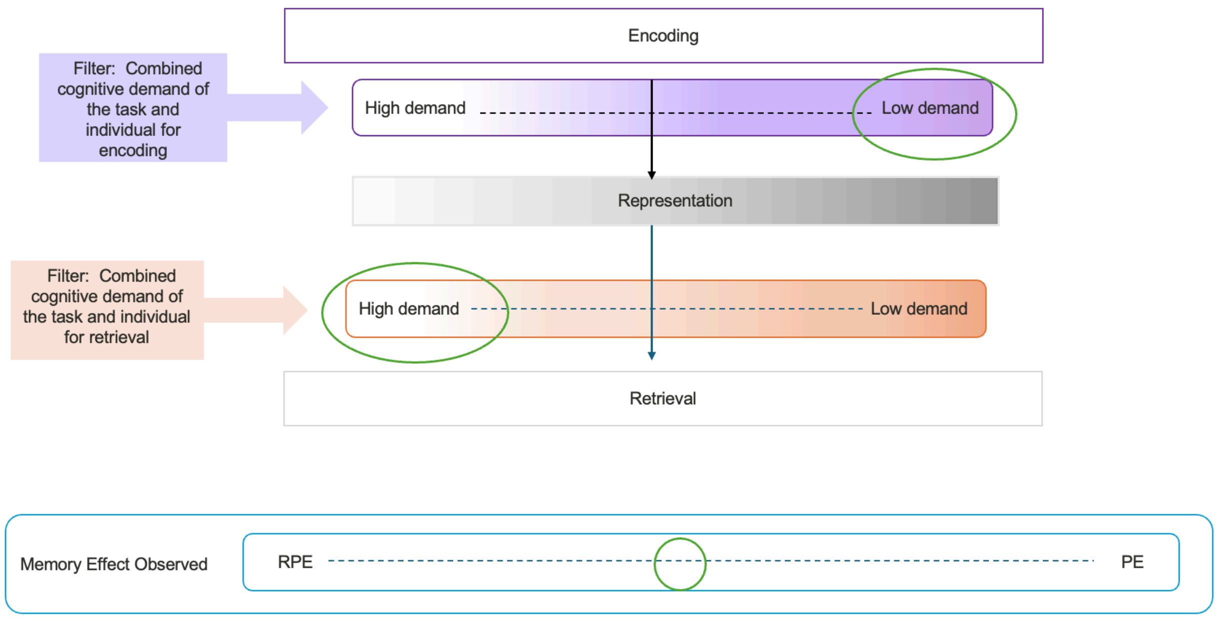


Note. The representation of this scenario is circled in green. In this scenario, the demands for the task are low; therefore, the learner is able to create distinctive encodings for *Produced* items, and access them during retrieval, resulting in a Production Effect.

Figure 16 below illustrates an attenuation of the Production Effect, where no advantage is observed for any Training Condition. The encoding stage remains consistent with that depicted in Figure 15, with distinctive representations successfully created by the participant. However, the key difference lies in the cognitive demands of the retrieval task. While the child was able to form a distinctive representation during encoding, the cognitive demand from the retrieval task does not allow the participant to tap into the distinctive representation. As a result, no memory advantage emerges for any Training Condition. This scenario does not lead to a Reverse Production Effect, as items from both Training Conditions are activated and retrieved to a similar extent.

A reversal or attenuation of the Production Effect arises when these combined factors result in higher cognitive demands. Under such conditions, fewer cognitive resources are available for encoding, leading to less stable representations for retrieval. In a word-learning task, such as the one in Zamuner et al. (2018), 4- to 6-year-old children were required to map a novel auditory word form to a visual referent while simultaneously articulating the word form (saying the novel word aloud). This process relies heavily on the phonological loop, which is essential for word learning as it temporarily maintains unfamiliar sounds in working memory while long-term representations are constructed (Baddeley et al., 1998). The articulatory rehearsal system supports this process by keeping materials active in memory. However, young children, with their still-developing working memory systems, have fewer cognitive resources available for processing. The additional cognitive load imposed by the concurrent production task during learning reduces the resources available for the phonological loop, thereby limiting the ability to maintain and rehearse information. This increased demand hinders the encoding of *Produced* items, preventing the formation of distinctive representations. Consequently, participants exhibit a memory advantage for *Heard* items, as the potential distinctiveness of *Produced* items is not achieved.

