# WORD LEARNING IN MANDARIN-SPEAKING CHILDREN WITHAND WITHOUT AUTISM SPECTRUM DISORDER: THE ROLE OF SEMANTIC HEAD

YU JIAYU

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

# Word learning in Mandarin-speaking children with and without Autism Spectrum Disorder: The role of semantic head

YU Jiayu

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2024

# CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_________________________ (Signed)

YU Jiayu (Name of student)

ABSTRACT

Characteristics of to-be-learned words are influencing factors in word learning but have been understudies in Mandarin-speaking children, especially in children with Autism Spectrum Disorder (ASD). Capitalizing on the properties of Mandarin, a language in which 65% of the vocabulary consists of compound words, the current study focuses on the role of semantic head, which provides learners with information about the semantic category of the word. For example, the right morpheme "花/hua2/(flower)" specifies the semantic category of the compound word "桃 花 /tao1 hua2/ (peach flower)". By using multiple tasks across several time points, the current thesis includes two studies that examined the effect of semantic head in word learning among Mandarin-speaking children with and without ASD. Furthermore, the predictive role of phonological memory and compound awareness has been examined.

By ameliorating methodological gaps observed in previous research, Study 1 shows that the facilitative role of semantic head was observed in encoding phonological representations for Mandarin-speaking typically developing children, while its effect was absent at the retention stage. The salience of semantic head was further highlighted by the naming errors, as although children showed difficulty in retrieving the complete phonological representations of the newly taugth words, they could produce the semantic heads of these words. Phonological memory and compound awareness predicted children's encoding performance. The role of compound awareness was more salient in predicting children's learning of words with semantic head, whereas phonological memory and compound awareness were equally important in predicting children's learning of words without semantic head. Moreover, instead of overnight improvement, children showed the evidence of forgetting in the naming task after a one-day delay, which may relate to preschoolers' sleep patterns. These findings support the necessity of taking word characteristics and learners' abilities into consideration when constructing a theoretical model that accounts for word learning for children aged over three years.

Study 2 focuses on Mandarin-speaking children with ASD. Many children with ASD have smaller vocabularies than their TD peers, while they do not exhibit significant deficits in many aspects of word learning in laboratory studies of word learning. Study 2 shows that children with ASD can establish form-referent links and form phonological representations at the encoding stage. However, they showed differences in retaining the phonological representations over a one-week delay. Sleep quality was related to the changes in retention performance one-day and one-week after training only in children with ASD. Moreover, children with ASD also showed extraordinary difficulties in the definition task at both encoding and retention stages. The effect of semantic head was not observed in verbal children with ASD at either the encoding or retention stage, a possible consequence of their deficits in lexical-semantic processing and procedural learning. Phonological memory and compound awareness explained group differences in selecting correct word forms and predicted the encoding performance of children with ASD, with phonological memory being more critical in encoding new words. These key findings yielded several clinical implications that may inform the study of vocabulary interventions for Mandarin-speaking children with ASD.

## PBULICATIONS ARISING FROM THE THESIS

Duan, W., **Yu, J**., & Sheng, L., (2024, August). The development of definitional skill in Mandarin-speaking Preschoolers: Evidence from a novel word learning experiment. 2-minute Madness Teaser Session at the 2nd PolyU Research Student Conference (PRSC 2024). Hong Kong.

Duan, W., **Yu, J**., & Sheng, L., (2024, August). The Relationship Between Mandarin-Speaking Preschoolers' Vocabulary Skills, Home Language Environment, and the Development of Definitional Skills. Oral presentation at the Second International Symposium on Child Language, Speech Acquisition and Cognitive Development. Ürümqi, Xinjiang.

**Yu, J** & Sheng, L. (2024, June). Word learning in Mandarin-speaking Children with and without Autism Spectrum Disorder: The role of word characteristics. Poster presentation at the 44th annual Symposium on Research in Child Language Disorders, Downtown Madison, Wisconsin.

**Yu, J**. & Sheng, L. (2023, November). The Role of Semantic Head in Word Learning: A Pilot Study from Mandarin-Speaking Preschoolers. Poster presentation at the Annual Convention of the American Speech-Language-Hearing Association, Boston, Massachusetts.

# ACKNOWLEDGEMENTS

People always become emotional when start to look back on the past, especially late in the night. When I firstly considered pursuing a PhD degree five years ago, I never thought I would complete my thesis in Hong Kong. This may be the magic of fate.

I was also wondering that if it is acceptable to write the acknowledgements in Chinese as majority of people that I would like to thank are native Mandarin speakers and some of them may not understand English. Anyway, I think I can write whatever I want only if I did express my thankfulness.

First and foremost, I would like to express my deepest gratitude to my chief supervisor, Professor SHENG Li, for her unwavering support, guidance, patience and encouragement throughout this journey. Her insightful feedback and constructive suggestions have been invaluable in shaping this research. She is always patient and meticulous in revising every version of this thesis. Prof. Sheng creates an ideal and warm environment. Whenever I needed help, she was always able to respond to me promptly. I am always looking forward to the personal meetings and lab meetings with her. In addition to academic activities, we also have many valuable and interesting moments that I can proudly share on social media. I am very fortunate to complete my PhD journey with Prof. Sheng.

I would like to express my heartfelt thanks to Professor LIANG Dandan and Language Development Research Center at Nanjing Normal University for assistance in recruiting participants for this thesis. Prof. Liang has directed my development and helped me realize my potential. I am sincerely grateful to Prof. Liang for her support over these years. I am indebted to Dr. SONG Yiqi for always being there and hearing my concerns and anxiety. I am truly thankful to TANG Mingyue and WANG Ruoting for helping my collect the data. Special thanks go to HU Han for answering statistical questions, to BAI Yuxuan and CHEN Haoyun for recording materials for this project. 非常感谢紫金幼儿园的黄园长和杨园长，星星荟的周老师，磨方成的杨老师，环南路小学的赵老师，心萌贝贝的赵老师，和金钥匙的黄校长和朱老师在数据收集过程中给予的帮助和关心。感谢所有参加本研究的儿童及家庭。祝愿孩子们能够健康快乐的成长。

I am also thankful to my colleagues and friends in the Language Learning and Bilingual Lab for creating an inspiring and motivating environment, including Yingnan, Luyuan, Haolun, Jingdan, Katherine, Jin Tong, Jiahan, Shuying, Jingbo, Zhao Xi, and Leo. Special thanks to NING Jinghong for her continuous support for data analysis and DUAN Weihua for data transcription.

I would like to express my sincere thanks to the Hong Kong SAR Government, the Faculty of Humanities, Graduate School and the Department of Chinese and Bilingual Studies (CBS) of the Hong Kong Polytechnic University for their financial support for this project.

I am fortunate to meet my colleagues and friends Hansy LIN Jueyao, PENG Yingying, Dr. WENG Yi, Dr. KONG Deran, and Dr. GU Chanyuan in CBS PolyU. Every meal, phone call and discussion we had together made my days in Hong Kong colorful and worthwhile!

I am truly thankful to WANG Yue for always answering my phones for hours when I was having an emotional breakdown. My heartful appreciation also goes to YU Kefeng, LIU Chang, and Cai Duyu for being my friends and giving emotional support. Special thanks to the KPOP group Seventeen for providing me with a space to escape from the stress with their music and content.

My heartfelt thanks go to my parents YU Zhibai, Cao Xiaofang and grandparents YU Weichao, Miao Yuruo, for their endless love, patience, and belief in me. Thank you for the willingness to hear my thoughts, for understanding, and for always standing by my side.

Thank me for not giving up.

## TABLE OF CONTENT

# LIST OF TABLES

# LIST OF FIGUES

# Chapter 1 Introduction

As an essential component of language, vocabulary constitutes the foundation of children's daily communication. The critical value of vocabulary acquisition and development is well-established and self-evident. The production of the first meaningful word is a milestone for children's language development. This milestone commonly emerges near the end of children's first year of life (Hoff-Ginsberg, 2014). Children's expressive vocabulary grows slowly but steadily between 12 to 24 months, from dozens to hundreds. By the end of the second year of life, typically developing (TD) children are usually able to say over 200 words(Frank et al., 2017; Owens, 2020). Children's vocabulary growth spurts around three-years-old and they are able to reach the threshold of 1000 (Owens, 2020). Receptive vocabulary typically develops faster than expressive vocabulary (Hoff-Ginsberg, 2014). Children are able to comprehend approximately 50 words at around 13 months of age (Benedict, 1979), and over 200 words around 18 months (He & Arunachalam, 2017). Around the time when children are about to formally enroll in school education (i.e., six to seven years old), they usually can say over 2000 words and comprehend over 20,000 words (Owens, 2020). In about 18 years, children acquire about 60,000 words (McMurray et al., 2012). The general pattern of vocabulary growth shows cross-language consistency. Li (2004) summarized findings from several studies on vocabulary development in Mandarin-speaking children, noting that the trajectory of vocabulary growth aligned with the patterns introduced earlier.

The growth of vocabulary is anchored in word learning. The process of learning spoken words is multifaceted and entails various components. Researchers have different opinions on the components and stages of the word learning process (Gordon et al., 2022; Hoover et al., 2010; Leach & Samuel, 2007). The current study did not intend to address the question "what is word learning" and adapted a functional perceptive to identify the components and stages of word learning. Adapting the simple functional framework of word learning (Gupta & Tisdale, 2009), word learning entails the following core components: learning the word form, learning the referent, and establishing the link between the form and referent. Taking the concrete noun as an example, during the first few exposures, children need to encode both auditory and visual information of the new word to create the initial phonological and semantic representation and the link between these two representations (Gordon et al., 2022; Gray et al., 2020). Generally, the auditory information refers to the phonological sequence that makes up the word form, and the visual information refers to the physical and functional features of the referent at a minimum (McGregor et al., 2007). This functional framework is supported by empirical research. Gray et al. (2022) created several latent variable models to examine the structural model of word learning based on multiple theoretical models and found that the model with independent phonological, semantic, and form-referent link indicators best fit the word learning performance of 167 school-age children (mean age = 92.82 months).

Learning outcomes measured immediately after learning may not adequately reflect learners' ability to learn words (McGregor et al., 2017). After this early stage of learning, newly created lexical representations should be retained in memory for later retrieval

(Wojcik, 2013). For example, the newly formed form-referent links is quite unstable with decreased referent-identification accuracy observed 5 minutes after learning (Horst & Samuelson, 2008). The importance of examining retention performance during word learning has been well emphasized in the recent years (Gordon et al., 2016; Henderson et al., 2013; James et al., 2017; Wojcik, 2013). Following the suggestion of moving beyond mapping (Wojcik et al., 2022), the current study would also include the retention stage to investigate children's word learning abilities. To sum up, the current study focused on the following components: the encoding of word forms and referents, the establishment of form-referent links, and the retention of newly learned words.

So how do children acquire their very first vocabulary? What happened during early word learning? Researchers believe that the first challenge for the word learning novice is to find the WORD within the speech stream (He & Arunachalam, 2017). There are various cues within children's language input to assist them in completing the task of finding word boundaries, including the transitional probabilities between syllables, the prosodic characteristics specific to each language (e.g., stress in English), and the child-directed speech (e.g., single word utterances) (Hoff-Ginsberg, 2014). Saffran and colleagues (1996) found that 8-month-old infants showed the ability to use the transitional probabilities of their native language to segment individual words. Once the word forms have been identified, the next task is to find the referent and establish the association. Young word learners have to address the referential ambiguity during the learning process, that is, there are infinite referents for an unknow form in any naming situation (Quine, 1960). There are several cues and mechanisms available to young word learners to overcome this challenge under different learning conditions, including social cues (e.g., pointing, speaker's direction of gaze), "innate" or early assumptions (e.g., mutual exclusivity, the whole-object assumption), the frequency of the form-referent co-occurrence (i.e., cross-situational word learning), and linguistic cues (e.g., syntactic bootstrapping) (He & Arunachalam, 2017; Wojcik, 2013). The significance of these cues and mechanisms has been emphasized in the initial stage of vocabulary growth, as evidenced by the fact that the majority of previous studies have investigated the availability of these cues in children aged under 3 years (He & Arunachalam, 2017).

As discussed earlier, around the age of three, children have accumulated an initial mass of vocabulary and entered a period of rapid vocabulary growth. From this stage onwards, researchers have investigated the factors that influence children's word learning, including the characteristics of to-be-learned words and the abilities of learners (Gray et al., 2014). Each word possesses various characteristics based on different criteria. For example, a word can be classified as a noun, verb, or adjective based on its part of speech. Within the current framework of word learning, characteristics of to-be-learned words can be related to the forms (e.g., phonological neighborhood density) or the referents (e.g., polysemous and homonymous words). Characteristics of to-be-learned words have received attention from researchers, and their effects have been observed from preschoolers (as young as three) (e.g., Storkel, 2001) to adults (e.g., James et al., 2019). These findings suggested that word learning is a life-long project, and that the influencing factors of word learning are developmentally variable. In other words, the frequency of the form-referent occurrences is a salient influential factor for infants, while word characteristics such as phonological neighborhood density and polysemy emerge as critical for preschoolers. Therefore, adopting a developmental perspective to identify the factors that influence children's word learning at different stages would be informative for constructing a theoretical model of word learning. Furthermore, when considering the influence of word characteristics, attention should be paid to language-specific characteristics, as languages differ in their lexical structures. Study 1 of this thesis (Chapter 2) investigated the effect of the semantic head, a type of word characteristics that specify the semantic category and is salient in Mandarin words, on word learning in Mandarin-speaking children.

Further evidence that supports the critical value of word characteristics comes from children who showed delayed vocabulary. For example, Gray et al. (2014) found that unlike in typically developing (TD) children, word form characteristics did not influence the naming performance of preschoolers with Specific Language Impairment (SLI). These findings (also see Floyd et al., 2021) also help to explain these children's delayed vocabulary development. The current study investigated the role of semantic head in word learning for Mandarin-speaking children with Autism Spectrum Disorder (ASD). The facilitative effect of semantic head has been observed in Mandarin-speaking children with SLI (Chen & Liu, 2014; Ma & Liang, 2019), but there is no evidence from children with ASD. Considering that the semantic head commonly specifies the semantic category, it would be informative to examine its effect on word learning in children with ASD, who are likely to exhibit lexical-semantic processing deficits (Naigles & Tek, 2017). Moreover, previous studies suggested that although verbal children with ASD often have delayed vocabulary development, they can nonetheless use the cues and mechanisms that are essential for early word learning (Arunachalam & Luyster, 2016). As discussed in the last paragraph, the weighting of various factors would vary at different stages of vocabulary development. Therefore, it is important to examine the effect of word characteristics on word learning in children with ASD aged over three years. Study 2 of this thesis (Chapter 3) investigated the effect of semantic head on word learning in Mandarin-speaking children with ASD.

Another salient feature of children's vocabulary development after age three is the significant individual differences in word learning (Hoff-Ginsberg, 2014; Samuelson, 2021). Hence, it is crucial to examine the relationship between learners' abilities and learning outcomes (Gray et al., 2014; Pomper et al., 2022). Identifying which abilities predict children's word learning performance has significant practical implications in vocabulary instruction, especially for children with delayed vocabulary development. Among various abilities, the role of phonological memory and vocabulary abilities has been emphasized (Gray et al., 2022). To be more specific, children who have better phonological memory skills (e.g., Gordon et al., 2022) and /or larger vocabulary size (e.g., James et al., 2017) would perform better in word learning tasks. Furthermore, Pomper et al. (2022) has found that the differences in cognitive and vocabulary abilities between TD children and children with Developmental Language Disorder (DLD) accounted for their group differences in word learning performance. For children with ASD, previous studies have found that receptive vocabulary size is a significant predictor of their word learning success (Hartley et al., 2020; McGregor et al., 2013), while no study to our knowledge has examined the role of phonological memory. The two studies in this thesis also investigated the role of phonological memory and language abilities in children's learning performance. Considering that we were interested in the role of semantic head, instead of general vocabulary abilities, such as vocabulary size, we investigated the role of compound awareness because morphological awareness, more specifically, compound awareness, should play a crucial role in learning novel words with semantic heads.

This thesis consists of four chapters and two studies. The current chapter (Chapter 1) introduces the research background, the framework to examine the word learning process and the key word-level and learner-level variables. The two studies contained in this thesis investigated whether the presence of semantic head affected Mandarin-speaking children's performance in learning novel words. Multiple aspects of word learning were examined, including how well children link forms to referents, receptive and expressive knowledge of phonological representations, and expressive knowledge of semantic representations. The retention of these phonological and semantic representations was measured at the following three time points: one-day, one-week, and one-month after training. Learners' abilities, that is, phonological memory and compound awareness, were examined at the encoding stage. Chapter 2 presents Study 1, which focuses on typically-developing (TD) Mandarin-speaking children. Chapter 3 presents Study 2, which focuses on Mandarin-speaking children with ASD. Each study is self-contained, with separate sections for Literature Review, Method, Results, Discussion and Conclusion. Chapter 4 provides a general discussion of the results from these two studies.

# Chapter 2 The effect of word characteristics on word learning: Does the presence of semantic head facilitate novel word learning in Mandarin-speaking children?

## 1 Introduction

### 1.1 Word characteristics

Existing studies examining the effect of word characteristics on word learning have commonly focused on form characteristics. Two interrelated characteristics of word forms, phonotactic probability (the likelihood of a sound sequence occurring in a language) and phonological neighborhood density (the number of phonologically similar words based on the substitution, addition or deletion of one phoneme) interactively influence the word learning performance of English-speaking children and adults (Gray et al., 2014; Hoover et al., 2010; James et al., 2019; Storkel & Hoover, 2011). For example, Hoover et al. (2010) investigated the interaction between phonotactic probability and neighborhood density in English-speaking preschoolers (three to five years old). Results showed that children's naming performance of novel words with low probability and low density was better than words with low probability and high density, and their naming performance of novel words with high probability and high density was better than words with high probability and low density, suggesting that these two factors converged in influencing children's word learning performance. James et al. (2019) examined the effect of characteristics of word forms on a large sample of school-aged English-speaking children. 232 children aged between 7;3-9;3 ([years; months], the same below) learned novel words with no or many orthographic neighbors. Novel words with no orthographic neighbors also had significantly lower phonological neighborhood density and phonotactic probability. As evidenced by the higher accuracy in cued form recall task after training, children learned words with more neighbors better than words with no neighbors. In general, the effect of word form characteristics (e.g., the facilitative effect of higher neighborhood density) was observed in English-speaking individuals. These findings have suggested that when a learner is creating the representation for a novel word, the existing representations within the learner's mental lexicon would be activated, and actively influence learner's learning performance.

In regard to the timeline of this effect, existing studies have found that word characteristics influence children's learning performance immediately after training. Could the influence of word form characteristics still be observed at the retention stage? To answer this question, James et al. (2019) also measured learning outcomes on the next day and one-week later and found that the high neighborhood density benefit was absent at both time points. Hoover et al. (2010) also found that the low density advantage for low probability words did not persist into the one-week retention stage for 5-year-old children. Therefore, the available findings suggested the effect from characteristics of word forms was salient at the encoding stage, but diminished at the retention stage.

When considering the influence of word characteristics, attention should be paid to language-specific features, as some languages possess unique linguistic properties that may influence learners' encoding and/or retention of new words. Ferman (2021) examined the morphological pattern that marks the categorical information of words, in Hebrew-speaking children with and without ASD. Since the current study focused on typically developing (TD) children, only TD group's performance would be discussed in this section (the same below). Participants learned two types of novel words, one type consisting of existing Hebrew morphophonological patterns, and the other type consisting of pseudo-Hebrew morphological patterns. A referent-selection task and a naming task were administered immediately after training. The facilitative role of existing morphological patterns was observed in TD children (mean age = 6;11) in both tasks.

Similar to the morphological pattern observed in Hebrew, Mandarin words possess morphological characteristics that could influence children's learning performance. The syllable structure of Mandarin is simple (Taylor & Taylor, 2014) and a large portion of Mandarin words are disyllabic compound words. Chen and Duan (2016) reported that there are 42,165 disyllabic compound words in the Dictionary of Modern Chinese (5th edition) out of more than 65,000 entries, accounting for about 65% of the Chinese lexicon. A subset of the compound words provides learners with information about the semantic category of the word. For example, the compound word "桃 花 /tao2 hua1/ (peach flower)" has the right morpheme "花/hua1/(flower)" specifying the semantic category, and the left morpheme "桃/tao2/(peach)" specifying the subcategory. Head morphemes are semantically transparent and highly productive, and once Mandarin-speaking children are aware of its essential role, they can apply this knowledge to accelerate their word learning (Chen et al., 2009). Evidence from Mandarin-speaking children's spontaneous speech indicates that children as young as three years old have been reported to produce invented compounds, such as "盒菜 /he2 cai4/ box of vegetables", by drawing analogies from real compounds, such as "盒饭 /he2 fan4/ box of rice" (Liang, 2017). Chen et al. (2009) found that school-aged Mandarin-speaking children's compound awareness (measured by the ability to identify the semantic head of a compound noun) was significantly related to their oral expressive vocabulary.