Figure 16 *Representation of attenuation or reversal of Production Effect for a child memorizing familiar words*



Note. The representation of this scenario is circled in green. While participants are able to create a distinctive representation, the high cognitive demand from the retrieval task does not allow the participant to tap into the distinctive representation, resulting in an attenuation of the effect and no advantage for any Training Condition.

6.1 Limitations of the current research

While the research presented in this dissertation provide important insights into the Production Effect in children, it is important to acknowledge the limitations. First, all the studies were conducted with English-speaking monolingual children, limiting the applicability of the findings to monolingual populations. This is important because language experience has been suggested to influence the Production Effect; thus, it is possible that bilingual children show different patterns to monolingual children. Second, only one study (Chapter 3 - López Assef et al., 2023) studied delayed testing with children, and this study was conducted online during the COVID-19 pandemic, which may have introduced external factors affecting children's performance. More research is needed to better understand the Production Effect and delayed testing in children. Third, across multiple studies (Chapters 2, 3, and 4 – López Assef et al., 2021, 2023, submitted), close to floor effects were observed in recall tasks with children, potentially affecting the detection of the Production Effect. Additional studies using different memory measurement tasks are recommended to minimize the risk of floor effects. Lastly, as discussed in Chapter 5 – López Assef et al. (in prep), the storybook methodology used may not have incorporated the critical features needed to boost children's learning of the novel words and reduce the task difficulty. Further studies are encouraged to help determine whether or to which extent the Production Effect is affected by the familiarity of the task in children.

6.1 Implications and Future Directions

Studies on the Production Effect and Reverse Production Effect with children are relatively new, and more work is needed to establish how the effects may or may not generalize across variables. Although there are many interesting avenues to consider for future studies, I will focus on discussing two areas of research related to the characteristics of the individual.

(1) Linguistic variables of the individual:

Language familiarity emerged as a key factor influencing the Production Effect in both adults and children. Although numerous studies have examined this phenomenon using familiar and unfamiliar stimuli, further investigation into the role of language background and experience in children could provide deeper insights into how familiarity and language skills shape the Production Effect. While Chapter 3 - López Assef et al. (2023) focused on the impact of vocabulary size, incorporating a broader range of controlled measures—such as multiple measures of vocabulary size or language exposure—could offer a more comprehensive understanding of how language experience affects the Production Effect. Furthermore, most studies on the Production Effect are done with English speaking participants and English stimuli. It would be interesting to look into learning in new languages, across different language backgrounds or across a continuum of language experience or proficiency.

(2) Cognitive measures of the individual:

Future research examining cognitive skills and additional measures is crucial for a deeper understanding of the Production Effect in both children and adults. So far, three studies have included cognitive or language measures (see Table 16 for details): Junttila and Ylinen (2020) measured memory, reasoning and phonological skills in 5–8-year-olds, Chapter 3 - López Assef et al., (2023) measured vocabulary size in 5-6-year-olds and Richtsmeier et al. (2024) measured language, literacy and inhibition skills and phonological short-term memory in 3-4-year-olds. None of these measures was found to predict the Production Effect. However, as a central prediction of the proposed framework is that an individuals' cognitive skills is related to the Production Effect, future research incorporating these factors are essential to better understand the source behind changes in the direction of the Production Effect (also see Baese-Berk et al., 2024 for a similar discussion).