So, would the effect of semantic head be observed in the process of word learning for Mandarin-speaking children? Chen and Liu (2014) manipulated the availability of semantic heads in their novel words and investigated the word learning performance of Mandarin-speaking TD children (mean age = 65.6 months) and children with specific language impairments (SLI) (mean age = 65.4 months). Children learned three types of novel words composed of real syllables: 2-syllable, 3-syllable, and 3-syllable with a semantic head. The training session consisted of two phases. In phase 1, children were exposed to an incidental word learning task, in which novel words were taught by listening to stories. A referent-identification task was used to assess the learning outcomes after phase 1. In phase 2, children were repeatedly exposed to the novel form-referent pairs one by one across ten rounds. After each round, the learning performance was measured by a naming task. As mentioned earlier, only TD group's performance would be discussed in this section. Results showed that Mandarin-speaking TD preschoolers learned novel words with semantic heads better than those without semantic heads. This benefit was found not only in establishing the initial associations in phase 1 but also in the naming task in phase 2.

Ma and Liang (2019) extended Chen and Liu (2014)'s study to school-aged (7 to 9 years old) Mandarin-speaking children with and without SLI. They designed four types of disyllabic words with real Mandarin morphemes that varied in the degree of semantic transparency, that is two meaningless morphemes (meaningless words), meaningless morpheme + categorical morpheme (categorical words), feature morpheme + meaningless morpheme (feature words), and feature morpheme + categorical morpheme (feature and categorical words). During the training phase, one novel form-referent pair was presented to children in each trial. After the passive exposure, a referent-identification task and a naming task were used to measure the learning outcome. A total of three rounds of training and testing were conducted. For TD group, significantly better learning outcomes for categorical words compared to meaningless words were observed in both tasks. Furthermore, in the referent-identification task, no significant differences were found among the three types of words with meaningful morphemes. In the naming task, the best performance was observed for feature and categorical words, and significantly better performance was found for categorical words compared to feature words.

These two studies provided the initial empirical evidence of the facilitative role of semantic head in the word learning for Mandarin-speaking children. However, there are some improvements that can be made to further clarify its role. First, the stimuli in both studies were not well controlled. The novel words with and without heads referred to objects from different semantic categories. In Chen and Liu (2014), the novel words with heads were from the categories of birds and flowers, whereas the novel words without heads were from the categories of toys and artifacts. This introduced a potential confound because we do not know if the boost in learning was due to the referents being inherently easier to identify and name or due to the facilitation of semantic head. Second, both studies used real Mandarin syllables to construct word forms without considering the potential influence of phonological characteristics of word forms (Gray et al., 2014). As discussed in the previous section, the influence of the characteristic of word forms (e.g., neighborhood density) has been well-established. This also introduced a potential confound since we cannot be sure that the observed advantage was from some syllables having higher phonological neighborhood density or from the presence of semantic head. Last but not the least, by only using the referent-identification task and naming task just after learning, the effect of semantic head on the semantic representations and retention remained unexplored. By ameliorating these methodological gaps, the current study aimed to comprehensively investigate the role of semantic head in Mandarin-speaking children's word learning.

### 1.2 Learners' abilities

Children's abilities are key factors that would influence their word learning performance. Among various abilities, the essential role of children's language and cognitive abilities has been well established (Archibald, 2018). Previous studies have found, not surprisingly, a positive correlation between children's vocabulary size and learning outcomes, suggesting that the larger their vocabulary size, the easier it is for children to learn new words (Adlof & Patten, 2017), especially for those who would show delayed vocabulary development (Hartley et al., 2020; McGregor et al., 2013).

The relationship between children's vocabulary size and word learning abilities can be seen as a manifestation of the Matthew Effect in the field of word learning (Gordon et al., 2022). James et al. (2019) also found that children's expressive vocabulary knowledge (measured by a written definition task) was a significant predictor of school-aged children's word learning outcomes. As for cognitive abilities, the role of working memory has been highlighted. Working memory is a type of memory process that temporarily stores and manipulates incoming visual and verbal information(Gray et al., 2022; Wang et al., 2017). The newly formed lexical representation would be stored in the working memory, waiting to be transferred into long-term memory (McGregor et al., 2017). Depending on the task requirement (maintenance or manipulation) and stimulus modality (verbal or visual), the type of working memory varies and includes at least the following types: phonological short-term memory, phonological working memory, visuospatial short-term memory, and visuospatial working memory (Wang et al., 2017). Gordon et al. (2022) examined word learning performance in a group of children aged from 4-to-6 years old and found that children's phonological short-term memory (as measured by the nonword repetition task) was a significant predictor of their learning of the phonological representations, indicating that children who had better phonological short-term memory skills would perform better in encoding new word forms.

Existing studies also examined the relative influence of working memory and language abilities. Adlof and Patten (2017) examined the contribution of phonological memory (measured by the nonword repetition task) and vocabulary abilities (i.e., receptive vocabulary size) to word learning in a group of children aged between 5 to 12 years. Results showed that phonological memory contributed more to children's receptive phonological and semantic knowledge, while vocabulary abilities contributed more to children's expressive semantic knowledge. Pomper et al. (2022) comprehensively measured children's working memory skills (including both phonological and visuospatial short-term and working memory) and found that phonological memory (reflected by the composite score of the nonword repetition and the backward digit tasks) was a significant predictor of children's performance in creating the form-referent links and encoding phonological representations of new words. Compared to other abilities, including visuospatial memory, vocabulary and sustained attention, the role of phonological memory was more salient (Pomper et al., 2022). Furthermore, Pomper et al. (2022) found that children's cognitive and language abilities did not predict the retention of word forms. Considering the nature of working memory, that is, temporarily stores and manipulates incoming visual and verbal information, it was not surprising that its role would be more salient at the encoding stage.

Given that previous studies have observed the predictive role of language and cognitive abilities at the encoding stage rather than the retention stage, the current study would also focus on the encoding stage. Empirical evidence showed that both language abilities and working memory, especially phonological memory, could predict children's word learning outcomes, and different skills may predict different aspects (e.g., Adlof & Patten, 2017). The current study also examined the role of phonological memory and language abilities in children's word learning performance. When selecting the language ability variable, instead of general vocabulary abilities, such as vocabulary size, the current study investigated the role of compound awareness because morphological awareness, more specifically, compound awareness, should play a crucial role in learning novel words with semantic heads. Along this line, the current study included an exploratory aim, which was to examine the relationship between word characteristics and learners' abilities, in other words, to investigate whether the skills that predicted learning performance would differ by the types of words.

### 1.3 Research questions

By teaching Mandarin-speaking children novel words with and without semantic head, the current study aimed to answer the following research questions:

Research question 1: would the effect of semantic head be observed at the encoding and retention stage of word learning?

Based on previous findings (Chen & Liu, 2015; Ma & Liang, 2019), it was hypothesized that the facilitative role of semantic head would be observed at the encoding stage, as evidenced by better performance on tasks that measured phonological (e.g., the naming task) and semantic representations (e.g., the referent identification task). As for the retention stage, it was hypothesized that the effect of semantic head would not be observed (e.g., James et al., 2019). We further hypothesized that the availability of semantic head would influence the error patterns of children's learning outcomes. To be more specific, in the naming task, learners would be more likely to make errors on both syllables for novel words without semantic head, while they would be able to produce the semantic head for words with semantic head.

Research question 2: could children's phonological memory and compound awareness predict their learning performance at the encoding stage? Would the predictive skills differ by word type?

Based on previous findings (e.g., Adlof & Patten, 2017; Gordon et al., 2022), it was hypothesized that the influence of these skills would be observed. We believe that the contribution of each skill would differ for words with and without semantic heads: compound awareness would contribute more to learning words with semantic heads, and phonological memory would contribute more to learning words without semantic heads.

## 2 Methods

### 2.1 Participants

This study obtained the ethical approval from the Institutional Review Board of The Hong Kong Polytechnic University (Reference number: HSEARS20221110003). Participants were recruited in two ways: (1) cooperating with local kindergartens to inform parents of this project, and then obtaining the parental enrollment from the kindergartens; (2) placing recruitment posters on the Internet and obtaining the enrollment directly from parents. A total of 64 children aged between 3;10 (year; month) to 7;02 were recruited from May to October 2023 in Jiangsu Province, Mainland China. The exclusion criteria for the encoding analysis were as follows: (1) withdrawal from this study (n = 4); (2) nonverbal IQ < 70 (n = 0); (3) did not complete all working memory tasks due to the lack of cooperation of the child (n = 1). In total, 59 children (31 girls) aged between 3;10 to 7;02 (year; month) were included in the encoding analysis. As for the retention analysis, two more exclusion criteria applied: (1) Time 2 retention exceeded one day (n = 1); (2) Time 3 retention exceeded 10 days (n = 5). 53 children (29 girls) aged between 3;11 to 7;02 were included in the retention analysis. Table 1-1 presents the demographic information and test scores of participants included into the final analysis.

Table 1-1 The demographic information and test scores of participants

| Domain | Encoding (n = 59) | | Retention (n = 53) | |
|---|---|---|---|---|
| | Mean (SD) | Range | Mean (SD) | Range |
| Age (months) | 63.10 (10.52) | 46 - 86 | 64.08 (10.04) | 47 - 86 |
| Nonverbal IQ [a] | 119.53 (16.41) | 76 - 149 | 119.57 (17.10) | 76 - 149 |
| Receptive vocabulary [b] | 105.02 (11.57) | 81 - 131 | 105.15 (11.15) | 81 - 127 |
| Expressive vocabulary [b] | 108.14 (12.12) | 77 - 134 | 108.00 (12.41) | 77 - 134 |
| Digit span (forward) | 10.71 (2.39) | 6 - 16 | 10.72 (2.49) | 6 - 16 |
| Digit span (backward) | 4.37 (1.82) | 0 - 10 | 4.53 (1.78) | 0 - 10 |
| Compound analogy | 10.68 (2.67) | 3 - 15 | 10.89 (2.70) | 3 - 15 |
| Compound structure | 15.20 (3.80) | 6 - 25 | 15.17 (3.98) | 6 - 25 |
| The duration between Time 1 and Time 3 (days) | - | - | 6.89 (0.47) | 6 - 8 |
| The duration between Time 1 and Time 4 (days) | - | - | 28.34(2.30) | 25 - 39 |
| Mandarin exposure [c] | 4.64 (0.61) | - | 4.60 (0.63) | - |
| | N | % | N | % |
| 20%-39% | 1 | 1.69 | 1 | 1.89 |
| 40% -59% | 1 | 1.69 | 1 | 1.89 |
| 60%-79% | 16 | 27.12 | 16 | 30.19 |
| 80%-100% | 41 | 69.49 | 35 | 66.04 |
| Maternal education [c] | 4.05 (0.90) | | 4.09 (0.90) | - |
| | N | % | N | % |
| middle school and below | 2 | 3.39 | 2 | 3.77 |
| high school | 2 | 3.39 | 1 | 1.89 |
| associate | 4 | 6.78 | 4 | 7.55 |
| bachelor | 34 | 57.63 | 29 | 54.72 |
| master and above | 17 | 28.81 | 17 | 32.08 |

a. Nonverbal IQ was measured by The Primary Test of Nonverbal Intelligence (McGhee & Ehrler, 2008) and presented as standardized scores. b. Children's receptive and expressive vocabulary were assessed by the Mandarin Oral Vocabulary Screener (Sheng et al., in development) and presented as standardized scores. c. Information from the questionnaire parents filled about the child's family background, general health, developmental history, exposure to dialects or other languages, and so on.

### 2.2 Novel word learning

The current study applied a combination of in-person and online participation to collect data. All children completed Time 1 word learning in person. Among all participants, 31% completed all retention tests in person, and the remaining participants completed at least one retention test online (mixed mode). Among those who participated in the current study in the mixed mode, 88% completed all retention tests online.

During the in-person sessions, children completed tasks in a quiet room under the examiner's instruction. Visual stimuli were presented by a 14-inch laptop, and auditory stimuli were presented through external speakers connected to the computer. All in-person sessions were recorded by a sound recorder (Aigo R2210). During the online sessions, parents or teachers helped children participate in tasks by logging into an online meeting platform using laptops or tablets, with few exceptions (6 of the children included in final analysis used mobile phones at least once to participate in the online sessions). The camera must remain on during the online assessment and the video was recorded. In addition, parents or teachers recorded the audio on the children's side and sent the recordings to the examiner after the test in most cases (79.8% of all online tests included into final analysis). Coding was based on these audios and/ or video recordings as well as the scoring sheet recorded by the examiner during tests.

#### 2.2.1 Materials

Considering the relatively young age of the participants, the number of to-be-learned words was determined to be eight, which is consistent with previous studies that investigated word learning abilities in preschoolers (e.g., Ferman, 2021; Gordon et al., 2022). Among the 8 novel words, 4 contained semantic heads (WH) and 4 did not contain semantic heads (OH).

The following steps were applied to design and choose the to-be-learned novel words (target words):

Step 1: Choose the potential semantic heads. Based on previous studies in Chinese populations (Chen & Liu, 2014; Jiang et al., 2011; Ma & Liang, 2019) and English speakers (McGregor et al., 2020), and the Chinese Communicative Development Inventory (Tardif et al., 2008), a list of 4 potential semantic heads were selected, including "羊/ yang2/ sheep, 虫/ chong2/ insect, 鸟/ niao3/ bird, and 果/ guo3/ fruit".

Step 2: Obtain the phonological features of all candidate semantic heads. The Database of Word-level Statistics (DoWLS) was used to retrieve the phonological features of all candidate semantic heads (Neergaard et al., 2022). According to Neergaard et al. (2019) the tonal complex-vowel segmented phonological segmentation neighborhoods, that is, the segmentation of a Mandarin syllable based on consonant-vowel-consonant-tone (C_V_C_T), was most compatible with the mental representation of native Mandarin adults. Therefore, the phonological neighborhood density of candidate semantic heads was retrieved under the C_V_C_T segmentation schemas.

Step 3: Choose real syllables that match the phonological features of the semantic heads. Based on the results of Step 2, a list of real syllables that share the same phonological neighborhood density (PND) with each candidate semantic head was decided. The real syllable candidates should share the same syllable structure or syllable length, and the same PND as the semantic heads, usually do not serve as semantic heads, and have no or limited meaning connection with potential unfamiliar referents. 4 triplets of real syllables were found, including "yang2/ sheep – wen1/ temperature – yuan3/ far", "chong2/ insect – nong2/ farming – gong4/ together", "niao3/ bird – piao3/ bleach – hui1/ grey" and "guo3/ fruit – duo3/ hide – huo2/ live". So far, all candidates for the second syllables of the disyllabic novel words have been decided, and the final choices for the second syllables of OH words were based on the results of Step 5.

Step 4: Choose the appropriate candidates for the first syllables of disyllabic novel words. To control for potential semantic association, the first syllables were chosen from the syllable gaps (pseudo syllables that follow the combination rules of the real Mandarin syllables, and with the existing diphone combinations within syllables, e.g., fao) from the DoWLS (Neergaard et al., 2022). These monosyllables to be selected consist of common Chinese consonants, vowels, and tones, but their combination does not exist in modern Mandarin. Based on the C_V_C_T segmentation schemas, these pseudo syllables and their phonological features were retrieved from the DoWLS. The following criteria were applied to screen the pseudo syllables: (1) initial consonants should be early acquired (acquired by 90% of children by 4 years of age), including "b, m, f, d, t, g, k, h, x, p, n, j, q" (Peng & Chen, 2020); (2) the vowels should not start with [y], which may to too hard to pronounce; (3) can be paired up with other syllables on PND and syllable length. In the end, 10 pairs of pseudo syllables were found, including "fua1 – wen1", "tia2 – tua2", "bua4 –bia4", "tia4 – tua4", "pun1 – bun1", "piou2 – tiou2", "pua4 – bia1", "mong1 – pong4", "mun4 – pong1" and "piou2 – muei2".

Step 5: Choose the disyllabic novel words. The first syllables of the disyllabic novel words came from the 10 pairs of pseudo syllables from Step 4, while the second syllables of disyllabic novel words came from the 4 triplets of real syllables from Step 3. For each semantic head condition (e.g., yang2/ sheep), a list of 60 disyllabic novel words can be generated. Native Chinese speakers recruited online rated the wordlikeness and familiarity of all 60 disyllabic novel words on a 5-point scale (1: very unlike Mandarin words/ very unfamiliar; 5: very much like Mandarin words/ very familiar) for one of the four sematic head conditions. Five fill-in words from other conditions were placed at the beginning and the end of each word list. In addition to the basic information, raters were required to answer a total of 140 questions (one question for wordlikeness and one question for familiarity for each word). If a rater selected the same option, such as 1, for more than 100 questions, it was considered as an ineffective evaluation of the materials, and his or her ratings would be excluded.

Table 1-2 The comparison of wordlikeness and familiarity between selected word pairs

| | WH words | Mean (SD) | OH words | Mean (SD) | *t* | *df* | *p* |
|---|---|---|---|---|---|---|---|
| Wordlikeness | piou2 yang2 | 2.40 (1.31) | tiou2 yuan3 | 2.60 (1.20) | -0.59 | 14 | 0.56 |
| | bua4 chong2 | 2.57 (1.12) | bia4 nong2 | 2.43 (1.12) | 0.41 | 13 | 0.69 |
| | tia2 niao3 | 2.00 (1.20) | tua2 hui1 | 2.14 (1.06) | -0.35 | 13 | 0.73 |
| | pong4 guo3 | 2.60 (0.95) | mong1 huo2 | 2.40 (1.08) | 0.56 | 14 | 0.58 |
| Familiarity | piou2 yang2 | 1.87 (0.96) | tiou2 yuan3 | 1.73 (0.85) | 0.62 | 14 | 0.55 |
| | bua4 chong2 | 2.21 (1.26) | bia4 nong2 | 2.43 (1.40) | -0.82 | 13 | 0.55 |
| | tia2 niao3 | 1.71 (0.88) | tua2 hui1 | 1.71 (1.03) | 0.00 | 13 | 1.00 |
| | pong4 guo3 | 1.60 (0.71) | mong1 huo2 | 2.13 (1.15) | -2.09 | 14 | 0.06 |

The rating results from 14-15 raters for each sematic head condition were included in the following analysis. Based on the rating results, pairs of disyllabic novel words with comparable wordlikeness and familiarity were chosen. With the aim of controlling the similarity between pairs of words, one combination of consonant and complex vowel would appear only once. For example, if the combination of "bua4 – bia4" with "chong2" was confirmed, the "pua4 – bia1" pair would be excluded.

Table 1-3 Characteristics of to-be-learned words

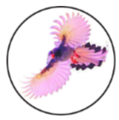
| Forms | | | | | | Referents | |
|---|---|---|---|---|---|---|---|
| Novel words with semantic head (WH words) | | | Novel words without semantic heads (OH words) | | | | |
| | | PND | | | PND | | |
| piou2<br>yang2 | piou2<br>yang2<br>*sheep* | 12<br>11 | tiou2<br>yuan3 | tiou2<br>yuan3<br>*far* | 12<br>11 |  |  |
| bua4<br>chong2 | bua4<br>chong2<br>*insect* | 12<br>11 | bia4<br>nong2 | bia4<br>nong2<br>*farming* | 12<br>11 |  |  |
| tia2<br>niao3 | tia2<br>niao3<br>*bird* | 14<br>18 | tua2<br>hui1 | tua2<br>hui1<br>*gray* | 14<br>18 |  |  |
| pong4<br>guo3 | pong4<br>guo3<br>*fruit* | 10<br>22 | mong1<br>huo2 | mong1<br>huo2<br>*live* | 10<br>22 |  |  |

Table 1-2 presents rating results of the final 4 pairs of target words. Results from paired-sample *t*-test showed no significant difference between wordlikeness and familiarity for the yang2/ sheep, chong2/ insect, and niao3/ bird pairs. As for the guo3/ fruit pair, the difference in wordlikeness was not significant, while the difference in familiarity was marginally significant. Since the overall familiarity was lower than 3, and mixed-effect regression that considered the random effect of the item would be used to analyze the results, the current study included this pair of words. A female native speaker of standard Mandarin recorded all materials in a soundproof recording booth, including sentences containing the target words for the training and testing sessions.

Table 1-3 presents the detailed characteristics of the eight target words. The pictures of unfamiliar referents were chosen from McGregor et al. (2020). Considering there are two possible correspondences between forms and referents for each semantic head condition, two versions of word learning tasks were determined, and participants were randomly assigned to either version.

### 2.2.2 Learning and testing procedures

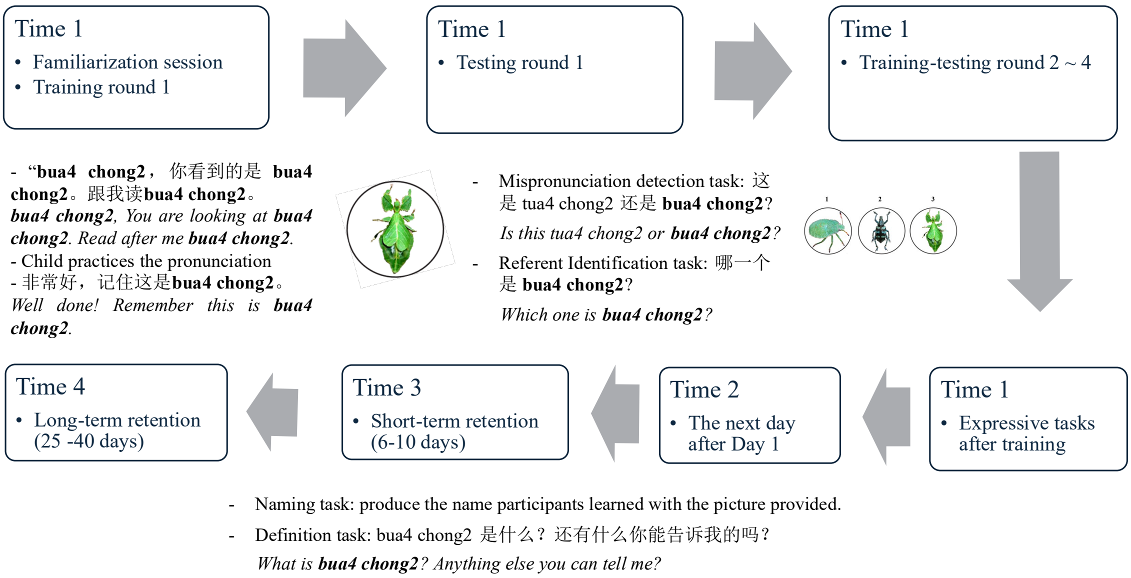



Figure 1-1 Word learning procedure

Figure 1-1 presents the word teaching-testing procedures. The whole process of word learning and testing consisted of four separate days and lasted for about one month. On the first day of learning (Time 1), participants learned eight disyllabic novel words. Before the formal teaching session, pictures of referents were presented to participants one by one. Participants were asked to label these referents to familiarize them with these unusual referents and to ensure that children did not have precise names for these referents. In most cases, children provided the category label as the name for these novel referents or gave "don't know" responses. In each teaching session, participants were exposed to each novel form four times with its corresponding referent being presented on the screen and were given the chance to practice its pronunciation. To be more specific, for each trial (word), participants firstly heard the following sentences: **bua4 chong2**，你看到的是 **bua4 chong2**。跟我读 **bua4 chong2**。[bua4 chong2, you are looking at bua4 chong2. Read after me bua4 chong2]. After participants had completed this pronunciation practice, they heard "非常好，记住这是 bua4 chong2。[Well done! Remember this is bua4 chong2.]". Then the next trial began until all eight target words were taught once. When the training round was over, the testing round began. A mispronunciation detection task and a referent identification task constituted the testing round. These two tasks were used to measure children's receptive learning outcomes and would be discussed later. After each teaching-testing round, children had a short break and were rewarded with stickers to keep their motivation. After four teaching-testing rounds, a naming task and a definition task were used to measure the overall learning performance. Before these two expressive tasks, there were at least 24 exposures to each word form for each participant. The word learning task on Day 1 lasted about 30 to 40 minutes.

The naming and definition tasks were used to measure the children's retention performance at the following expected time points: the next day (about 24 hours) of Day 1 (Time 2), one week after Day 1 (Time 3), and one-month after Day 1 (Time 4), which have been often used in previous studies. However, due to practical factors, there were variations in the intervals for Time 3 and Time 4, which were reported in Table 1-

1. The word learning tests at Time 2, Time 3 and Time 4 lasted about 5 minutes.

### 2.2.3 Learning outcomes

The mispronunciation detection task was used to measure the receptive knowledge of phonological representations. Before this task, participants were told that they would see one picture and hear two words, and they were required to listen carefully and tell the examiner which one was the right name of the referent on the picture. Then participants heard the following sentence: "Is this tua2 chong2 or bua4 chong2" with the referent and number 1 and 2 presented on the screen. The foil form (i.e., tua2 chong2) differs from the target form on the initial consonant by two phonetic features (Gordon et al., 2022) and was kept the same in each testing round. Although the number of 1 and 2 were presented on the screen to help children make the choice, some children found it difficult to give the answer one or two, and instead directly produced the form they thought to be correct (oral response). Although the examiner asked these children to choose 1 or 2 after the oral response, when confirming the response offline, we decided that if children produced the oral response spontaneously, we would take the oral response as the children's choice. If children did not produce an oral response, we would take their choice of 1 or 2 as the answer. In addition, if children indicated that they did not hear the question clearly, they would be given one opportunity to hear the two forms again.

The referent identification task was used to measure the establishment of a link between the form and referent. Three pictures (the target, the trained non-target from the same category, and an untrained distractor from the same category) were presented, and participants were asked to choose the corresponding picture by either pointing to it or say the number above picture. Children's final choice in this task was taken as the answer.

The naming task was used to measure the expressive knowledge of phonological representations. Pictures of target referents used during training were presented, and participants were asked to produce the name they learned. If children just produced the category of the target referent (e.g., yang2/ sheep), then the examiner would provide the prompt "我们刚刚学习过的名字是什么？(What is its name we just learned)?". In addition, considering that the naming task may be challenging for preschoolers, children would be given the chance to correct their response if they showed uncertainty about their response. Children's final responses were scored (Gordon et al., 2022). The same segmentation schemas (C_V_C_T) as discussed in the section 2.2.1 Materials were used to calculate the proportional accuracy for the two syllables of each target word.

The errors children made in the naming task were analyzed using the following coding scheme: (1) the position of the error: only in the first syllable, only in the second syllable, in both syllables, or no response; (2) the specific type of error at each position: substituting the target syllable with syllabled from the other target word in the same semantic category, making phonologically related errors with at least one phoneme overlapping with the target syllable, other errors that were neither substitution nor phonological errors, and no response. Table 1-4 presents the coding scheme and examples for the naming error analysis.

Table 1-4 Coding scheme and examples for the naming error analysis

| Error position- Error type | Response | Target |
|---|---|---|
| S1 [a]- Sub [c] | tiou2 yang2<br>yuan3 yang2 | piou2 yang2 |
| S1-Pho [d] | ping2 yang2 | piou2 yang2 |
| S1-Other [e] | hui1 yang2 | piou2 yang2 |
| S1-NR [f] | yang2 | piou2 yang2 |
| S2 [b]-Sub | bia4 chong2 | bia4 nong2 |
| S2-Pho | mong1 he2 | mong1 huo2 |
| S2-Other | Not attested | |
| S2-NR | Not attested | |
| S1-Sub + S2-Sub | tiou2 yuan3 | piou2 yang2 |
| S1-Pho + S2-Sub | pia4 chong2 | bia4 nong2 |
| S1-Other + S2-Sub | pao4 guo3 | mong1 huo2 |
| S1-NR + S2-Sub | yang2 | tiou2 yuan3 |
| S1-Sub + S2-Pho | Not attested | |
| S1-Pho + S2-Pho | bong1 tuo3 | pong4 guo3 |
| S1-Other + S2-Pho | bia4 hong2 | mong1 huo2 |
| S1-NR + S2-Pho | not available | |
| S1-Sub + S2-Other | not available | |
| S1-Pho + S2-Other | tuo2 lu4 | tiou2 yuan3 |
| S1-Other + S2-Other | tua2 huo1 | bua4 chong2 |
| S1-NR + S2-Other | hui1 | bia4 nong2 |
| S1-Sub + S2-NR | Not attested | |
| S1-Pho + S2-NR | tiao2 | tiou2 yuan3 |
| S1-Other + S2-NR | gua4 | tiou2 yuan3 |

Note: a. S1: the first syllable; b. S2: the second syllable; c. Sub: substitution errors; d. Pho: phonological errors; e. Other: neither substitution nor phonological errors; f. NR: no response.

The definition task was used to measure expressive semantic knowledge after training. Participants were asked to answer the question "piou2 yang2 是什么？/ What is piou2 yang2/ sheep?" without the referent being presented. If children produced the categorical information of the target word spontaneously, the examiner would ask whether there was anything else they would like to say. If children did not spontaneously produce the categorical information, the examiner would ask "这是一个什么东西？/ What is this thing?". In general, children were encouraged to produce as much information as possible.

Table 1-5 Scoring protocol for definition task

| Score | Description |
|---|---|
| 0 | - No response or I do not know<br>- Repeat the target word<br>- Did not provide any correct information |
| 1 | - Only produce the categorical information<br>OR<br>- Only produce **one type** of accurate observable features (e.g., color, or body parts of animals)<br>OR<br>- Only produced reasonably related words (e.g., real animals from the same category, or other trained novel words) |
| 2 | - Categorical information + one piece of accurate semantic information<br>OR<br>- Categorical information + reasonably related words<br>OR<br>- Produce more than one accurate physical description (e.g., color and body parts of animals)<br>OR<br>- One type of accurate physical description + reasonably related words<br>OR<br>- Other responses that do not meet the criteria of 3 points |
| 3 | Categorical information + more than one type of accurate observable features (e.g., color and body parts) |

The scoring protocol of children's definition performance was modelled on studies that examined children's definition skills for real words (Dosi & Gavriilidou, 2020; Duff, 2019). Table 1-5 presents the scoring protocol for the sematic information children produced during the task. It is important to note that the definition task was relatively challenging, as the examiner did not give practice trials or models before the test. Therefore, the scoring criteria were adjusted to be more lenient. For example, some children produced related words when asked whether they had anything else to say. The current protocol accepted these reasonably related words (e.g., other novel words that have been taught, or real words from the same category) as a piece of accurate semantic information produced by children, assuming that children knew these words are semantically connected.

### 2.3 Language and cognitive abilities measures

Tasks measuring children's phonological memory, nonverbal IQ, oral vocabulary abilities, and compound awareness were administered on different days after children completed the main word learning tasks or retention tests.

#### 2.3.1 Phonological memory

Children's phonological memory was assessed by the forward and backward digit span task from Wechsler Intelligence Scale for Children 4$^{th}$ edition (Chinese Version) (Li & Zhu, 2012) . Children were instructed to repeat a sequence of digits in the same or the reverse order after the examiner orally produced these digits (randomly chosen from 1 to 9). The forward digit span task started from two digits without the practice trials. If children succeeded in one of two trials of the two-digit sequence, the length of sequence would be increased to three digits. If children failed in both trials of, for example, six-digit sequence, the test discontinued. The general rule for the backward digit span was the same except that there were two practice trials of two-digit sequence. All children completed the forward task first, then the backward task. Considering this task was designed to test children aged above 6-year-old, during the formal test, the examiner provided more instruction on the trials of two-digit sequence when younger children showed difficulty in understanding the requirements of the task. All children completed the phonological memory test during the offline session.

Table 1-1 presents scores of these two tasks for children included in the final analysis.

A single composite phonological memory score based on the total number of correct trials of the forward digit span task (a maximum score of 20) and the backward digit span task (a maximum score of 18). For children included in the final analysis, their raw scores (the total number of correct trials) of these two tasks were firstly converted to z scores and then averaged (Pomper et al., 2022).

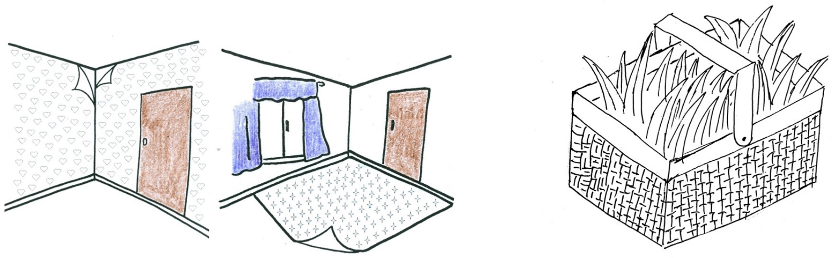
### 2.3.2 Nonverbal IQ

The Primary Test of Nonverbal Intelligence (PTONI) was used to assess children's nonverbal intelligence (McGhee & Ehrler, 2008). Children were asked to choose one picture that was not like the other pictures. All children completed the nonverbal IQ test during the offline session. The standard scores of the PTONI were used to screen participants. Children who scored lower than 70 were excluded from the final analysis.

### 2.3.3 Mandarin oral vocabulary

Children's oral vocabulary was assessed by a quick vocabulary assessment tool for Mandarin Chinese (Mandarin Oral Vocabulary Screener, Sheng et al., in development). In the receptive subtest, four pictures were presented on the screen for each word and children heard the target word once. Children were asked to point to or say the number of the picture that matched the word they had just heard. In the expressive subtest, children were asked to say the name of the picture presented to them one by one. The expressive subtest always followed the receptive subtest. Both subtests include three practice trials and sixteen testing trials. This vocabulary test has a normative sample of over 1000 children with acceptable internal consistency (Cronbach's $\alpha$= .78) for both receptive and expressive subtests. 57.6% of the children completed the vocabulary test during the offline session and the remaining children complete the vocabulary test in their first online session.

### 2.3.4 Compound awareness



a. Example for the compound analogy task b. Example for the compound structure task

Figure 1-2 Illustration of the compound analogy and structure task

Children's compound awareness was assessed by a compound analogy task and a compound structure task (Lam & Sheng, 2016; Yang et al., 2021). The compound analogy task taps into children's ability to draw analogies from existing compounds to create novel compounds. Two drawings were displayed for each trial, one showing a familiar object (i.e., the left one on Figure 1-2a) and the other showing a novel object adapted from the familiar object (i.e., the right one on Figure 1-2a). Children were instructed to give a name for the novel object. This task includes two practice trials and fifteen test trials. In the practice trials, the examiner pointed to the familiar object and played the audio "贴在墙上的纸叫墙纸。那么贴在地上的纸该叫什么?/Paper pasted on the wall is called wallpaper, then what should we call paper that is pasted on the floor? (floor paper)". Feedback would be given in the practice trials. If children produced the expected answer, the examiner would give a positive feedback and repeat the sentence with the expected name of the novel object. If children produced an unexpected answer or said, "I don't know", the examiner would give more explanation and model the answer. In the test trials, no feedback would be given. Participants were required to complete all trials, and their performance was indicated by the total number of correct trials with a maximum score of 15.

The compound structure task taps into children's awareness that the second morpheme in a compound word specifies the semantic category. Children heard the following instructions with the drawing presented (e.g. a basket filled with grass in Figure 1-2b):" 这些草被装在篮子里。我们给这些草起个名字，草篮和篮草，你看哪个更好？(篮草)/ Look at the picture. Here is some grass we put in a basket. Now, let's give a name to the grass. Which name is better: grass basket or basket grass? (basket grass)" (Chen et al., 2009). The overall procedure of the compound structure was similar to the compound analogy task, with two practice trials and thirteen test trials for each version. Feedback would be given only during the practice trials. Since there were two potentially named objects within one drawing (e.g., basket and grass) and both names were provided (e.g., grass basket for basket and basket grass for grass), two versions were used in this task. Children completed one of two versions on two separate days that were at least one week apart. The order of the two versions was counterbalanced. Participants were required to complete all trials of the two versions and their performance was indicated by the total number of correct trials with a maximum score of 26. Compound awareness tests were not given at Time 1 when novel words were taught. The first compound structure task always followed the compound analogy. 40.7% of children completed all compound awareness tests during the offline session and the remaining children completed at least one of compound awareness tests online.

Table 1-1 presents scores of these two tasks for children included into final analysis.

Referring to the method applied by Pomper et al. (2022), a single composite compound awareness score based on the total number of correct trials of the compound structure task and compound analogy task was calculated. For children included in the final analysis, their scores (the total number of correct trials) of these two tasks were firstly converted to z scores and then averaged.

## 3. Results

Analyses and visualization of results were completed by R Version 4.4.4 (R Core Team, 2024) in R studio (Posit team, 2024) using the packages of tidyverse (Wickham et al., 2019), lme4 (Bates et al., 2015), BruceR (Bao, 2023), ggplot2 (Wickham, 2016), and rcartocolor (Nowosad, 2018), and Jamovi (The jamovi project, 2024) (The Jamovi project, 2023) using the module GAMLj (Gallucci, 2023).

### 3.1 Coding and reliability

All participants' learning outcomes were coded offline using audio/video recordings and scoring sheets. 20 percent of participants' data were double coded for reliability check. Reliability was 98% for the mispronunciation detection task and 100% for the referent-identification task. The reliability of overall phoneme accuracy for naming was 95%. The scoring consistency of definition was 96%. Inconsistencies were resolved based on the audio/video recordings and incorrect codings were corrected.

### 3.2 The effect of semantic head

Mixed-effect logistic regression was used to analyze children's performance on the mispronunciation detection and referent identification task since their responses were coded as correct (1) or incorrect (0). The binomial test was used to compare children's performance in the mispronunciation detection and referent identification task to chance level. Mixed-effect linear regression was used to analyze children's naming and definition performance. Simple coding was used to contrast the variable of Type (-0.5 for OH, +0.5 for WH), and dummy coding was used to contrast the variable of Round and Time. All models contained age as the covariate. The minimal random effect structure that best supported the model fit was determined for each task. The *p*-values were corrected with Bonferroni adjustment for post-hoc comparisons.

### 3.2.1 Encoding

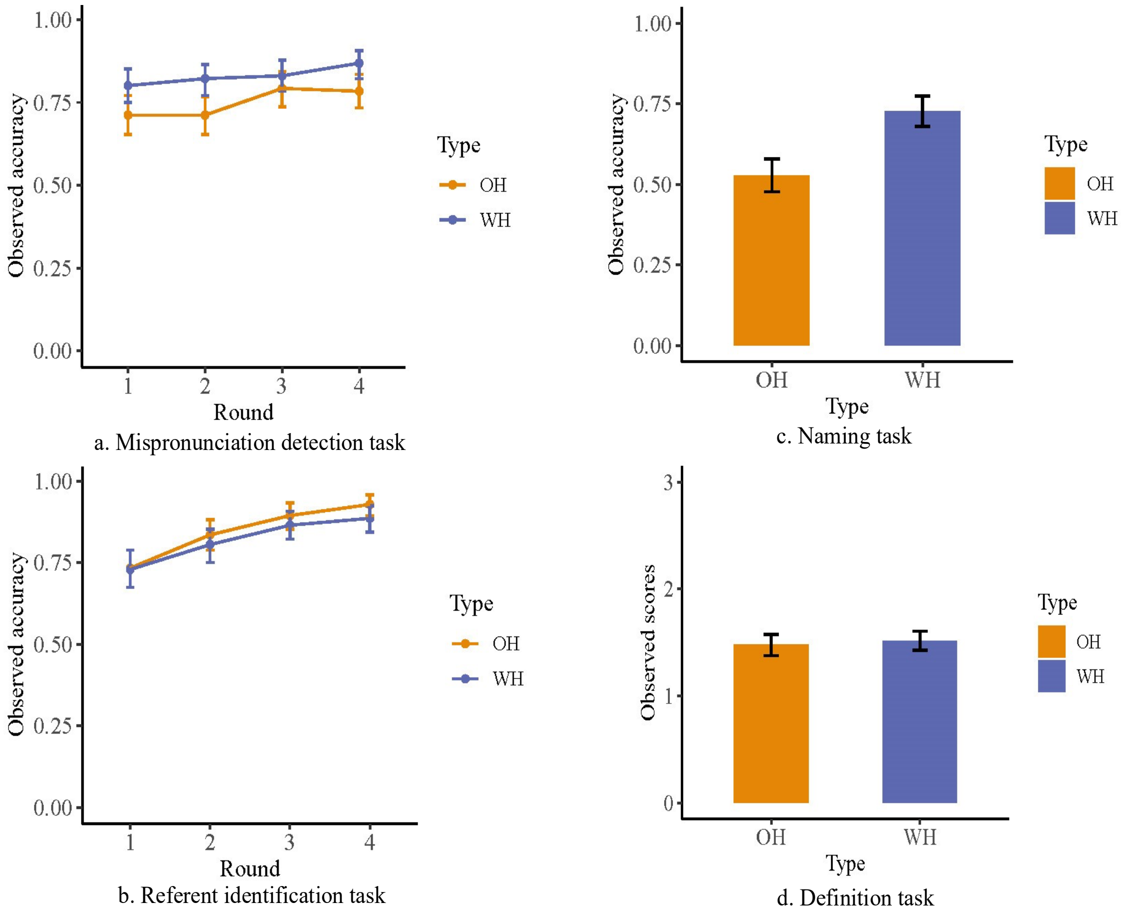



Figure 1-3 Children's learning performance at Time 1; Error bars represent 95% Confidence Intervals, the same below.

Figure 1-3 presents the observed learning performance of the four tasks at the encoding stage and Table 1-6 summarizes the statistical results. Results of the binomial test showed that children performed significantly differently from chance level (0.5 for the mispronunciation detection and 0.33 for the referent identification task) in choosing the correct forms and referents of WH and OH words (all *p*s < 0.001).