Furthering our knowledge on the Production Effect will lead to a better understanding of the role of speech production on language learning across development, will provide further insights into the type of information that is encoded in lexical representations or in memory, and the type of information that is available for later retrieval. The Production Effect also has pedagogical applications as a mnemonic device. The goal is to have evidence-informed methods that can be integrated into educational settings to establish best-practices for children (and adults) to learn or memorize language and/or new languages. This could be especially effective because speech is a simple, accessible task that requires no additional materials, making it a useful teaching tool regardless of a school or family's socioeconomic context. This was especially evident during my time conducting testing at the Canada Science and Technology Museum. One of our key goals is not only to engage participants in our study but also to share our research and knowledge with the public. We do this, in part, through study inserts (see Appendix 1)—pamphlets that provide information about the study and include related activities for families to do both at the museum and at home. Parents were consistently interested and receptive, often sharing stories about how they and their children have used speech production as a tool for memorization. Understanding the contexts in which speech production aids and does not aid! language learning will inform teaching methods and resources, making the learning process easier for children.

References

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von 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,* 1-36. https://doi.org/10.1007/s10648-020-09527-z

Baddeley, A. D. (1986). Working memory. Oxford, England: Oxford University Press. https://doi.org/10.1126/science.1736359

Baddeley, A. D. (2002). Is working memory still working?. *European Psychologist*, *7*, 85–97. https://doi.org/10.1027/1016-9040.7.2.85

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A., Gathercole, S., Papagno, C., & Bjork, Robert A. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*, *105*(1), 158-173. https://doi.org/10.1037/0033-295x.105.1.158

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

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641-1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Camos, V., and Barrouillet, P. (2014). Attentional and non-attentional systems in maintaining verbal information in working memory: the executive and phonological loops. *Frontiers in Human Neuroscience*, *8*, 900. https://doi.org/10.3389/fnhum.2014.00900

Camos, V., Lagner, P. L., and Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*. *61*(3), 457–469. https://doi.org/10.1016/j.jml.2009.06.002

Carey, S., and Bartlett, E., (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15,* 17-29.

Courage, Bakhtiar, Fitzpatrick, Kenny, & Brandeau. (2015). Growing up multitasking: The costs and benefits for cognitive development. *Developmental Review*, *35*, 5-41. https://doi.org/10.1016/j.dr.2014.12.002

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Forrin, N. D., & MacLeod, C. M. (2016). Auditory presentation at test does not diminish the production effect in recognition. *Canadian Journal of Experimental Psychology*, *70*(2), 116–124. https://doi.org/10.1037/cep0000092

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

Gualtieri, S., & Finn, A. S. (2022). The sweet spot: When children's developing abilities, brains, and knowledge make them better learners than adults. *Perspectives on Psychological Science*, *17*(5), 1322-1338. https://doi.org/10.1177/17456916211045971

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

Junttila, K., & Ylinen, S. (2020). Intentional training with speech production supports children's learning the meanings of foreign words: A comparison of four learning tasks. *Frontiers in Psychology*, *11*, 1108. https://doi.org/10.3389/fpsyg.2020.01108

Kan, P. F., & Kohnert, K. (2008). Fast mapping by bilingual preschool children. *Journal of Child Language*, *35*(3), 495-514. https://doi.org/10.1017/S0305000907008604

Kan, P. F., & Kohnert, K. (2012). A growth curve analysis of novel word learning by sequential bilingual preschool children. *Bilingualism*, *15*(3), 452-469. https://doi.org/10.1017/S1366728911000356

Kan, P. F., Sadagopan, N., Janich, L., & Andrade, M. (2014). Effects of speech practice on fast mapping in monolingual and bilingual speakers. *Journal of Speech, Language, and Hearing Research*, *57*(3), 929-941. https://doi.org/10.1044/2013_JSLHR-L-13-0045

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

Kohonen, V., & Service, E. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition?. *Applied Psycholinguistics*, *16*(02), 155-172. https://doi.org/10.1017/S0142716400007062

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*. https://doi.org/10.1186/s40359-017-0198-8