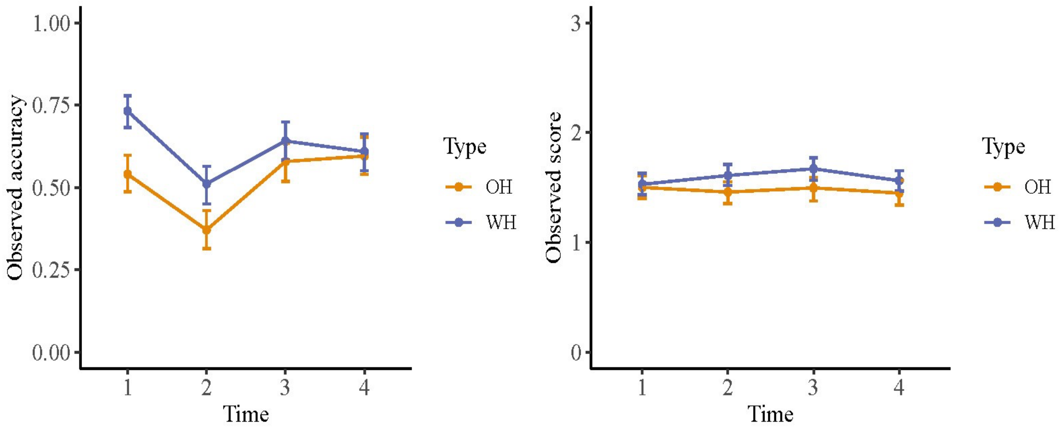
Table 1-6 Result summaries of learning performance at Time 1

| | | | | |
|---|---|---|---|---|
| | | $\chi^2$ | *df* | *p* |
| Mispronunciation detection task [a] | Type | 7.21 | 1 | 0.01 |
| | Round | 0.60 | 3 | 0.90 |
| | Age | 5.56 | 1 | 0.02 |
| | | Prob [e]. | SE [f] | 95% CI [g] |
| | WH | 0.89 | 0.02 | [0.83,0.92] |
| | OH | 0.81 | 0.03 | [0.74,0.87] |
| | | $\chi^2$ | *df* | *p* |
| Referent identification task [b] | Type | 0.60 | 1 | 0.44 |
| | Round | 68.56 | 3 | < 0.001 |
| | Age | 1.90 | 1 | 0.17 |
| | | Prob [e]. | SE [f] | 95% CI [g] |
| | Round 1 | 0.78 | 0.04 | [0.69,0.84] |
| | Round 2 | 0.86 | 0.03 | [0.80,0.91] |
| | Round 3 | 0.92 | 0.02 | [0.87,0.95] |
| | Round 4 | 0.94 | 0.01 | [0.90,0.96] |
| | | *F* | *df* | *p* |
| Naming task [c] | Type | 6.25 | [1,7.96] | 0.04 |
| | Age | 0.67 | [1,58.53] | 0.42 |
| | | Mean [e] | SE [f] | 95% CI [g] |
| | WH | 0.73 | 0.07 | [0.58,0.88] |
| | OH | 0.53 | 0.07 | [0.38,0.68] |
| | | *F* | *df* | *p* |
| Definition task [d] | Type | 0.55 | [1,413] | 0.46 |
| | Age | 3.10 | [1,59] | 0.08 |

Note: a. The model contained Type and Round as fixed effects, by-participant intercept and by-item random slope for round as random effects. b. The model contained Type and Round as fixed effects, by-participant and by-item intercept as random effects; c. The model contained Type as the fixed effect, by-participant and by-item intercept as random effects; d. the model contained Type as the fixed effect, by-participant intercept as the random effect. e. Estimated marginal means/probability would be presented for significant effects and on the second half of the model summary tables for each task; f. SE = standard error; g: CI: Confidence Interval.

Results of the mixed-effect logistic regression showed that the main effect of Type was statistically significant in the mispronunciation detection task ($\beta$ = 0.59, $z$ = 2.68, $p$ = 0.01) and the naming task ($\beta$ = 0.20, $t$ = 2.50, $p$ = 0.04), suggesting that probability of identifying the correct word forms of WH words was higher than OH words, and the phoneme accuracy of WH words was higher than OH words. The main effect of Round was statistically significant in the referent-identification task, suggesting that the probability of identifying the correct referents increased as training rounds increased.

### 3.2.2 Retention




a. Naming task b. Definition task

Figure 1-4 Children's learning outcomes across the four time points

Figure 1-4 presents the observed naming and definition performance across the four time points and Table 1-7 summarizes the statistical results. The main effect of Time and the interaction between Type and Time were statistically significant in the naming task. The post-hoc analysis showed that the effect of Type (OH-WH) was not observed at Time 2 ($t$ =-1.31, $p$ = .21), Time 3 ($t$ = -.59, $p$ = .56), and Time 4 ($t$ =-.13, $p$ = .90). The accuracy of children's phonological representation of WH words at Time 1 was better than Time 2 ($t$ = 6.42, $p$ < .001), and the performance at Time 2 was worse than Time 3 ($t$ = -3.80, $p$ = .001). No significant difference was found between Time 3 and

Time 4 ($t$ = .98, $p$ = 1.00). Significant differences were found between Time 1 and Time 3 ($t$ = 2.61, $p$ = .05), between Time 1 and Time 4 ($t$ = 3.59, $p$ = .002), and between Time 2 and Time 4 ($t$ = -2.83, $p$ = .03).

The accuracy of children's phonological representation of OH words at Time 1 was better than Time 2 ($t$ = 4.94, $p$ < .001), and the performance at Time 2 was worse than Time 3 ($t$ = -6.06, $p$ < .001). No significant difference was found between Time 3 and Time 4 ($t$ = -.49, $p$ = 1.00). In addition, no significant difference was found between Time 1 and Time 3 ($t$ = -1.12, $p$ = 1.00), Time 1 and Time 4 ($t$ = -1.61, $p$ = .65), while the performance at Time 2 was worse than Time 4 ($t$ = -6.54, $p$ < .001).

Table 1-7 Result summaries of learning performance across the four time points

| | | *F* | *df* | *p* |
|---|---|---|---|---|
| Naming task [a] | Type | 1.27 | [1,8.28] | 0.29 |
| | Time | 26.80 | [3,1583.88] | < 0.001 |
| | Type * Time | 5.36 | [3,1583.88] | 0.001 |
| | Age | 2.50 | [1,52.44] | 0.12 |
| | | Mean [c] | SE [d] | 95% CI [e] |
| | WH-Time 1 | 0.73 | 0.08 | [0.56,0.90] |
| | WH-Time 2 | 0.51 | 0.08 | [0.34,0.68] |
| | WH-Time 3 | 0.64 | 0.08 | [0.47,0.81] |
| | WH-Time 4 | 0.61 | 0.08 | [0.44,0.79] |
| | OH-Time 1 | 0.54 | 0.08 | [0.37,0.71] |
| | OH-Time 2 | 0.37 | 0.08 | [0.20,0.54] |
| | OH-Time 3 | 0.58 | 0.08 | [0.40,0.75] |
| | OH-Time 4 | 0.60 | 0.08 | [0.42,0.77] |
| | | *F* | *df* | *p* |
| Definition task [b] | Type | 4.90 | [1,7.31] | 0.06 |
| | Time | 1.55 | [3,1635.90] | 0.20 |
| | Type * Time | 1.33 | [3,1635.90] | 0.27 |
| | Age | 13.46 | [1,52.98] | < 0.001 |

Note: a. The model contained Type, Time and their interaction as the fixed effect, by-participant random slope for type and by-item intercept as random effects; b. The model contained Type, Time and their interaction as the fixed effect, by-participant and by-item intercept as random effects; c. Estimated marginal means/probability would be presented for significant effects and on the second half of the model summary tables for each task; d. SE = standard error; e: CI: Confidence Interval.

### 3.2.3 Naming error analysis

Table 1-8 Summaries of naming error analysis

| Error position | Error type | WH | OH |
|---|---|---|---|
| S1 [a] | Sub [c] | 5.69% | 2.90% |
| | Pho [d] | 20.12% | 7.43% |
| | Other [e] | 7.72% | 0.54% |
| | NR [f] | 9.35% | 3.80% |
| | **SUM** | **42.89%** | **14.67%** |
| S2 [b] | Sub | 1.02% | 7.43% |
| | Pho | 0.00% | 0.91% |
| | Other | 0.00% | 0.00% |
| | NR | 0.00% | 0.00% |
| | **SUM** | **1.02%** | **8.33%** |
| Both syllables | S1-Sub + S2-Sub | 7.93% | 11.23% |
| | S1-Pho + S2-Sub | 1.02% | 4.35% |
| | S1-Other + S2-Sub | 0.41% | 3.26% |
| | S1-NR + S2-Sub | 0.00% | 2.72% |
| | S1-Sub + S2-Pho | 0.00% | 0.00% |
| | S1-Pho + S2-Pho | 0.00% | 0.36% |
| | S1-Other + S2-Pho | 0.00% | 0.36% |
| | S1-NR + S2-Pho | 0.00% | 0.00% |
| | S1-Sub + S2-Other | 0.00% | 0.00% |
| | S1-Pho + S2-Other | 0.20% | 1.09% |
| | S1-Other + S2-Other | 2.44% | 5.25% |
| | S1-NR + S2-Other | 0.20% | 1.27% |
| | S1-Sub + S2-NR | 0.00% | 0.00% |
| | S1-Pho + S2-NR | 1.22% | 0.36% |
| | S1-Other + S2-NR | 0.00% | 0.72% |
| | **SUM** | **13.41%** | **30.98%** |
| No response | | **42.68%** | **46.01%** |

Note: a. S1: the first syllable; b. S2: the second syllable; c. Sub: substitution errors; d. Pho: phonological errors; e. Other: neither substitution nor phonological errors; f. NR: no response.

Table 1-8 presents the percentage of each type of error for WH and OH words. The analysis of naming errors was based on the participants who were included into the retention analysis and responses were pooled together across time points. The most common errors of WH words were errors only in the first syllable and no response, followed by errors in both syllables, and errors only in the second syllable. As for OH words, the most common errors were no response, followed by errors in both syllables, errors only in the first syllable and errors only in the second syllable. After excluding the overall no response, the main error type of WH words was errors only in the first syllable (74.82%), while the main error type of OH words was errors in both syllables (57.38%). This general pattern suggested that when children came across difficulties in retrieving the phonological representations of newly learned words, they could still produce the correct second syllables (i.e., the semantic head) of WH words in most cases but may face challenges in producing both syllables of OH words. Furthermore, after excluding no response, over half of errors (53.69%) in OH words were the substitution of the second syllables, suggesting that children preferred to name target referents with the semantic head.

### 3.3 Learners' abilities

To investigate whether phonological memory and compound awareness could predict children's learning outcomes and whether they contributed differently to words with and without semantic head, this study referred to the method applied by Hartley et al. (2020), which involved a series of model comparisons. The comparisons started from comparing a model containing either phonological memory or compound awareness with the baseline model only containing the random effect. If a significant improvement of the model fit was observed, further comparisons would be conducted with two abilities to determine which predictors contribute more to the model fitness. Table 1-9 summarizes the results of all model comparisons in the four tasks at the encoding stage. For WH words, we found that phonological memory alone could significantly predict children's performance on the mispronunciation detection task and the referent-identification task. Compound awareness alone could significantly predict their performance on all four tasks. In the mispronunciation detection task, adding both phonological memory and compound awareness did not significantly improve the model fit compared to the phonological memory only condition, while it showed a marginal significant improvement over the compound awareness only condition, suggesting that the role of phonological memory was more salient. Similarly, the role of compound awareness was more salient in the referent identification task.

Table 1-9 The role of phonological memory and compound awareness at encoding stage

| | | WH | | OH | |
|---|---|---|---|---|---|
| | | $\chi^2(1)$ | *p* | $\chi^2(1)$ | *p* |
| Mispronunciation detection task | PM [a] vs Baseline [d] | 8.09 | 0.004 | 4.07 | 0.04 |
| | CA [b] vs Baseline [d] | 6.74 | 0.009 | 2.30 | 0.13 |
| | PM [a] vs PM & CA [c] | 2.29 | 0.13 | N. A [h] | |
| | CA [b] vs PM & CA | 3.64 | 0.056 | | |
| Referent identification task | PM [a] vs Baseline [e] | 7.76 | 0.005 | 5.65 | 0.02 |
| | CA [b] vs Baseline [e] | 10.25 | 0.001 | 4.22 | 0.04 |
| | PM [a] vs PM & CA | 4.85 | 0.03 | 1.24 | 0.27 |
| | CA [b] vs PM & CA | 2.36 | 0.12 | 2.67 | 0.10 |
| Naming task | PM [a] vs Baseline [f] | 0.97 | 0.32 | 3.18 | 0.07 |
| | CA [b] vs Baseline [f] | 3.68 | 0.055 | 4.49 | 0.03 |
| Definition task | PM [a] vs Baseline [g] | 2.94 | 0.09 | 5.88 | 0.02 |
| | CA [b] vs Baseline [g] | 5.63 | 0.02 | 4.78 | 0.03 |
| | PM [a] vs PM & CA [g] | N. A [h] | | 1.82 | 0.18 |
| | CA [b] vs PM & CA [g] | | | 2.92 | 0.09 |

Note: a. PM: model contains phonological memory; b. CA: model contain compound awareness; c. PM & CA: model contain both phonological memory and compound awareness; d. The baseline model of the mispronunciation detection task contained the by-participant intercept as the random effect; e. The baseline model of the referent identification contained the by-participant and by-item intercept as random effects; f. The baseline model of the naming task contained the by-participant and by-item intercept as random effects; g. The baseline model of the definition task contained the by-participant intercept as the random effect; h. N.A: not applicable.

For OH words, phonological memory alone could significantly predict children's performance on the mispronunciation detection task, referent-identification task and definition task. Compound awareness alone could significantly predict their performance on the referent-identification task, naming task and definition task. In both referent-identification task and definition task, adding both phonological memory and compound awareness did not significantly improve the model fit compared to phonological memory only condition and the compound awareness only condition, suggesting that the contributions of phonological memory and compound awareness were similar.

## 4. Discussion

### 4.1 The effect of semantic head on word learning

To answer the first research question, that is, would the presence of semantic head influence children's learning performance at the encoding and retention stage, the current study firstly analyzed children's learning performance at Time 1. Among the four tasks children completed at Time 1, children performed better in identifying the correct word forms of WH words than OH words (the mispronunciation detection task) and the accuracy of overall phonological representations of WH words was higher than OH words (the naming task). These findings suggested that the presence of semantic head in the novel words has contributed to better phonological representations in children, which was consistent with our hypothesis and previous findings (Chen & Liu, 2014; Ferman, 2021; Ma & Liang, 2019).

It was unexpected that the advantage of learning WH words was not observed in the referent-identification task. This may be attributed to the current design of the learning procedure and the task requirement. The order of tasks may have an influence. Chen and Liu (2014) only used the referent identification task at phase 1 (after an incidental word learning task). The referent identification task was the first task in Ma and Liang (2018) and Ferman (2021). In the current study, the referent identification task always followed the mispronunciation detection task. Although there was no re-exposure of one-to-one form-to-referent associations in the mispronunciation detection task, the retrieval of phonological knowledge with referents provided may have inadvertently consolidated the associations. In addition, compared to previous studies, the design of distractors in our study was different, which may influence the complexity of the task. Chen and Liu (2014) and Ferman (2021) used three distractors from different categories from those of the target words, while the current study used two distractors from the same category of the target words, which could have made this task harder. Taking these differences in experimental design into consideration, it was not surprising that the word type effect was not observed in the referent-identification task in the current study. Despite the lack of word type effect, children performed well on this task, suggesting that they could establish the links between forms and referents of a majority of the novel words.

The definition task addressed the semantic representations of novel words, where the effect of word type was not observed, either. Figure 1-5 presents the distribution of scores in the definition task at Time 1. In general, children scored 1 or 2 points on this task, suggesting that they could provide one or two pieces of correct semantic information. There were no obvious differences between the two types of words in scoring 1, 2 and 3 points, while there were more 0s for OH than WH words. This subtle difference suggested that when asked to verbally describe referents with only forms provided, children were more likely to come across difficulties in retrieving the correct information of target referents for OH words and even may not be able to provide the correct categorical information. These difficulties could be attributed to the limited semantic association between forms and referents in OH words.

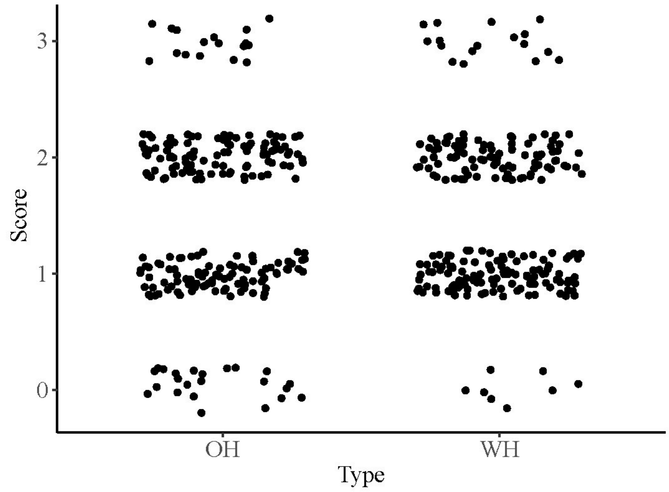



Figure 1-5 Distribution of definition scores at Time 1

In addition, the free definition task is challenging for children within this age range (mainly preschoolers). Previous studies that examined definitional skills of real words showed that children aged older than 8 years can produce more abstract and formal definitions that include at least the superordinate and one feature of the targe word (Dosi & Gavriilidou, 2020), meeting the criteria for 2 points in the current study. Although Gavriilidou et al. (2022) found better definitional performance on compound nouns than derived nouns in Greek-speaking participants (n = 322, age range = 5 years to 68.4 years), the current study did not find an obvious advantage of WH words in the definition task, which may relate to children's under-developed definitional skills.

To summarize the findings so far, the facilitative effect of semantic head was observed on forming phonological representations of newly learned words at the encoding stage. Why would the availability of semantic head be facilitative? Previous studies have argued that these characteristics provided supportive cues (e.g., the schematic structure) to reduce the load of initial learning (Ferman, 2021; James et al., 2019). Taking a step further, in the current study, the facilitative effect of semantic head observed on forming the phonological representations may stem from the lower cognitive load of WH words than OH words. According to the Cognitive Load theory (Sweller et al., 2011), when learners initially process to-be-learned information, this process would occupy the resources of working memory. The load of working memory is related to the characteristics of to-be-learned information (intrinsic cognitive load) and how the information would be taught (extraneous cognitive load) (Sweller et al., 2011). WH and OH words were taught in the same way, suggesting the extraneous cognitive load was the same for the two types of words. As for the intrinsic cognitive load, the processing load of WH words is likely to be lower than OH words. Recalling that many characteristics (e.g., phonological neighborhood density and familiarity) were controlled when designing these novel words, the main difference between WH words and OH words should be that the associations between forms and referents were closer in WH words than in OH words. This explicit cue may alleviate processing demands of WH words, allowing children to allocate more resources to process the unfamiliar parts of the to-be-learned words (e.g., the first syllable) and leading to better phonological representations of the whole words. On the contrary, without the semantic head, children had to encode two unfamiliar and unrelated syllables and map them with the referents, resulting in the higher processing demands of OH words than WH words within the same limited learning period.

Further evidence supporting the aforementioned argument came from the naming error analysis. Consistent with the hypothesis, the availability of semantic head also has influenced children's errors in the naming task. The most common errors in WH words were errors only in the first syllable and the most common errors in OH words were errors in both syllables. These patterns suggested that when children failed to retrieve the full phonological representations of WH words, they could at least produce the semantic head and make mistakes mainly on the first syllables (i.e., pseudo syllables in Mandarin). When children failed to retrieve the correct phonological representations of OH words, children would make mistakes on both syllables. Furthermore, we observed that the common of substitution error in the second syllable of OH words (58.69% after excluding no response), suggesting that children preferred to name target referents with the semantic head.

To investigate whether the role of semantic head would be observed at the retention stage, the current study compared the naming and definition performance of two types of words across four time points. For naming, although higher phoneme accuracy of WH words was found at Time 1, the effect of Type was not observed in one-day retention (Time 2), short-term retention (Time 3) and long-term retention (Time 4). These findings suggested that the initial advantage of WH words in forming the phonological representations was not maintained into the retention stage. The absence of semantic head facilitation in the retention stage was consistent with (James et al., 2019), in which the advantage of more orthographic neighbors was not observed in the one-day and one-week retention. So why would the effect of word characteristics not persist into the retention stage? An examination of the temporal change patterns of the two types of words indicated an overall similarity, that is, a significant decrease from Time 1 to Time 2, followed by a significant increase from Time 2 to Time 3, and no difference between Time 3 and Time 4. Nevertheless, differences did exist. Figure 1-4a demonstrates that the difference in phoneme accuracy between the two types of words gradually diminished, especially from Time 2 to Time 3. Moreover, there were no differences in phoneme accuracy for OH words at Time 3 and Time 4 compared to Time 1, suggesting that children maintained the phonological representations of OH words at later retention points. On the contrary, significant differences between Time 1 and Time 3 and between Time 1 and Time 4 were observed for WH word. The decreased phoneme accuracy at Time 3 and Time 4 (compared to Time 1) suggested that children failed to recover some of the phonological representations of WH words learned at Time 1 at later retention points. Although not statistically significant, we can observe from Figure 1-4a that there was a clear trend of WH advantage at Time 2. Therefore, the absence of a semantic head advantage during longer-term retention may relate to the smaller recovery at Times 3 and 4 from the significant dip at Time 2 for WH words.

What happened from Time 2 to Time 3 that may have led to this difference? It is possible that the two types of words may experience different sleep-based consolidation (James et al., 2019). If no additional exposure is provided, the retention of new words mainly relies on the sleep-related consolidation process (Davis & Gaskell, 2009). According to the complementary learning account, the new lexical representations would be re-activated during sleep, interact with existing lexicon, and be transferred from the hippocampus to the neocortex and subcortical structures (Davis & Gaskell, 2009). Words with no neighbors have limited association with the existing lexicon and would experience the shift from the hippocampus to subcortical structure (James et al., 2019). In contrast, for words with many neighbors, there are two possible routes: these words are speedily encoded into the neocortical structure and do not rely on sleep-based consolidation; alternatively, because these words have more neighbors, their consolidation process is interfered by the existing lexicon (James et al., 2019). In terms of the current findings, we may speculate that the OH words were more likely to be transferred from the hippocampus to subcortical structure in the one-week retention period, whereas the consolidation of WH words was interfered by the existing lexicon during the same period.

In regard to the retention of the semantic representations, there was a trend indicating that children scored higher on WH words than OH words, although this difference was not statistically significant ($p$ = .06). Although there were no references to previous studies, it was not unexpected that the effect of semantic head was not observed at the retention stage, especially given the challenging nature of the definition task for most preschoolers. Furthermore, compared to the retention pattern of phonological representations, children's definition performance across the four time points was quite stable, suggesting that children did retain a few pieces of semantic information for both

WH and OH words. James et al. (2023) also found that the retention pattern of phonological knowledge differed from that of semantic knowledge. On their cued naming task, school-aged children performed significantly better at one-day delay (compared to immediately after training) and one-week delay (compared to the two previous time points). As for their cued definition task, no significant differences were found among the three time points. These findings and our current findings suggested that new phonological representations and semantic representations may differ in the trajectories of retention.

As a brief answer to the first research question, Study 1 found that the facilitative effect of semantic head was observed at the encoding stage but was absent at the retention stage. In other words, the phonological representations of OH words encoded by children were maintained after one-week delay, whereas children showed some evidence of forgetting WH words. Shifting the focus from semantic head and interpreting these findings in the context of OH words, we believe that children may have encountered a level of "desirable difficulty" in learning OH words. Learners may encounter "desirable difficulty" under certain learning conditions, where the greater the challenges they face during the initial stage of learning (e.g., the encoding stage of word learning) and the more effort they make, the better the retention and generalization of the successfully encoded knowledge will be (BjorK & Kroll, 2015). For example, Schneider et al. (2002) trained native English-speaking adults to learn French-English word pairs under different learning conditions, including the grouping methods of training materials (i.e., blocked by category or mixed together). Grouping word pairs under the same semantic category would reduce the contextual interference and create an optimal learning condition (Schneider et al. (2002). An initial advantage was observed in the categorical grouping condition, where word pairs from the same semantic category (e.g., body parts) were trained together, while the better long-term retention was observed in the mixed grouping condition, where word pairs from different semantic categories (e.g., body parts, transportation, food and so on) were trained together.

When it comes to the current study, the differences in the intrinsic cognitive load between two types of words has been discussed in the previous section. The processing demands of OH words is higher than WH words at the encoding stage, as OH words do not include any morphemes that provide cues about the semantic characteristics of the referents. The higher processing demand implies that learners need to devote more cognitive effort and perform in-depth processing of target words (Pyke et al., 2024). Higher processing demands initially put children at a disadvantage when learning OH words. However, the deeper processing enabled children to form more stable and independent representations of OH words, as evidenced by the maintained phonological representations of OH words after one-week and one-month delay. To sum up, the availability of semantic head has alleviated processing demands of WH words at the encoding stage, resulting in an initial advantage. In contrast, the absence of semantic head in OH words has led to the increased cognitive effort and deeper processing, resulting in more stable retention.

## 4.2 The predictive role of phonological memory and compound awareness

To answer Research Question 3, that is, whether phonological memory and compound awareness would predict children's learning outcomes of WH words and OH words at the encoding stage, and whether differences in the predictive skills would be observed by word type, children's learning outcomes of WH words and OH words on each task were analyzed separately. For WH words, phonological memory and compound awareness predicted children's performance in identifying the correct word forms and referents. Compound awareness could also predict children's naming and definition performance. Furthermore, the role of phonological memory was more salient in identifying the correct word forms and the role of compound awareness was more salient in identifying the correct referents. The predictive role of compound awareness was worth-noting as it was identified in all tasks. This was consistent with our hypothesis and suggested that children's compound awareness would be implicitly activated when asked to retrieve the lexical representations of these new words.

For OH words, phonological memory predicted children's performance on word form identification, referent identification, and definition. Compound awareness predicted children's performance on referent identification, naming, and definition. With weaker association between forms and referents, it was hypothesized that children would rely more on general cognitive abilities, such as phonological memory to form representations of OH words at the encoding stage. The results showed that overall, phonological memory and compound awareness played largely equal roles in predicting OH words' learning outcomes. We speculated that due to the dominance of compound words in Mandarin, children may unconsciously analyze the lexical structures when learning new words. It is possible that this analysis may result in the awareness that the novel OH words do not contain any semantic heads, thus requiring more involvement of phonological memory to sub-vocally rehearse and store the relevant phonological and semantic information and commit it to memory.

**4.3 The absence of overnight improvement in Mandarin-speaking preschoolers**

Shifting the focus away from the effect of semantic head, the current study found a significant decrease in phoneme accuracy (naming task) for both types of words at Time 2, which was inconsistent with previous findings. Henderson et al. (2012) found that compared to just after learning, 7-12 years old children showed significant improvement in the cued-recall test of word forms after 24 hours. Lucas and Norbury (2014) also found a significant improvement of naming performance in a group of 7-12 years old children at one-day retention. Similarly, Fletcher et al. (2020) found an over-night improvement in 8-12 years old children in the naming task. The review study conducted by Mason et al. (2021) revealed an overnight improvement in word learning among school-aged children (older than 7 years old) across four studies. Based on these consistent previous findings (also see Henderson et al., 2014; James et al., 2019; Knowland et al., 2019; Schimke et al., 2023), an increase of phoneme accuracy at Time 2 was expected, since the overnight sleep had happened at that time point.

As discussed in the previous section (see 4.2), the mechanism underlying the overnight improvement was that newly formed representations were consolidated during sleep (Davis & Gaskell, 2009). Contrary to previous findings, the current study found the decreases of phoneme accuracy after the one-day delay, suggesting that forgetting happened one-day after training (Storkel, 2015). Storkel (2015) used the term "memory evolution" to refer to the events that would happen when memory is transferred. There are two sides of "memory evolution": consolidation or forgetting (Storkel, 2015). The evidence of forgetting mainly came from the one-week retention. Storkel et al. (2013) found that three-year old children (mean age = 3;8) and five-year-old children (mean age = 5;4) showed a significant decrease in the naming performance one-week after training. Worse performance on the referent identification task one-week after training was found in 3 to 6 years old children (Vlach & Sandhofer, 2014; Wojcik, 2021). For one-day retention, Spanò et al. (2018) found a group of children (mean age = 33.38 months) did not show significant improvement at 24-hour delay (wake group). Rothwell et al. (2023) found that TD children (mean age = 52.31 months) did not show the overnight improvement as well.

Why did some studies observe the overnight improvement while others observe no change or even forgetting? Age may be the answer. In studies that found forgetting or the absence of overnight improvement, including the current study, participants were mainly preschoolers (aged between 3 to 7 years old), whereas in studies that found the overnight improvement, participants were school-aged (aged above 7 years). Is it possible, then, that the effect of overnight sleep on consolidating newly learned words is limited for preschoolers? It is possible because changes in general sleep habits and nighttime sleep patterns occur from preschool to school age. In term of general sleep habits, preschoolers would have more regular and longer daytime naps, whereas daytime naps would be rare for school-aged children (Kang et al., 2022). Additionally, significant changes happened in children's nighttime sleep patterns from preschool to school-age (Rothwell et al., 2023). Montgomery-Downs et al. (2006) analyzed overnight polysomnographic measures of over 500 children aged from 3 to 8 years and found significant differences in central apneas, sleep cyclicity, body position and periodic leg movements between children aged 3 to 5 years and children aged older than 6 years.

For preschoolers, daytime napping or sleep immediately after learning is more beneficial for the consolidation of newly learned words (Mason et al., 2021). Horváth et al. (2015) found that for 16-month-olds infants, sleep (i.e., daytime napping) after training was beneficial for learning outcomes measured at a delayed time point (e.g., 2-3 hours after training). Gómez and Edgin (2015) also summarized that for children aged 18-24 months and older, daytime napping benefited children's retention of learning outcomes. Therefore, although the facilitative role of sleep on retaining newly learned words was observed in both preschool and school-aged children, the optimal timing of sleep (daytime napping or overnight sleep) that would facilitate consolidation may change as children grow up. However, these arguments should be treated with caution, as empirical evidence that directly compares one-day retention between preschoolers and school-aged children is needed.

**5. Conclusion**

Mandarin-speaking children were taught novel words with and without semantic heads. The presence of sematic head may have alleviated the load of forming the phonological representations of WH words for children at the encoding stage. The salience of semantic head was further highlighted by the analysis of naming errors, as although children showed difficulty in retrieving the full phonological representations, they were able to produce the semantic head. The effect of semantic head was not observed at the retention stage for both phonological and semantic representations. The lower processing demands of WH words provided an initial advantage during encoding, while the higher processing demands of OH words led to increased cognitive effort and deeper processing, resulting in more stable retention. Phonological memory and compound awareness predicted children's learning performance at the encoding stage. The role of compound awareness was more salient in predicting children's learning of WH words, whereas phonological memory and compound awareness were equally important in predicting children's learning of OH words. Moreover, the current study found that instead of overnight improvement, children showed significant forgetting in the naming task at one-day delay, which may relate to the different sleep patterns of preschoolers.

# Chapter 3 Word learning abilities in Mandarin-speaking children with Autism Spectrum Disorder: The Role of Semantic Head

## 1 Introduction

Closely related to deficit in communicative interaction, children with Autism Spectrum Disorder (ASD) often manifest language impairments as one of the earliest symptoms (Naigles, 2013; Sukenik & Tuller, 2023). As a spectrum disorder, the significant heterogeneity within the autistic population is well-known, with 25-30% of children with ASD having no or minimal language abilities (Anderson et al., 2007), and 45% of individuals with ASD have substantial language impairments (Charman et al., 2011). The current study focused on verbal children with ASD, that is, those who possess a certain level of language proficiency. Delayed vocabulary development has been documented in many autistic children and has inspired researchers to investigate their word learning abilities (Arunachalam & Luyster, 2016; Hou & Su, 2022).

### 1.1 Word learning profiles of children with ASD

The process of learning oral vocabulary is multifaceted and entails diverse components. According to the functional framework of word learning proposed by Gupta and Tisdale (2009), this process involves: learning the phonological sequence that makes up the word form, learning the physical and functional features of the referent (McGregor et al., 2007), and establishing the link between the form and referent. Learning outcomes measured just after learning may not adequately reflect learners' abilities to learn words (McGregor et al., 2017), as newly formed lexical representations (at least including phonological and semantic representations, as well as the link between them) (Storkel

& Hoover, 2011) should be retained in memory for later retrieval. Deficits in word learning may manifest as a failure to create an initial lexical representation at the encoding stage, a failure to retain the new lexical representation in long-term memory, or both (Gordon et al., 2021).

The available evidence generally supported that verbal children with ASD demonstrated the ability to use typical mechanisms or cues to establish the form-referent links at the encoding stage, as evidenced by their success in the referent identification task. Both preschool-aged (Carter & Hartley, 2021) and school-aged (Hartley et al., 2019) children with ASD showed the ability to apply the mutual exclusivity principle to establish the form-referent links. Hartley et al. (2020) examined cross-situational word learning in a group of school-aged children with ASD (mean age = 8.79 years) and no group differences were found in the referent identification task. Furthermore, although children with ASD have well-documented impairments in gaze following and joint attention (Arunachalam & Luyster, 2016), contrary to intuition, recent evidence has suggested that preschoolers with ASD (Bean Ellawadi & McGregor, 2016; Hani et al., 2013; Luyster & Lord, 2009) and school-aged children with ASD (Bang & Nadig, 2020; Field et al., 2019; Jing et al., 2014; Jing et al., 2013) can attend to social cues, such as gaze and pointing, to establish the form-referent links. Additional empirical findings (Gladfelter & Barron, 2020; Lucas & Norbury, 2014; Lucas et al., 2017; Walton & Ingersoll, 2013) have further suggested that verbal children with ASD can establish the form-referent links at the encoding stage, despite their deficits in vocabulary development.

Compared to the wealth of evidence on form-referent mapping, our understanding of the phonological and semantic representations that verbal children with ASD formed during the encoding stage was limited. To the best of our knowledge, four studies have examined phonological representations of new words and found that both preschool-aged (Walton & Ingersoll, 2013) and school-aged children with ASD (Gladfelter & Goffman, 2018; Lucas et al., 2017) did not perform significantly different from TD controls (mean age = 11.29 years) in the naming task. Additionally, Lucas and Norbury (2014) found that their children with ASD (mean age = 10.57 years) performed better than the TD group in the naming task. With respect to the semantic representations, the available findings were all from school-aged children with ASD and were inconsistent. Lucas et al. (2017) found that the ALI (autism language impaired) group (mean age = 11.23 years) produced significantly less accurate semantic features. In contrast, Bang and Nadig (2020) found that children with ASD (mean age = 8.83 years) performed similarly to their TD peers on the definition task. Gladfelter and Goffman (2018) also found that school-aged children with ASD and TD children did not differ in the number of semantic features provided on a definition task. However, a more detailed analysis of the same dataset revealed that children with ASD tended to produce more global semantic features (such as categorical properties) than TD controls (Gladfelter & Barron, 2020). To summarize, existing studies have suggested that forming phonological representations of new words was not a challenge for verbal children with ASD, whereas forming semantic representations requires further investigation.

Encoding is the first step. The next question to be addressed was whether children with

ASD differed from TD children in retaining the newly learned words. Three studies found no group differences between children with ASD and TD controls in five-minute retention of newly established form-referent links (Carter & Hartley, 2021; Hartley et al., 2019, 2020). For the one-day retention, Lucas and Norbury (2014) did not find significant differences between the two groups in the referent identification task and naming task, either. Bang and Nadig (2020) found no group differences in the one-week retention of semantic knowledge, with both TD children and children with ASD showing a significant decrease from the encoding to retention phase. Norbury et al. (2010) found different retention patten between school-age children with ASD and their TD peers, that is, the TD group performed better on the naming and definition task one month after training than just after training, whereas children with ASD showed the reverse pattern. Fletcher et al. (2020) did not find group differences on measures related to phonological representations but did find that children with ASD (mean age = 125.55 months) produced significantly fewer semantic features than their TD peers one month after training. Based on available evidence, the long-term retention (i.e., over a month) of phonological and semantic representations may be problematic for verbal children with ASD.

Overall, by reviewing existing findings, verbal children with ASD do not have significant deficits in many components of word learning. This begs the question why they still show delays in vocabulary development. One potential answer to this question is that the specific areas of weakness in word learning have not been identified in some verbal children with ASD. Most previous studies only used one or two tasks (e.g., referent identification) to measure learning performance at one time point (e.g., just after learning) (Wojcik, 2013), resulting in a limited understanding of the strength of phonological and semantic representations in children with ASD at the encoding stage. It is possible that there was a group of children with ASD who were able to establish the form-referent links but had difficulty in forming robust phonological and/or semantic representations. Insufficient investigation of retention performance is another gap in this field (Arunachalam & Luyster, 2016; Hou & Su, 2022). We must be cautious about making the conclusion that verbal children with ASD did not have retention deficits, because as compared to the abundance of studies on encoding, few studies examined retention beyond one day. A detailed description of the retention performance at different time points, especially longer-term retention points, is critical for this population. To address these research gaps, it is critical to use multiple tasks across several time points in a single study to measure the word learning performance in verbal children with ASD and gain a clearer understanding of their strengths and weaknesses (Hou & Su, 2022).

### 1.2 Word characteristics and word learning in children with ASD

Another potential answer that may explain the delayed vocabulary development in verbal children with ASD is that they may be less sensitive to some influencing factors of word learning, such as the characteristics of to-be-learned words. Overall, the role of word characteristics remains to be explored in children with ASD. To the best of our knowledge, two published studies have investigated how word characteristics affect learning outcomes in verbal children with ASD. Ferman (2021) examined the effect of a morphological pattern that marks the categorical information of words in Hebrew-speaking children with ASD (mean age = 6;9). The TD controls matched the ASD group on age, nonverbal IQ, vocabulary, and grammar abilities. Participants learned two types of novel words, one type consisting of existing Hebrew morphological patterns, and the other type consisting of pseudo-Hebrew morphological patterns. A referent identification task and a naming task were administered immediately after training. Consistent with most previous studies, this group of children with ASD did not differ from their TD peers in establishing the form-referent links and forming the phonological representations. The facilitative role of existing morphological patterns was observed in children with and without ASD in both tasks, but with a smaller facilitative effect in naming for children with ASD (the gap in the naming accuracy for two types of words was smaller), suggesting that children with ASD were not as sensitive to this facilitative cue as their TD peers.

This reduced sensitivity to facilitative cues from word characteristics has been reported in another group of verbal children with ASD. Based on the finding that it was easier for the TD population to learn novel polysemous words (words with multiple related meanings) than to learn homonymous words (words with multiple unrelated meanings), Floyd et al. (2021) investigated this "polysemy over homonymy advantage" in school-aged children with ASD. The chronological age (mean = 11.68 years) and vocabulary age (mean = 10.1 years) of children with ASD were significantly higher than TD children (mean chronological/ vocabulary age = 7.74 years), but the two groups performed similarly in the homonymy condition. TD children demonstrated an advantage in learning polysemous words compared to homonymous words in the referent identification task, at both the encoding stage and one-week later, whereas this advantage was not observed in verbal children with ASD, who showed a slight delay in vocabulary development. Both Ferman (2021) and Floyd et al. (2021) found that verbal children with ASD would be less sensitive to the facilitative characteristics of to-be-learned words. As the first empirical studies that examined the effect of word characteristics on children with ASD, these findings enrich our understanding of the word learning mechanisms in the autistic population. Although there may not be significant group differences in overall learning performance, children with ASD may still be less sensitive to factors that are facilitative for TD children.

So far, no previous studies examined the role of word characteristics in Mandarin-speaking children with ASD. According to the China Autism Education and Rehabilitation Status Report V (2024), there are over 10 million individuals with ASD in China, of which 3 to 5 million are children between 3 to 14 years, and the number is still growing at a rate of nearly 200,000 per year. Despite such a high prevalence rate, there is a scarcity of studies on the language abilities of Chinese-exposed children with ASD (Su et al., 2018). To our knowledge, only 3 published articles that investigated the utilization of social cues on word learning in the same group of school-aged Mandarin-speaking children with ASD (Jing & Fang, 2014; Jing et al., 2014; Jing et al., 2013), and the results were consistent with previously reviewed studies. The specific aspects of learning Mandarin words in verbal children with ASD remained to be addressed. For Mandarin-speaking children, the facilitative effect of semantic heads has been observed for preschoolers and school-aged children with and without SLI (Chen & Liu, 2014; Liang, 2017). Using a more rigorous experimental design, Yu and Sheng (in preparation) also found the facilitative role of semantic head in forming phonological representations for Mandarin-speaking TD children. Considering that the semantic head commonly specifies the semantic category, it would be critical to examine its effect on word learning for Mandarin-speaking children with ASD, who are likely to have lexical-semantic processing deficits.

Using the same word learning task as Yu and Sheng (in preparation), the current study investigated the effect of semantic head on different aspects and stages of word learning for Mandarin-speaking children with ASD, aiming to deepen our understanding of their lexical acquisition mechanism.