Lewandowsky, S., & Oberauer, K. (2015). Rehearsal in serial recall: An unworkable solution to the nonexistent problem of decay. *Psychological Review, 122*(4), 674. https://doi.org/10.1037/a0039684

López Assef, B. & Zamuner, T. S. (In prep). Children aged 4- and 6-years old learning new words in a storybook setting: recognition is better when just listening rather than producing aloud for younger children. *Developmental Psychology.*

López Assef, B. & Zamuner, T. S. (submitted). Task effects in children's word recall: Expanding the reverse production effect. *Journal of Child Language.*

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., McDonald, M., Bernard, A., & Zamuner, T. S. (2024). Effects of speech production training on memory across short and long delays in 5-and 6-year-olds: A pre-registered study. *Applied Psycholinguistics*, 1-24. https://doi.org/10.1017/S0142716424000109

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). https://doi.org/10.16995/glossa.8582

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The Production Effect: delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology, 25*(4), 530-535. https://doi.org/10.1016/j.cub.2014.11.068

Michas, I. C., & Henry, L. A. (1994). The link between phonological memory and vocabulary acquisition. *British Journal of Developmental Psychology*, *12*(2), 147-163. https://doi.org/10.1111/j.2044-835X.1994.tb00625.x

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *The Quarterly Journal of Experimental Psychology Section A*, *44*(1), 47-67. https://doi.org/10.1080/14640749208401283

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 98-107. https://doi.org/10.1080/14640749508401378

Perez Repetto, L., Jasmin, E., Fombonne, E., Gisel, E., & Couture, M. (2017). Longitudinal study of sensory features in children with autism spectrum disorder. *Autism Research and Treatment*, 1934701. https://doi.org/10.1155/2017/1934701.

Pritchard, V. E., Heron-Delaney, M., Malone, S. A., & MacLeod, C. M. (2019). 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

Richtsmeier, P. T., Gladfelter, A., & Moore, M. W. (2024). Contributions of Speaking, Listening, and Semantic Depth to Word Learning in Typical 3-and 4-Year-Olds. *Language, Speech, and Hearing Services in Schools*, 1-14. https://doi.org/10.1044/2024_LSHSS-23-00105

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 21-50. https://doi.org/10.1080/14640749208401314

Shams, S., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Thorn, A. S., & Gathercole, S. E. (2001). Language differences in verbal short-term memory do not exclusively originate in the process of subvocal rehearsal. *Psychonomic Bulletin & Review*, *8*(2), 357-364. https://doi.org/10.3758/BF03196173

Vlach, H., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology*, *3*, 46. https://doi.org/10.3389/fpsyg.2012.00046

von Kriegstein, K., & Giraud, A. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology*, *4*(10), e326–e326. https://doi.org/10.1371/journal.pbio.0040326

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. (2018). 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

Appendix A – Study Inserts

Figure A1. Study insert for Chapter 2 – López Assef et al., (2021)

The Living Lab

Making breakthroughs in child research studies
Come visit our friendly researchers!

Which learning strategy best helps children remember words: saying, listening, or looking?

When learning new words, does it help if a child says the word aloud? Many of us would answer, 'Yes, of course!'. New research suggests that there are some cases where saying the word aloud does not help with learning. Thus, we are interested in investigating the situations in which repeating a word contributes to word learning.

We ask children to play a computer game with us in order to teach Mr. Wiggles (our teddy bear mascot) words. Each child is presented with pictures and are asked to either (a) *look* at the pictures, (b) *look and listen* – observe the pictures and hear the word presented by the computer; or (c) *look and say* – observe pictures and name the object they see aloud. After a short play break, children are asked to teach Mr. Wiggles new words by naming as many objects from the game as they can remember. By having children learn words in 3 different ways (*look, look and listen, look and say*), we can find out which learning strategy is most effective.

So far, results show that younger children (3-4) recall more words that were heard, and older children (5-6) recall more words that they've said out loud. This research will help us better understand how children learn language and help guide how language is taught in classrooms. While these findings are interesting, they do not tell us the whole story! We are also interested in investigating bilingual children, to determine if there are benefits or disadvantages to learning more than one language at an early age.