### 1.3 Learners' abilities and word learning in children with ASD

The importance of learners' abilities (e.g., phonological memory) in word learning has been well documented in TD children (Yu & Sheng, in preparation), but received limited attention in children with ASD (Hou & Su, 2022). Only two studies examined the role of receptive vocabulary, nonverbal IQ, and age and found that receptive vocabulary was a significant predictor of their word learning success (Hartley et al., 2020; McGregor et al., 2013). The predictive role of phonological memory in word learning for children with ASD remained to be examined. Wang et al. (2017) conducted a meta-analysis to examine the working memory abilities in individuals with ASD and found significant working memory impairments in this population. Therefore, it is crucial to investigate the relationship between phonological memory and word learning in children with ASD, as it may shed light on the mechanisms underlying word learning difficulties in this population.

Examining the role of learners' abilities also showed the potential to explain group differences observed in word learning outcomes. By adding cognitive and language abilities (e.g., phonological memory and so on) as the covariates into to a model, Pomper et al. (2022) found that the previously observed group differences between TD children and children with Developmental Language Disorder (DLD) was no longer significant in some tasks that measured their word learning performance, indicating that the differences in the word learning abilities between DLD and TD groups were closely related to the differences in their general cognitive abilities. To the best of our knowledge, similar analysis has not been conducted to examine group differences between TD and ASD groups in word learning.

The current study investigated the role of phonological memory and compound awareness at the encoding stage for verbal children with ASD, as their predictive roles have been observed in the current word learning task for TD children (Yu & Sheng, in preparation). First, we would follow the method of Pomper et al. (2022) to examine whether phonological memory and compound awareness would address group differences observed in children's word learning performance at the encoding stage. Meanwhile, the roles of phonological memory and compound awareness in learning performance of children with ASD and TD children would be analyzed separately. By comparing with TD children, the current study also investigated whether skills that predict encoding performance differed between the two groups.

With regard to the retention performance, existing studies have attributed the observed group difference to the failure of sleep-based consolidation in children with ASD, which may result from their poor sleep quality (Fletcher et al., 2020; Hou & Su, 2022; Norbury et al., 2010). Sleep-based consolidation is an important pathway for learners to retain newly learned words (Davis & Gaskell, 2009; Schimke et al., 2023). Sleep problems are common in the autistic population (Fletcher et al., 2020; Schreck & Richdale, 2020). In addition to the poorer sleep quality as reported on a parent questionnaire (Children's Sleep Habits Questionnaire), Fletcher et al. (2020) also found different sleep patterns between the children with ASD and their TD peers by home polysomnography recordings, and this difference was related to the group differences observed in the stabilization of new semantic knowledge. This finding suggested that the sleep quality of children with ASD may be related to their retention performance of newly learned words. The available research directly examining the relationship between sleep quality and word retention in children with ASD is limited. To fill this gap, the current study collected sleep data from the Simplified version of Children's Sleep Habits Questionnaire and investigated whether parent reported sleep quality would associate with the retention performance in children with ASD.

**1.4 Research questions**

Using the same novel word learning task as Yu and Sheng (in preparation), the current study aimed to answer the following questions:

Research questions on encoding performance: would Mandarin-speaking children with ASD show difficulties in forming phonological and semantic representations and establishing the links between forms and referents at the encoding stage compared to TD peers? Would phonological memory and compound awareness explain the group differences observed in children's word learning performance at the encoding stage? Would the skills predicting encoding outcomes differ by group?

Based on previous studies (e.g., Carter & Hartley, 2021; Gladfelter & Barron, 2020; Hani et al., 2013), it was hypothesized that children with ASD would show the ability to establish the links and perform like TD controls in forming phonological representations. They may show difficulties in forming semantic representations. Phonological memory and compound awareness were expected to explain group differences observed at the encoding stage. We believed that children with ASD may rely more on phonological memory to learn new words because of their lexical-semantic processing deficits.

Research questions on retention performance: would Mandarin-speaking children with ASD show difficulties in retaining phonological and semantic representations of newly learned words? Would parent-reported sleep quality explain the differences observed between TD and ASD children's retention performance? Would parent-reported sleep quality be related to the retention performance?

Based on previous studies (Norbury et al., 2010; Fletcher et al., 2020), it was hypothesized that children with ASD would show difficulty in retaining phonological and semantic representations after a long-term delay (about one month). Sleep quality was expected to explain group differences at the retention stage. If changes were observed in the retention performance, it was hypothesized that these changes would be related to parent-reported sleep quality.

Research question on word characteristic effect: would the presence of semantic head influence the word learning in Mandarin-speaking children with ASD?

Based on previous findings (Ferman, 2022; Floyd et al., 2021), it was hypothesized that the effect of semantic head would not be observed in children with ASD at both encoding and retention stage.

**2 Methods**

**2.1 Participants**

This study obtained the ethical approval from the Institutional Review Board of The Hong Kong Polytechnic University (Reference number: HSEARS20221110003). Participants were recruited either through local kindergartens and private institutions that offer intervention service for children with special needs, or by placing recruitment posters on the Internet. A total of 114 children aged between 3;10 (year; month) to 7;02 were recruited from May to October 2023 in Jiangsu Province, Mainland China. 64 children were recruited for the TD group and 50 children were recruited for the ASD group. TD group in the current study was a subset of children in Study 1 (Yu & Sheng, in preparation).

The exclusion criteria for the encoding analysis were as follows: (1) withdrawal from this study (TD group : n =4; ASD group: n = 4); (2) nonverbal IQ < 70 (TD group: n =0; ASD group: n = 4); (3) parents or teachers did not provide proof of formal ASD diagnosis from the hospital (ASD group: n = 2); (4) the child's score on the SCQ questionnaire was lower than 12 (children aged over 4 years) or 11 (children aged lower than 4 years) (TD group: n = 5; ASD group: n = 15). The simplified Chinese version of the SCQ were completed by the parents. Liu et al. (2022) validated the internal consistency, sensitivity and specificity of the simplified Chinese version SCQ. Therefore, this study used the cut-offs provided by Liu et al. (2022) to include children in the ASD group and to exclude children in the TD group. Furthermore, the two groups were matched on the raw scores of PTONI at the group level to ensure comparability in basic cognitive skills, such as reasoning and visual information processing. 7 children from TD group with PTONI raw scores higher than 50 were excluded. In total, 47 TD children (23 girls, 3;11 to 7;02) and 25 children with ASD (2 girls, 4;01 to 7;01) were included in the encoding analysis. As for the retention analysis, two more exclusion criteria were applied: (1) Time 2 retention exceeded one day (TD: n = 1; ASD: n = 1); (2) Time 3 retention exceeded 10 days (TD: n = 3; ASD: n = 2). 43 TD children (22 girls, 3;11 to 7;02) and 22 children with ASD (1 girls, 4;01 to 7;01) were included in the retention analysis.

Table 2-1 presents the demographic information, test scores and comparison results of participants included in the final analysis. The independent sample *t*-test was used to compare the two groups of children across various abilities. Student's *t* was reported for data that generally followed the assumption of normality and the assumption of equal variance. Welch's *t* was reported for data that generally followed the normality assumption while violated the assumption of equal variance. Mann-Whitney *U* was reported for data that violated the assumption of normality. Significant group differences were observed except for the raw scores of PTONI and sleep quality. In general, this group of children with ASD were older than the TD group. The ASD group's receptive and expressive vocabulary, compound awareness and phonological memory skills were significantly lower than those of the TD group.

Table 2-1 Demographic information, test scores and comparisons between groups

| Domain | TD group (n = 47) | | ASD group (n = 25) | | Comparisons | |
|---|---|---|---|---|---|---|
| | Mean (SD) | Range | Mean (SD) | Range | Statistics | *p* |
| Age (months) | 62.72 (10.10) | 47 – 86 | 68.92 (9.80) | 49 – 84 | 2.5 [a] | 0.02 |
| Nonverbal IQ [d] (standard score) | 115.98 (15.01) | 76 – 143 | 100.16 (18.22) | 74 – 143 | -3.95 [a] | <.001 |
| Nonverbal IQ [d] (raw score) | 34.28 (8.20) | 20 – 50 | 31.28 (7.78) | 15 – 48 | -1.5 [a] | 0.14 |
| Receptive vocabulary [e] | 104.62 (12.30) | 81 – 131 | 71.52 (22.37) | 31 – 110 | -6.87 [b] | <.001 |
| Expressive vocabulary [e] | 109.74 (11.09) | 77 – 134 | 80.04 (15.20) | 54 – 112 | -8.63 [b] | <.001 |
| Digit span (forward) | 10.70 (2.33) | 6 – 15 | 7.80 (2.12) | 4 – 13 | -5.19 [a] | <.001 |
| Digit span (backward) | 4.47 (1.83) | 0 – 10 | 2.12 (2.59) | 0 – 9 | 272.5 [c] | <.001 |
| Compound analogy | 10.64 (2.71) | 3 – 15 | 5.68 (4.82) | 0 – 15 | -4.76 [b] | <.001 |
| Compound structure | 14.94 (3.35) | 10 – 24 | 12.88 (2.13) | 8 – 16 | 378 [c] | 0.01 |
| SCQ [f] | 6.57 (3.17) | 0 – 11 | 18.20 (4.33) | 12 – 28 | 13.00 [a] | <.001 |
| Sleep quality [g] | 47.30 (6.53) | 38 – 62 | 47.68 (5.68) | 36 – 61 | 0.23 [a] | 0.82 |
| The duration between Time 1 and Time 3 (days) [g] | 6.86 (0.47) | 6 – 8 | 7.50 (0.86) | 7 – 10 | 282 [c] | <.001 |
| The duration between Time 1 and Time 4 (days) [g] | 28.44 (2.46) | 25 – 39 | 28.64 (1.09) | 27 – 32 | 389 [c] | 0.21 |
| Mandarin exposure [h] | 4.60 (0.65) | - | 3.96 (1.21) | - | 430 [c] | 0.03 |
| | N | % | N | % | - | - |
| 20%-39% | 1 | 2.1 | 5 | 20.0 | - | - |
| 40% -59% | 1 | 2.1 | 3 | 12.0 | - | - |
| 60%-79% | 14 | 29.8 | 5 | 20.0 | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| 80%-100% | 31 | 66.0 | 12 | 48.0 | - | - |
| Maternal education [h] | 4.09 (0.90) | | 3.16 (1.14) | | 287.5 [c] | <.001 |
| | N | % | N | % | - | - |
| middle school and below | 2 | 4.3 | 4 | 16.0 | - | - |
| high school | 1 | 2.1 | 1 | 4.0 | - | - |
| associate | 2 | 4.3 | 8 | 32.0 | - | - |
| bachelor | 28 | 59.6 | 11 | 44.0 | - | - |
| master and above | 14 | 29.8 | 1 | 4.0 | - | - |

a. Student's *t*; b. Welch's *t*; c. Mann-Whitney *U*; d. Nonverbal IQ was measured by The Primary Test of Nonverbal Intelligence (PTONI, McGhee & Ehrler, 2008); e. Children's receptive and expressive vocabulary were assessed by the Mandarin Oral Vocabulary Screener (Sheng et al., in development) and presented as standardized scores; f. The simplified Chinese version of the Social Communication Questionnaire; g. Based on the final retention sample: 43 TD and 22 ASD; h. Information from the questionnaire parents filled about the child's family background, general health, developmental history, exposure to dialects or other languages, and so on.

**2.2 Novel word learning**

The current study applied a combination of in-person and online methods to collect data. All children participated in Day 1 of word learning in person. Among participants included in the final analysis, 26.4% completed all tasks in person, and the remaining participants completed at least one retention test online (mixed mode). Among those who participated in this study in mixed mode, 56.6% completed all retention tests online. Children completed all in-person sessions in a quiet room under the instruction of the examiner. The examiner recorded all offline sessions by a sound recorder (Aigo R2210). A 14-inch laptop with external speakers was used to present all visual and auditory stimuli. Parents or teachers helped children to participate in the online sessions by logging into an online meeting platform using laptops or tablets, with few exceptions (12 of the children included in the final analysis used mobile phones at least once to participate in the online tests). The camera must remain on during the online sessions and the video was recorded. In addition, parents or teachers recorded the audio on the children's side and sent the recordings to the examiner after the test in most cases (82.1% of all online tests included in the final analysis).

Children learned two types of novel pseudo-disyllabic novel nouns, one with semantic head (WH words) and one without semantic head (OH words). The first syllables of both types of words were pseudo-Mandarin syllables that matched each other in phonological neighborhood density and syllable length. The second syllables of WH words were real Mandarin syllables that served as semantic heads (e.g., "虫/ chong2/ insect"), while the second syllables of OH words were real syllables that did not serve as semantic heads. The second syllables of the two types of words also matched each other in phonological neighborhood density and syllable length. Eight words were taught in total, four for each type. The four word pairs are (WH -- OH):"piou2 yang2/sheep/羊 -- tiou2 yuan3/far/远", "bua4 chong2/insect/虫-- bia4 nong2/farming/农", "tia2 niao3/bird/鸟 -- tua2 hui1/gray/灰", and "pong4 guo3/fruit/果 -- mong1 huo2/live/活". The referents are unfamiliar sheep, insects, birds and fruits from McGregor et al. (2020).

The whole learning procedure spanned over four separate days and lasted one month. On the first day of learning (Time 1), children were trained to learn the eight target words. The whole training session started with a familiarization session to ensure children did not have precise names for all referents. During the familiarization session, children were asked to try their best to give a name to the picture on the screen. Then the formal teaching session began, consisting of four rounds of labelling and testing. In each labelling session, children heard each novel forms four times within sentences with its referent presented on the screen and had one opportunity to practice the pronunciation (e.g., "bua4 chong2，你看到的是 bua4 chong2。跟我读 bua4 chong2。/bua4 chong2, you are looking at bua4 chong2. Read after me bua4 chong2."; "非常好，记住这是 bua4 chong2。/Well done! Remember this is bua4 chong2."). The labelling session ended when all eight words had been taught. After the labeling session, a mispronunciation detection task and a referent-identification task constituted the testing session. These two tasks were used to measure children's receptive learning performance. The mispronunciation detection task was a two-alternative forced choice test in which the foil form differed from the target form on the initial consonant by two phonetic features (Gordon et al., 2022). The referent identification task was a three-alternative forced choice test in which one foil was the trained non-target from the same category and the other foil was an untrained distractor from the same category.

After the fourth labelling and testing round, children's overall learning performance was measured by a naming task and a definition task. In total, there were 24 exposures to each word form before the expressive measurement of word learning. The naming and definition tasks were also conducted one day after Time 1 (Time 2), one week after Time 1 (Time 3), and one month after Time 1 (Time 4) to examine the retention performance. Children's naming performance were scored on phoneme accuracy based on the segmentation schemas of C_V_C_T (consonant-vowel-consonant-tone). Children's definition performance was scored based on a 0-to-3-point scoring protocol. A response would score 0 if it did not contain any correct information about the target words; 1 if it contained one piece of correct information, 2 if it contained two pieces of correct information of different types (e.g., categorical information and color); and 3 if it contained correct categorical information and two or more types of observable features (e.g., color and body parts of animals). More detailed descriptions of materials and word learning procedure can be found in Yu and Sheng (in preparation).

**2.3 Measures of language, cognitive abilities and sleep quality**

Children completed tasks that measured their nonverbal IQ, oral vocabulary, phonological memory and compound awareness after they completed the main word learning tasks on separate days. The Mandarin Oral Vocabulary Screener (MOVS, Sheng et al., in development) was used to measure children's oral vocabulary. MOVS has a normative sample of over 1000 children with acceptable internal consistency (Cronbach's α= .78) for both receptive and expressive subtests. 72.2% of the children completed the vocabulary test during the offline session and the remaining children completed the vocabulary test in their first online session.

The digit span task from Wechsler Intelligence Scale for Children 4$^{th}$ edition (Chinese Version) (Li & Zhu, 2012) were used to measure children's phonological memory. All children completed the phonological memory test during the offline session. A single composite phonological memory score was calculated based on the total number of correct trials of the forward and backward digit span subtest. For children included in the final analysis, their raw scores (the total number of correct trials) of these two tasks were firstly converted to z scores and then averaged (Pomper et al., 2022).

A compound analogy task and a compound structure task (Lam & Sheng, 2016; Yang et al., 2021) were used to measure children's compound awareness. Yu and Sheng (in preparation) gave a detailed illustration of these two tasks. Children did not complete these two tasks at Time 1 and always completed the compound analogy before the compound structure task. 43.1% of the children completed both compound awareness tasks during the offline session and the remaining children completed at least one of compound awareness tests online. A single composite compound awareness score was calculated based on the total number of correct trials of the two tasks. For children included in the final analysis, their raw scores (the total number of correct trials) of these two tasks were firstly converted to z scores and then averaged (Pomper et al., 2022).

The Chinese version of the Children's Sleep Habits Questionnaire (CSHQ) was completed by parents or the main caregivers to measure children's sleep quality. Parents rated their children's daily sleep habits on a scale of "never", "sometimes" and "always" based on the most recent week's performance, including bedtime resistance, sleep-onset delay, sleep duration, sleep anxiety, night-time awakenings, parasomnias, sleep disordered breathing, and daytime sleepiness (Liu et al., 2014). The lower the total scores on this questionnaire, the better the quality of sleep. The internal consistency (0.72) and test-retest reliability (0.77) of the CSHQ have been reported in Chinese preschoolers (Liu et al., 2014).

**3. Results**

Mixed-effect regression was used to analyze children's word learning performance. Simple coding was used to contrast the variable of Group (-0.5 for ASD, +0.5 for TD) and Type. Dummy coding was used to contrast the variable of Round and Time. The *p*-

values were corrected with Bonferroni adjustment for post-hoc comparisons. Following Pomper et al. (2022), two-step modeling to add covariates was used to investigate the role of phonological memory and compound awareness during the encoding stage, and the role of sleep quality during the retention stage. For tasks that measured the encoding performance (Time 1), Model-1 included Age (in months) and Mandarin exposure as covariates. Model-2 additionally included phonological memory and compound awareness as covariates to examine whether any changes would occur to group differences. For retention performance, Model-1 included Age (in months), Mandarin exposure and the duration between Time 1 and Time 3 (in days) as covariates. Model-2 additionally included the scores of sleep quality as covariates to examine whether any changes would occur.

The analysis of whether different skills would predict learning outcomes between TD and ASD group at the encoding stage was modeled after Hartley et al. (2020). Separate models were built for TD and ASD groups in each task. Like Yu & Sheng (in preparation), a series of comparisons was conducted to compare models including one or more abilities.

R (R Core Team, 2024) and Jamovi (The jamovi project, 2024) were used for analyses and visualization of results. The following packages were used: lme4 (Bates et al., 2015), BruceR (Bao, 2023), ggplot2 (Wickham, 2016), rcartocolor (Nowosad, 2018), tidyverse (Wickham et al., 2019), and GAMLj (Gallucci, 2023).

### 3.1 Coding and reliability

All participants' learning outcomes were coded offline by the audio/ video recordings and scoring sheets. 20 percent of participants' data were double coded for the reliability check. Reliability was 95% for the MPD task and 100% for the referent identification task. The reliability of overall phoneme accuracy was 94%. The scoring consistency of definition was 95%. Inconsistencies were resolved based on the audio recordings and incorrect codings were corrected.

### 3.2 Encoding

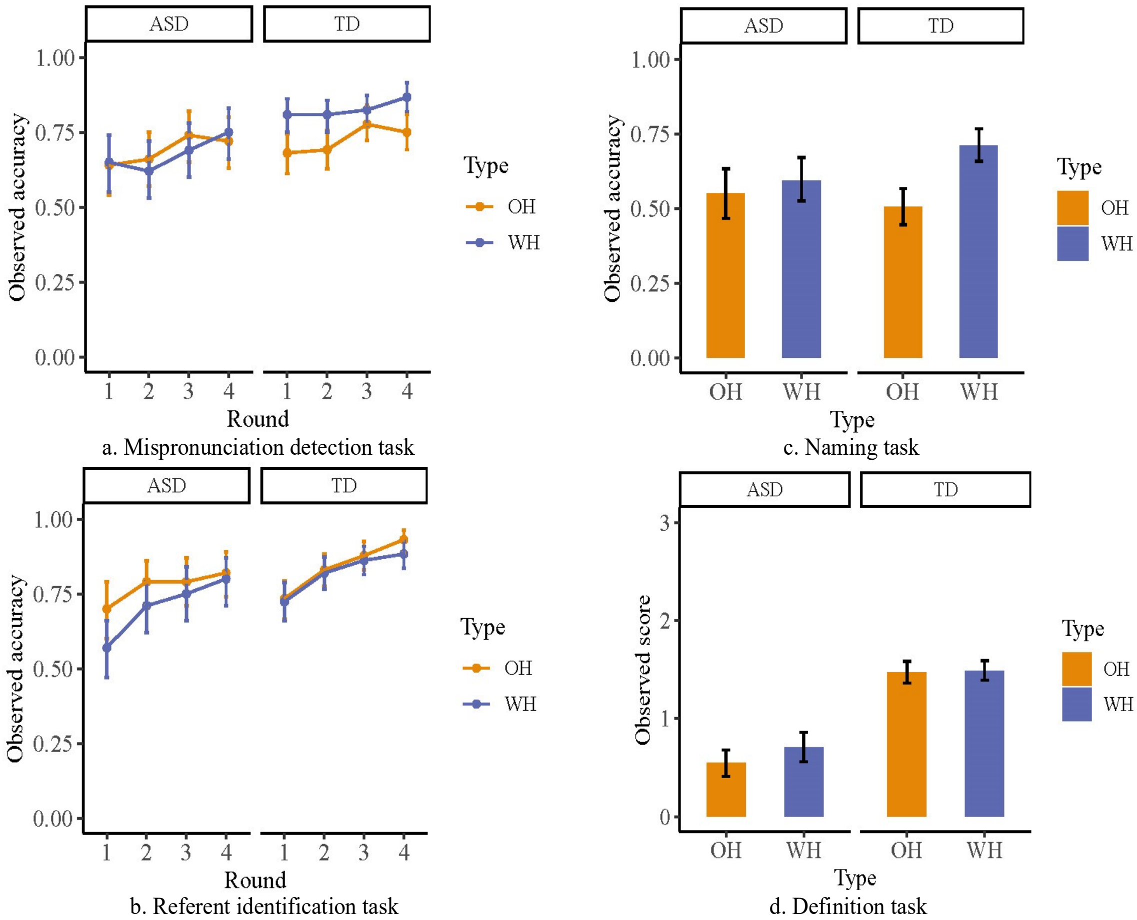



Figure 2-1 TD and ASD groups' learning performance at Time 1

#### 3.2.1 Mispronunciation detection task

Figure 2-1a presents the two groups' performance on the mispronunciation detection task. Results of the binomial test showed that both TD children and children with ASD performed significantly differently from the chance level (0.5) in choosing the correct forms of WH and OH words (all *p*s < 0.001). Table 2-2 presents detailed results of

Model-1 and Model-2.

Model-1 showed that the main effect of Group, the main effect of Round, and the interaction between Group and Type was statistically significant. Post-hoc analysis showed that children with ASD showed a lower probability of identifying correct forms of WH words than TD children ($z$ = -3.42, $p$ < .001). The probability of identifying correct forms of OH words did not differ significantly between the two groups ($z$ = -.35, $p$ = .73). For TD children, the probability of identifying correct word forms of OH words was lower than that of WH words ($z$ = -2.36, $p$ = .02). The probability of identifying the correct word form of OH words and WH words did not differ for children with ASD ($z$ = .32, $p$ = .75).

Table 2-2 Result summary of mispronunciation detection task

| | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|
| | $\chi^2$ | *df* | *p* | $\chi^2$ | *df* | *p* |
| Group | 4.56 | 1.00 | 0.03 | 0.01 | 1.00 | 0.93 |
| Type | 1.12 | 1.00 | 0.29 | 1.13 | 1.00 | 0.29 |
| Round | 14.99 | 3.00 | 0.002 | 14.97 | 3.00 | 0.002 |
| Group * Type | 11.57 | 1.00 | < .001 | 11.55 | 1 | < .001 |
| Age | 2.25 | 1.00 | 0.13 | 1.65 | 1.00 | 0.20 |
| Mandarin exposure | 8.53 | 1.00 | 0.003 | 7.01 | 1.00 | 0.01 |
| Phonological memory | -- | -- | -- | 7.88 | 1.00 | 0.01 |
| Compound awareness | -- | -- | -- | 0.29 | 1.00 | 0.59 |
| | Prob. [a] | SE [b] | 95% CI [c] | Prob. [a] | SE [b] | 95% CI [c] |
| TD-WH | 0.84 | 0.03 | [0.78,0.89] | 0.82 | 0.03 | [0.76,0.88] |
| TD-OH | 0.74 | 0.04 | [0.66,0.81] | 0.72 | 0.04 | [0.63,0.79] |
| ASD-WH | 0.71 | 0.05 | [0.61,0.80] | 0.762 | 0.04 | [0.66,0.84] |
| ASD-OH | 0.73 | 0.05 | [0.63,0.81] | 0.78 | 0.04 | [0.68,0.85] |

Note: Both Model-1 and Model-2 included Group, Type, Round and the interaction between Group and Type as fixed effects, and by-participant and by-item intercept as random effects; a. Estimated marginal means/probability would be presented for significant effects and on the second half of the model summary tables for each task; b. SE = standard error; c: CI: Confidence Interval.

After additionally controlling phonological memory and compound awareness (Model-2), the main effect of Group was no longer significant. The interaction between Group and Type was still significant. The probability of identifying correct forms of WH words ($z$ = -1.45, $p$ = .15) and OH words ($z$ = 1.30, $p$ = .19) was not significantly different between the two groups. The significant effect of Type was still present for TD children ($z$ = -2.36, $p$ = .02) and absent for children with ASD ($z$ = .32, $p$ = .75).

### 3.2.2 Referent identification task

Figure 2-1b presents the two groups' performance on the referent identification task. Results of the binomial test showed that both TD children and children with ASD performed significantly differently from the chance level (0.33) in choosing the correct referents of WH and OH words (all $p$s < 0.001). Table 2-3 presents detailed results of Model-1 and Model-2. Both Model-1 and Model-2 showed the main effect of Round was statistically significant.

Table 2-3 Result summary of referent-identification task

| | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|
| | $\chi^2$ | *df* | *p* | $\chi^2$ | *df* | *p* |
| Group | 2.07 | 1.00 | 0.15 | 3.07 | 1 | 0.08 |
| Type | 0.95 | 1.00 | 0.33 | 0.96 | 1 | 0.33 |
| Round | 74.24 | 3.00 | < .001 | 73.95 | 3 | < .001 |
| Group * Type | 1.31 | 1.00 | 0.25 | 1.33 | 1 | 0.25 |
| Age | 0.15 | 1.00 | 0.70 | 0.82 | 1 | 0.37 |
| Mandarin exposure | 1.30 | 1.00 | 0.25 | 0.11 | 1 | 0.74 |
| Phonological memory | -- | -- | -- | 15.39 | 1 | < .001 |
| Compound awareness | -- | -- | -- | 4.46 | 1 | 0.04 |

Note: Both Model-1 and Model-2 included Group, Type, Round and the interaction between Group and Type as fixed effects, by-participant and by-item intercept as random effects.

### 3.3.3 Naming task

Figure 2-1c presents children's performance on the naming task. Table 2-4 presents detailed results of Model-1 and Model-2. Model-1 showed that the interaction between Group and Type was statistically significant. Post-hoc analysis showed that the phoneme accuracy of WH words ($t$ = -1.23, $p$ = .22) and OH words ($t$ = 1.27, $p$ = .20) did not differ significantly between the two groups. The significant effect of Type was observed in the TD group ($t$ = -2.61, $p$ = .02), indicating that the accuracy of phonological representations of WH words was higher than OH words. The effect of word type was not observed in the ASD group ($t$ = -.51, $p$ = .62).

Table 2-4 Result summary of naming performance at Time 1

| | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|
| | *F* | *df* | *p* | *F* | *df* | *p* |
| Group | < 0.001 | [1,71.63] | 0.98 | 9.52 | [1,71.53] | 0.003 |
| Type | 3.42 | [1,8.25] | 0.10 | 3.47 | [1,8.29] | 0.10 |
| Group * Type | 7.02 | [1,496.70] | 0.01 | 7.02 | [1,469.65] | 0.01 |
| Age | 0.10 | [1,71.63] | 0.75 | 0.71 | [1,71.53] | 0.40 |
| Mandarin exposure | 3.36 | [1,71.63] | 0.07 | 1.02 | [1,71.53] | 0.32 |
| Phonological memory | -- | -- | -- | 4.77 | [1,71.53] | 0.03 |
| Compound awareness | -- | -- | -- | 16.73 | [1,71.53] | < .001 |
| | Mean [a] | SE [b] | 95% CI [c] | Mean [a] | SE [b] | 95% CI [c] |
| TD-WH | 0.70 | 0.06 | [0.57,0.83] | 0.65 | 0.06 | [0.51,0.76] |
| TD-OH | 0.49 | 0.06 | [0.36,0.62] | 0.43 | 0.06 | [0.31,0.56] |
| ASD-WH | 0.62 | 0.07 | [0.47,0.76] | 0.74 | 0.07 | [0.59,0.88] |
| ASD-OH | 0.57 | 0.07 | [0.43,0.72] | 0.69 | 0.07 | [0.55,0.84] |

Note: Both Model-1 and Model-2 included Group, Type and their interaction as fixed effects, by-participant and by-item intercept as random effects; a. Estimated marginal means/probability would be presented for significant effects and on the second half of the model summary tables for each task; b. SE = standard error; c: CI: Confidence Interval.

After controlling phonological memory and compound awareness (Model 2), the main effect of Group was statistically significant ($\beta$ = -0.18, $t$ = -3.08, $p$ = 0.003), and the interaction between Group and Type was still significant. Children with ASD performed better in naming OH words than TD children ($t$ = 3.84, $p$ < .001) after differences in phonological memory and compound awareness were accounted for. No group differences were observed for WH words ($t$ = 1.46, $p$ = .15). The significant effect of Type was observed in the TD group ($t$ = -2.61, $p$ = .02) but not in the ASD group ($t$ = -.51, $p$ = .61).

### 3.3.4 Definition task

Figure 2-1d presents children's performance on the definition task. Table 2-5 presents detailed results of Model-1 and Model-2. Model-1 showed that the main effect of Group was statistically significant. After controlling phonological memory and compound awareness (Model 2), the effect of Group was still statistically significant.

Table 2-5 Result summary of definition performance at Time 1

| | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|
| | *F* | *df* | *p* | *F* | *df* | *p* |
| Group | 34.75 | [1,72] | < .001 | 10.74 | [1,72] | 0.002 |
| Type | 3.56 | [1,504] | 0.06 | 3.56 | [1,504] | 0.06 |
| Group * Type | 2.39 | [1,504] | 0.12 | 2.39 | [1,504] | 0.12 |
| Age | 2.79 | [1,72] | 0.10 | 2.13 | [1,72] | 0.15 |
| Mandarin exposure | 2.42 | [1,72] | 0.12 | 1.2 | [1,72] | 0.28 |
| Phonological memory | -- | -- | -- | 4.28 | [1,72] | 0.04 |
| Compound awareness | -- | -- | -- | 1.83 | [1,72] | 0.18 |

Note: Both Model-1 and Model-2 included Group, Type and their interaction as fixed effects, by-participant as the random effect.

### 3.3.5 Learners' abilities

Table 2-6 summarizes the results of all model comparisons in the four tasks at the encoding stage. In the mispronunciation detection task, phonological memory predicted TD children's performance, while neither ability predicted ASD children's performance. For referent identification, both phonological memory and compound awareness predicted TD children's performance, and their contributions were similar; phonological memory predicted ASD children's performance. In the naming task, compound awareness predicted TD children's performance. Both phonological memory and compound awareness predicted ASD children's naming performance, and their contributions were similar. In the definition task, neither ability predicted TD children's performance, while both phonological memory and compound awareness predicted ASD children's performance, and the contribution of compound awareness was more salient.

Table 2-6 The role of phonological memory and compound awareness on the encoding stage

| | | TD | | ASD | |
|---|---|---|---|---|---|
| | | $\chi^2(1)$ | *p* | $\chi^2(1)$ | *p* |
| Mispronunciation detection task | PM [a] vs Baseline [d] | 8.21 | 0.004 | 2.41 | 0.12 |
| | CA [b] vs Baseline [d] | 2.49 | 0.11 | 0.84 | 0.36 |
| Referent identification task | PM [a] vs Baseline [e] | 8.03 | 0.01 | 11.04 | <0.001 |
| | CA [b] vs Baseline [e] | 7.75 | 0.01 | 2.79 | 0.09 |
| | PM [a] vs PM & CA [c] | 4.30 | 0.04 | N.A [h] | |
| | PM [a] vs PM & CA [c] | 4.58 | 0.03 | | |
| Naming task | PM [a] vs Baseline [f] | 2.64 | 0.10 | 7.87 | 0.01 |
| | CA [b] vs Baseline [f] | 8.10 | 0.004 | 15.32 | <0.001 |
| | PM [a] vs PM & CA [c] | N.A [h] | | 13.77 | <0.001 |
| | CA [b] vs PM & CA | | | 6.31 | 0.01 |
| Definition task | PM [a] vs Baseline [g] | 2.90 | 0.09 | 4.55 | 0.03 |
| | CA [b] vs Baseline [g] | 0.91 | 0.34 | 5.72 | 0.02 |
| | PM [a] vs PM & CA [c] | N.A [h] | | 4.07 | 0.04 |
| | CA [b] vs PM & CA [c] | | | 2.90 | 0.09 |

Note: a. PM: model only contained phonological memory as the fixed effect, the same below; b. CA: model only contained compound awareness as the fixed effect, the same below; c. PM & CA: model contain both phonological memory and compound awareness as fixed effects, the same below; d. The baseline model contained by-participant and by-item intercept as random effects; e. The baseline model contained by-participant and by-item intercepts as random effects.; f. The baseline model contained by-participant and by-item intercepts as random effects.; g. The baseline model contained by-participant intercept as the random effect; h. N.A: not applicable.

### 3.3 Retention

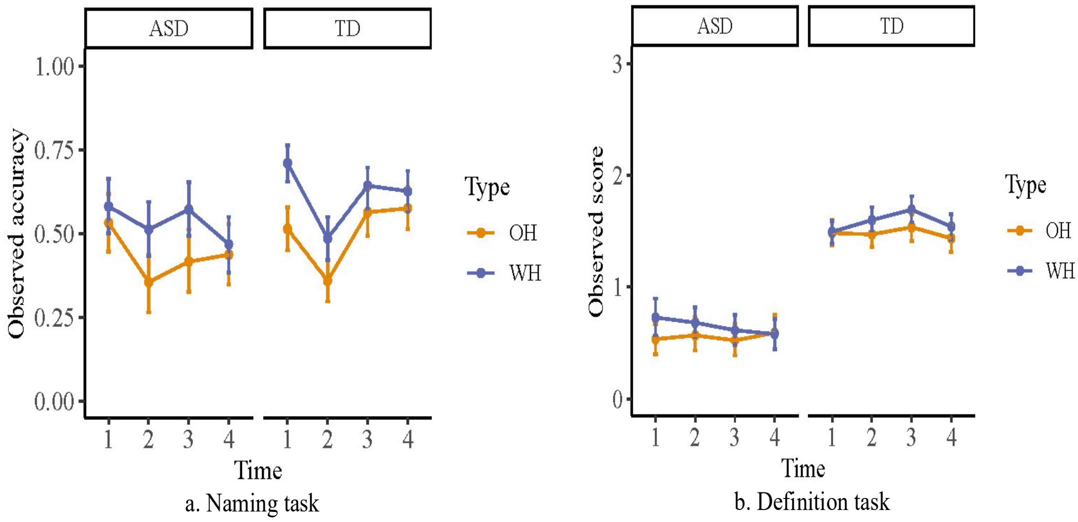



Figure 2-2 TD and ASD children's naming and definition performance across four time points

#### 3.3.1 Naming task

Figure 2-2-a presents children's performance on the naming task across four time points. Table 2-7 presents detailed results of Model-1 and Model-2.

Model-1 showed that the main effect of Time and the interaction between Group and Time were statistically significant. Post-hoc analysis of the significant interaction between Group and Time showed that group differences (ASD - TD) were not significant at the three retention points: Time 2 ($t$ = .35, $p$ = .73), Time 3 ($t$ = -1.32, $p$ = .19), and Time 4 ($t$ = -1.87, $p$ = .06).

For TD children, naming performance at Time 1 was better than Time 2 ($t$ = 7.01, $p$ < .001), and Time 2 performance was worse than Time 3 ($t$ = - 6.69, $p$ < .001) and Time 4 ($t$ = -6.60, $p$ < .001). No significant difference was found between Time 1 and Time 3 ($t$ = .33, $p$ = 1.00), Time 1 and Time 4 ($t$ = .42, $p$ = 1.00), and Time 3 and Time 4 ($t$ = 0.09, $p$ = 1.00).

For children with ASD, naming performance at Time 1 was better than Time 2 ($t$ = 3.26, *p* = .01) and Time 4 (*t* = 2.78, *p* = .03). No significant difference was found between Time 2 and Time 3 (*t* = - 1.60, *p* = .66), Time 3 and Time 4 (*t* = 1.12, *p* = 1.00), Time 1 and Time 3 (*t* = 1.66, *p* = .58), Time 2 and Time 4 (*t* = -.47, *p* = 1.00).

Table 2-7 Result summary of naming performance across four time points

| | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|
| | *F* | *df* | *p* | *F* | *df* | *p* |
| Group | 0.92 | [1,64.51] | 0.34 | 1.02 | [1,64.51] | 0.32 |
| Time | 16.71 | [3,2007.94] | < .001 | 16.71 | [3,2007.94] | < .001 |
| Type | 1.86 | [1,8.00] | 0.21 | 1.86 | [1,800] | 0.21 |
| Group * Time | 4.50 | [3,2007.94] | 0.004 | 4.50 | [3,2007.94] | 0.004 |
| Group * Type | 0.23 | [1,2007.94] | 0.63 | 0.21 | [1,2007.94] | 0.63 |
| Type * Time | 1.88 | [3,2007.94] | 0.13 | 1.88 | [3,2007.94] | 0.14 |
| Group * Type * Time | 2.17 | [3,2007.94] | 0.09 | 2.17 | [3,2007.94] | 0.09 |
| Age | 0.80 | [1,64.51] | 0.38 | 0.79 | [1,64.51] | 0.38 |
| Mandarin exposure | 1.44 | [1,64.51] | 0.24 | 1.18 | [1,64.51] | 0.28 |
| Time 1-3 duration | 0.12 | [1,64.51] | 0.73 | 0.08 | [1,64.51] | 0.78 |
| Sleep quality | -- | -- | -- | 0.16 | [1,64.51] | 0.69 |
| | Mean [a] | SE [b] | 95% CI [c] | Mean [a] | SE [b] | 95% CI [c] |
| TD-Time 1 | 0.61 | 0.06 | [0.49,0.73] | 0.61 | 0.06 | [0.49,0.73] |
| TD-Time 2 | 0.42 | 0.06 | [0.30,0.54] | 0.42 | 0.06 | [0.30,0.54] |
| TD-Time 3 | 0.60 | 0.06 | [0.48,0.72] | 0.60 | 0.06 | [0.48,0.72] |
| TD-Time 4 | 0.60 | 0.60 | [0.48,0.71] | 0.60 | 0.06 | [0.48,0.72] |
| ASD-Time 1 | 0.57 | 0.07 | [0.42,0.71] | 0.56 | 0.07 | [0.42,0.71] |
| ASD-Time 2 | 0.44 | 0.07 | [0.30,0.58] | 0.44 | 0.07 | [0.30,0.58] |
| ASD-Time 3 | 0.50 | 0.07 | [0.36,0.65] | 0.50 | 0.07 | [0.36,0.64] |
| ASD-Time 4 | 0.46 | 0.07 | [0.32,0.60] | 0.46 | 0.07 | [0.32,0.60] |

Note: Both Model-1 and Model-2 included Group, Type, Time, all two-way interactions and the three-way interaction as fixed effects, by-participant and by-item intercept as random effects; a. Estimated marginal means/probability would be presented for significant effects and on the second half of the model summary tables for each task; b. SE = standard error; c: CI: Confidence Interval.

After adding sleep quality as a covariate (Model 2), the main effect of Time and the interaction between Group and Time remained statistically significant. The results of the post-hoc analysis were consistent with those of Model 1.

Changes in the retention of phonological representation were observed. Pearson correlation was used to examine whether sleep quality was associated with the changes (i.e., increases or decreases) in phoneme accuracy during the retention stage in the two groups of children. Table 2-8 presents the results of correlation analysis.

Table 2-8 Correlation analysis between sleep quality and changes in retention

| | TD group | | ASD group | |
|---|---|---|---|---|
| | *r* | *p* | *r* | *p* |
| T2-T1 | -0.02 | 1.00 | -0.55 | 0.02 |
| T3-T1 | -0.09 | 1.00 | -0.53 | 0.03 |
| T4-T1 | -0.22 | 0.45 | -0.34 | 0.35 |

Note: a: *p* value corrected for multiple comparisons.

No significant relationship was observed between sleep quality and changes in phoneme accuracy in the TD group. A significant negative correlation was observed between sleep quality and changes in phoneme accuracy from Time 1 to Time 2, and from Time 1 to Time 3, in children with ASD. As shown in Figure 2-3, for children with ASD, higher scores on the Children's Sleep Habits Questionnaire (i.e., poorer sleep quality) were associated with decreases in phoneme accuracy at Time 2 and Time 3 as compared to Time 1.