This research is conducted at the Canada Science and Technology Museum, Ottawa. The research is supported by the University of Ottawa and a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to Tania Zamuner, Primary Investigator at the Centre for Child Language Research (CCLR).

Email: livinglab@uottawa.ca **Web: uOttawaLivingLab.weebly.com**

TO TRY AT THE MUSEUM:
Active Play in Exploratek

Play creates fun opportunities to **learn new words**. While exploring *Exploratek*, participate in hands on activities with your child. If there is an object that they do not know the name of, name it for them and have them name the object as well. Does your child later remember the object? Do they remember an attribute (e.g. the sound it makes, the way it works, etc.)?

Invent in the ZOOOM Zone

While exploring the ZOOOM area with your child, check out the building stations (as displayed below).

With your child, create one of the objects and name it aloud. Encourage your child to repeat. After playing with their creation, ask your child for the name of their creation. Do they remember the object name? Do they talk about the object when retelling their visit to the museum?

ACTIVITIES TO TRY AT HOME:
What Did You See?

When taking a walk with your child, **ask them to name** different things that they see around them. Afterwards when returning home, **ask your child what they saw** throughout their walk. Do they name things they haven't mentioned, or are they more likely to remember items that they've named out loud?

CANADA SCIENCE AND TECHNOLOGY MUSEUM

uOttawa

Figure A2. Study insert for Chapter 4 – López Assef et al., (submitted)

The Living Lab
Making breakthroughs in child research studies
Come visit our friendly researchers!

uOttawa CANADA SCIENCE AND TECHNOLOGY MUSEUM Follow Us On twitter

Can some actions distract us from learning?

When trying to memorize information, it is very common to practice by saying words aloud. New research with children has suggested that there are some cases in which saying a word does not improve their memorization. While this is surprising, research has also suggested that the memory benefit from saying words aloud depends on many factors like the age of the person, cognitive skills and characteristics of the information we are trying to memorize (is this new or previously known information, is it in another language). For children, attention and cognitive skills seem to be a key factor, doing something (in this case saying the word aloud) while memorizing could distract them from learning. In this study, we are trying to identify which type of actions can sometimes distract children from learning.

In the Living Lab, we invite children to play a game on the computer where they learn new words for new friends. Children are presented some pictures of animals. For half of these, they only hear the name of the animal and for the other half they also either (a) repeat the name after hearing it, (b) stick out their tongue after hearing it, or (c) touch their nose after hearing it. By learning words with these 3 different tasks, we are trying to identify which is more distracting from learning a new word.

To see which action distracts children more from learning words, we use two methods: we use a camera that tracks eye movements to see how well children recognize the animals, and we also ask them to say the words they remember.

We predict that if children remember fewer words that were learned while speaking aloud compared to words learned with the other actions, this might mean that it is language-specific actions (and not performing any other action while learning) that cause a heavier burden on their brains, distracting them from memorizing. Understanding what makes learning easier for children can help us improve how children are taught and make their learning journey better.

ACTIVITIES TO TRY AT HOME:

Silly Simon Says!

Play the game "*Simon Says touch the* ____", and add a silly action to certain words, either new words or words in another language. For example, for the word "*door*" we touch the door, but for the word "*tapis*" we touch the rug and stick our tongue out!

Learning While Exploring!

When out for a walk, say the names for different objects that your child has not learned yet, or that he or she does not know in a different language! Have your child repeat some of the words and then do a fun action such as sticking out their tongue whenever you say the word aloud. When you arrive home, ask them if they remember any of the words!

email: livinglab@uottawa.ca
web: uottawalivinglab.weebly.com

This research is being conducted at the Canada Science and Technology Museum (CSTM) by researchers from the Centre for Language Research (CCLR) at the Department of Linguistics at the University of Ottawa. The Director of the CCLR is Dr. Tania Zamuner.