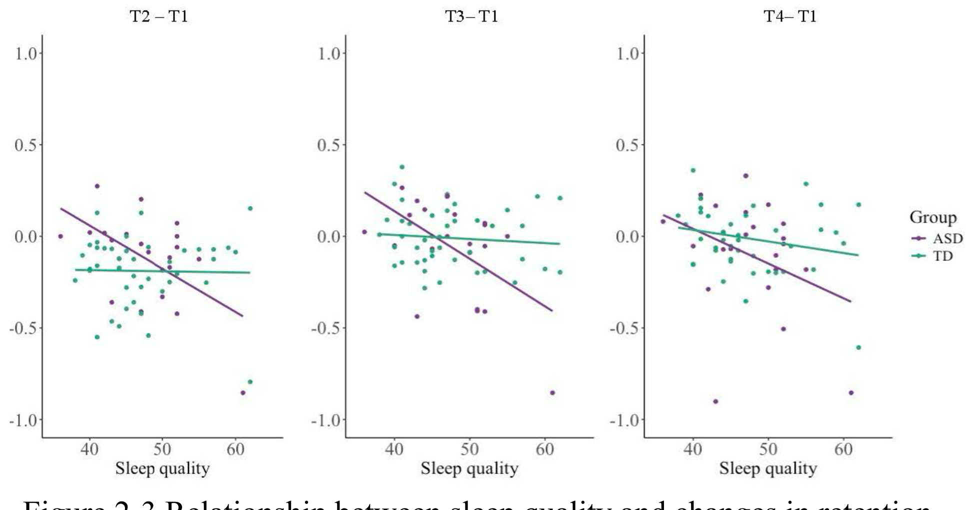



Figure 2-3 Relationship between sleep quality and changes in retention

### 3.3.2 Definition task

Figure 2-2b presents children's performance on the definition task across the four time points. Table 2-9 presents detailed results of Model-1 and Model-2.

Table 2-9 Result summary of definition performance across four time points

| | Model 1 | | | Model 2 | | |
|---|---|---|---|---|---|---|
| | *F* | *df* | *p* | *F* | *df* | *p* |
| Group | 35.56 | [1,64.98] | <0.001 | 37.55 | [1,64.98] | < .001 |
| Time | 0.90 | [3,2007.92] | 0.44 | 0.90 | [3,2007.92] | 0.44 |
| Type | 4.30 | [1,7.73] | 0.07 | 4.31 | [1,7.73] | 0.07 |
| Group * Time | 2.59 | [3,2007.92] | 0.051 | 2.59 | [3,2007.92] | 0.051 |
| Group * Type | 0.01 | [1,2007.92] | 0.94 | 0.01 | [1,2007.92] | 0.94 |
| Type * Time | 0.51 | [3,2007.92] | 0.68 | 0.51 | [3,2007.92] | 0.68 |
| Group * Type * Time | 1.68 | [3,2007.92] | 0.17 | 1.68 | [3,2007.92] | 0.17 |
| Age | 12.55 | [1,64.98] | < .001 | 12.66 | [1,64.98] | < .001 |
| Mandarin exposure | 6.45 | [1,64.98] | 0.01 | 5.10 | [1,64.98] | 0.03 |
| Time 1-3 duration | 4.29 | [1,64.98] | 0.04 | 3.60 | [1,64.98] | 0.06 |
| Sleep quality | -- | -- | -- | 1.34 | [1,64.98] | 0.25 |
| | Mean [a] | SE [b] | 95% CI [c] | Mean [a] | SE [b] | 95% CI [c] |
| TD-Time 1 | 1.46 | 0.08 | [1.30,1.63] | 1.47 | 0.08 | [1.30,1.64] |
| TD-Time 2 | 1.51 | 0.08 | [1.34,1.68] | 1.52 | 0.08 | [1.35,1.68] |
| TD-Time 3 | 1.59 | 0.08 | [1.42,1.75] | 1.60 | 0.08 | [1.43,1.76] |
| TD-Time 4 | 1.46 | 0.08 | [1.30,1.63] | 1.47 | 0.08 | [1.30,1.64] |
| ASD-Time 1 | 0.68 | 0.12 | [0.44,0.92] | 0.67 | 0.12 | [0.42,0.91] |
| ASD-Time 2 | 0.68 | 0.12 | [0.43,0.92] | 0.66 | 0.12 | [0.42,0.90] |
| ASD-Time 3 | 0.62 | 0.12 | [0.38,0.86] | 0.60 | 0.12 | [0.36,0.85] |
| ASD-Time 4 | 0.64 | 0.12 | [0.39,0.88] | 0.62 | 0.12 | [0.38,0.86] |

Note: Both Model-1 and Model-2 included Group, Type, Time, all two-way interactions and the three-way interaction as fixed effects, by-participant and by-item intercept as random effects; a. Estimated marginal means/probability would be presented for significant effects and on the second half of the model summary tables for each task; b. SE = standard error; c: CI: Confidence Interval.

Model-1showed that the main effect of Group was statistically significant ($\beta = 0.78$, $t = 5.23$, $p < 0.001$), and the interaction between Group and Time was marginally significant.

Post-hoc analysis of interaction between Group and Time showed that the TD group provided more accurate sematic information than the ASD group across all time points (all *p*s < .001). Children with ASD performed similarly across the four time points (all *p*s = 1.00). TD children provided more accurate sematic information at the Time 3 (Time 1-3: $t = -3.02$, $p = 0.02$; Time 3-4: $t = 3.09$, $p = 0.01$).

After adding the sleep quality as the covariates, the main effect of Group was still statistically significant ($\beta = 0.80$, $t = 5.38$, $p < 0.001$), and the interaction between Group and Time was still marginally significant. The results of the post-hoc analysis were consistent with those of Model 1.

## 4. Discussion

### 4.1 Group differences in learning performance at the encoding stage

Among the four tasks children completed at the encoding stage (Time 1), the mispronunciation detection task and naming task addressed the phonological representations of new words. Children with ASD performed worse than TD children in identifying correct word forms of WH words (the mispronunciation detection task), however, on the arguably more difficult task of naming, the accuracy of autistic children's phonological representations of new words did not differ from younger TD controls. Recalling the word learning procedure used in the current study, the mispronunciation detection task was conducted four times and was always following the explicit labelling session, suggesting that this receptive measure can be considered as a part of the training. The naming task was conducted after children completed the last referent-identification task and can be seen as a measure of children's overall learning outcomes of phonological representations at Time 1. Therefore, although group differences were observed during the training, the naming performance of children with ASD did not differ from TD controls after training. Overall, consistent with previous findings (e.g., Lucas & Norbury, 2014), forming the phonological representations of new words was not challenging for verbal children with ASD. Another finding that was consistent with previous studies came from the results of the referent identification task. Group differences were also not observed in the referent-identification task, suggesting that children with ASD could establish the links between forms and referent.

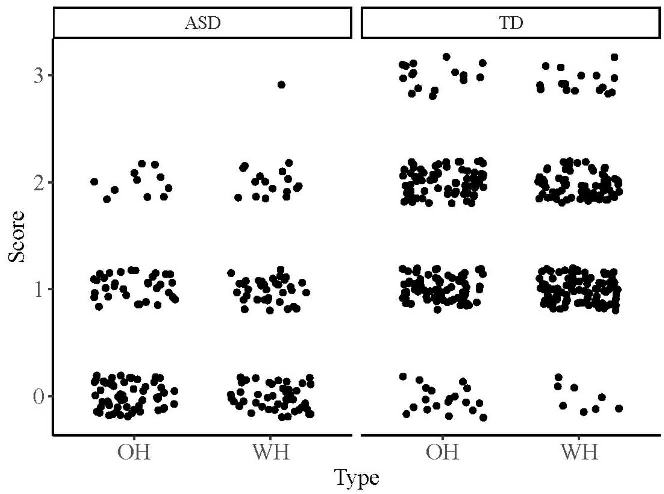



Figure 2-4 Distribution of definition scores of the two groups at Time 1

Salient group differences were observed in the definition task, which measures expressive semantic knowledge of target words. Children with ASD performed significantly worse than TD controls on the definition task. Figure 2-4 presents the distribution of definition scores for the two groups of children at Time 1. As seen in this figure, most children with ASD scored 0 on this task. But this does not necessarily mean that children with ASD could not form semantic representations of newly learned words.

To give a verbal definition of a word, children need to search their mental lexicon, extract and organize related information and report them (Wehren et al., 1981). Failure in any of these steps may result in failure to complete the definition task. It is possible that the floor effect observed in our children with ASD was due to their inability to complete the definition task. 5 children with ASD scored 0 for all trials because of no response, saying "I don't know.", or repeating target words. Although this task was also difficult for our young TD children (1 child from the TD group scored 0 for all trials because of no response or saying, "I don't know."), most of them showed the understanding that they needed to produce categorical information or other descriptive information which they believed to be related to the target words to complete the task. In short, TD children generally understood the task though many of their responses were still immature. Previous studies that investigated the development of definition skills of real nouns also showed that preschoolers can complete this task (To et al., 2013). Therefore, the failure to complete the definition task in some of the ASD children suggested that their definition skills were underdeveloped.

For children with ASD who were able to give verbal definitions, a majority of them were able to get 1 point, suggesting that they were only able to provide one piece of correct semantic information of target words. This finding was consistent with some previous studies that also used the definition task to measure school-aged autistic children's definition performance (e.g., Gladfelter & Barron, 2020; Lucas et al., 2017). McGregor et al. (2012) also found that school-aged children with ASD who showed language impairments scored lower in giving verbal definition of real words. Results of the current study provided the first empirical evidence from preschoolers, suggesting that providing the expressive semantic knowledge of newly learnt words was extremely challenging for verbal children with ASD.

Given that the current study used a challenging free definition task, we should be cautious to make the argument that verbal children with ASD may not be able to adequately encode the features of referents and/or store them into the mental lexicon. Children's performance would change if more cues were provided (e.g., directly asking children about the color of the target word), or using a receptive task (e.g., asking children to choose the semantic category of the target word).

We also intended to examine the role of phonological memory and compound awareness in children's encoding performance. Significant differences were observed between the two groups in both phonological memory and compound awareness. After controlling phonological memory and compound awareness, the group difference was no longer significant in the mispronunciation detection task, and children with ASD performed better than TD controls in the naming task. However, in the definition task, phonological memory and compound awareness did not cancel out the group differences. These results suggested that, with the exception of the definition task, if children with ASD were able to achieve a similar level of these cognitive and language skills as TD children, they would not show deficits in some aspects of word learning and may even perform better than TD controls.

Further analysis of the contribution of phonological memory and compound awareness to learning performance at the encoding stage indicated that the two groups relied on different abilities. For TD children, consistent with Yu & Sheng (in preparation), phonological memory and compound awareness contributed similarly to children's encoding performance, with 3 of the encoding tasks' performance predicted by these abilities. As for children with ASD, 3 of the encoding tasks' performance could be predicted by phonological memory and compound awareness, and there was a tendency for these children to rely more on phonological memory. Phonological memory effect was present in all three tasks wherein the predictive roles of these two abilities were identified. This finding suggests that training on phonological memory may enhance the ability of encoding new words for children with ASD. As for the definition task, a task where children with ASD showed significant difficulties, the current results showed that the role of compound awareness was more salient. This finding was not surprising since category information is included in compound words. If children with ASD could realize that the second morphemes of some words (WH words in the current study) represented semantic categories, they would be able to produce categorical information in the definition task. This finding also suggests that training on compound awareness may enhance children's ability to identify the semantic category of new words.

**4.2 Group differences in learning performance at the retention stage**

A naming and a definition task were used to measure the retention of the newly learned words one-day after training (Time 2), about one-week after training (Time 3) and about one-month after training (Time 4). Group differences were not observed in the naming task at Time 2, Time 3 and Time 4, suggesting that in general, children with ASD did not perform more poorly than TD controls in the accuracy of phonological representations at these three time points. However, the two groups of children showed different retention patterns. For TD children, a significant decrease from Time 1 to Time 2 was observed. Their naming performance significantly improved from Time 2 to Time 3 and was retained until Time 4, suggesting that TD children retained phonological representations of newly learned words at the short-term (about one-week) and long-term (about one-month) delay.

For children with ASD, there was also a significant decrease in phoneme accuracy at Time 2 (compared to Time 1). However, an improvement from Time 2 to Time 3 was not observed, and a significant decrease was observed at Time 4 (compared to Time 1), suggesting that children with ASD did not retain phonological representations of newly learned words at the long-term (about one-month) delay. Moreover, although not statistically significant, a trend that children with ASD performed worse than TD children was observed at Time 4 ($p$ = .06), providing additional support for the argument that children with ASD have deficits in the long-term retention of phonological representations of newly learned words. Consistent with an existing study that examined one-month retention of school-aged children with ASD (Norbury et al., 2010), the current findings suggested that children with ASD may not be able to retain the newly learned words after a longer delay.

Why would children with ASD show deficits in long-term retention? We hypothesized that poor long-term retention may stem from the sleep problems of children with ASD. Based on the results of Children's Sleep Habits Questionnaire, we did not see group differences in sleep quality between the two groups (see Table 2.1). Adding sleep quality to models that examined retention performance did not change the findings. It seems that our current sample of children with ASD did not have significant sleep problems, and the parent-report sleep quality could not explain their deficits in long-term retention. It is possible that as an indirect measure of children's sleep quality, the data from the parent-reported questionnaire (CSHQ) may be not precise enough to identify sleep problems of children with ASD. Besides, CSHQ requires that parents rely on the most recent week's sleep patterns to rate children's sleep habits, which is not consistent with the duration of one-month retention. Therefore, it is possible that the results of the CSHQ provided limited power to explain the deficits in long-term retention in children with ASD.

However, sleep quality was not entirely unrelated to the retention performance in children with ASD. The results of correlation analysis suggested that only in the ASD group, sleep quality was significantly related to the changes in retention performance one-day and one-week after training. Poorer sleep quality was associated with decrease of phoneme accuracy at one-day and one-week delay. This finding further suggested that sleep quality is more likely to influence the retention of phonological representations of newly learned words in children with ASD than in TD children. Comparing the retention pattern between the two groups of children, a lack of a significant increase from Time 2 to Time 3 in children with ASD was observed. This difference may be related to the influence of sleep quality on the retention of newly learned words for children with ASD.

Another possible explanation for the lack of a significant increase from Time 2 to Time 3 in the ASD group's naming performance may be related to the effect of the effect of re-exposure or retrieval practice at Time 2 (James et al., 2023). Although direct one-to-one re-exposure of the form and referent was not available at Time 2, the practice of completing the naming and definition task may have played a role. Generally speaking, when children were asked to complete a task, for example, producing the name of a referent, they had to make efforts to retrieve relevant information within their memory or mental lexicon. This retrieval practice would strengthen the newly formed representations. The improvement observed in TD group from Time 2 to Time 3 may be attribute to the effect of re-exposure or retrieval practice at Time 2 (James et al., 2023). The role of retrieval practice has been well-studied in the field of word learning and the facilitative role of retrieval practice (e.g., spaced repeated retrieval) during the training process has been observed in TD children and children with DLD (Leonard et al., 2022). The retrieval of newly learned words after the training (12-hour interim test) was found to be beneficial to the 24-hour retention of word forms for adults (McGregor, 2014). James et al. (2019) and James et al. (2023) also found a significant improvement from one-day delay to one-week delay of newly learned words in school-aged children. While more direct evidence is needed to examine the effect of one-day retrieval on the follow-up retention, in the current study, we cannot rule out this retrieval effect. Therefore, we cautiously speculate that the retrieval practice at Time 2 strengthened lexical representations of new words for TD children, thereby improving their performance at

Time 3. Similarly, the retrieval practice at Time 3 strengthened these representations again, allowing TD children to retain these representations even after one month.

Are individuals with ASD less sensitive to the effect of retrieval practice, especially spaced retrieval? To the best of our knowledge, there is no evidence from the field of word learning. Previous research that examined motor learning showed that the advantage of distributed practice (like spaced retrieval) was not observed in children with ASD (Wek & Husak, 1989). Based on the current analysis, the possibility that children with ASD would be less sensitive to the effect of spaced retrieval cannot be ruled out until empirical evidence is available.

For semantic knowledge, group differences were observed in all follow-up definition tasks. Again, sleep quality did not explain group differences. For TD children, the best overall definition performance was observed at Time 3, as evidenced by the significant increase from Time 1 to Time 3, and the significant decrease from Time 3 to Time 4. On the contrary, although statistical results showed that children with ASD performed similarly across the four time points, there was a tendency for decrease from Time 1 to Time 4. These differences may derive from their difficulties in retaining semantic representations.

To sum up, the analysis of retention performance of Mandarin-speaking children with ASD showed that their long-term retention (over one-week) of both phonological and semantic representations was problematic. Parent-reported sleep quality was related to the changes in retention performance one-day and one-week after training only in children with ASD. To fully understand these weaknesses, future studies can consider the following directions: firstly, following Fletcher et al. (2020), researchers should use more fine-grained measurements of sleep characteristics (e.g., polysomnography recordings) to analyze the relationship between these characteristics and children's retention performance; secondly, tracking children's sleep quality over a longer period of time to further investigate the direct relationship between sleep quality and children's retention performance; thirdly, the effects of one-day retrieval on short- and long-term retention should be examined directly in TD children, along with the effects of spaced retrieval during training and one-day retrieval in children with ASD.

### 4.3 The effect of semantic head on learning performance

The third research question addressed whether children with ASD showed sensitivity to the facilitative cues from word characteristics like TD children. Results showed that for children with ASD, the effect of semantic head was observed in neither the mispronunciation detection task and the referent-identification task during the training, nor the naming task and definition task after the training. These findings suggested that, unlike TD children, the presence of semantic heads did not effectively alleviate the processing load of new words for children with ASD at the encoding stage. At the retention stage, neither the interaction between group and type nor the three-way interaction between group, type and time was significant in both naming and definition task, suggesting that the effect of semantic head was not observed for both groups. In short, the presence of semantic head did not affect the word learning performance in Mandarin-speaking children with ASD.

The current results were consistent with our hypothesis and previous studies (Ferman, 2021; Floyd et al., 2021), suggesting that verbal children with ASD would be less sensitive to facilitative word characteristics. Why are children with ASD less sensitive to these characteristics? Considering the word characteristics that have been examined are meaning related, their deficits in lexical-sematic processing may be one of the possible answers (Naigles & Tek, 2017). Pastor-Cerezuela et al. (2016) found that school-aged children with ASD produced significantly fewer correct words and generated smaller clusters than age- and nonverbal IQ matched TD controls in the verbal semantic fluency task, wherein participants had to produce as many words from a certain category as they can within one or two minutes. Group differences were observed in this task even for adults with ASD who showed normal language development (Ehlen et al., 2020; Inokuchi & Kamio, 2013; Spek et al., 2009). These findings suggested that individuals with ASD may show persistent difficulty in processing categorical-related semantic information. In the current study, the presence of sematic heads in novel words provided learners with categorical information, which may be less useful for children with ASD.

Another possible answer to this question came from the procedural-declarative model of language learning (Ullman, 2016). Ferman (2021) applied this framework to discuss the finding that children with ASD who showed normal language development (ALN) were less efficient in applying derivational morphological knowledge for learning new words. Ferman (2021) designed two types of novel words, one type consisting of existing Hebrew morphological patterns and pseudo-root, and the other type consisting of nonexistent Hebrew morphological patterns and pseudo-root. Ferman (2021) proposed that learning words consisting of nonexistent-Hebrew morphological patterns mainly involved the declarative memory, since they did not follow any morphological word formation rules. On the contrary, learning words consisting of existing Hebrew morphological patterns involved both declarative and procedural memory, since they followed the morphological word formation rules. Based on this argument, Ferman (2021) further suggested that though not as efficient as TD children in applying derivational morphological knowledge, children with ALN showed similar and even better learning performance, which was consistent with the hypothesis that the weaker procedural system would be compensated by the intact declarative system (Walenski et al., 2014). The current study showed similarity with Ferman (2021) not only in the materials but also in the findings. OH words were not following familiar morphological word formation rules, whereas WH words were following morphological word formation rules. Although the effect of semantic heads was not observed in children with ASD, they did not differ significantly from TD controls in naming performance. Therefore, the insensitivity children with ASD exhibited to the semantic head might be attributed to their lack of awareness of this word formation rule, possibly stemming from their weaker procedural learning system.

## 5. Conclusion

In the current study, Mandarin-speaking children with and without ASD were taught novel words with and without semantic head. This group of children with ASD were significantly older than TD controls and showed delayed vocabulary development. They can establish the form-referent links and form phonological representations at the encoding stage. However, they show difficulties in retaining the phonological representations over a one-week delay. Parent-reported sleep quality was related to the changes in retention performance one-day and one-week after training only in children with ASD. Further investigations of fine-grained measurements of sleep characteristics (e.g., polysomnography recordings), long-term sleep quality, and the effects of retrieval practices on learning outcomes will be essential to explain their retention deficits. Children with ASD also showed extraordinary difficulties in the definition task at both encoding and retention stages. Further investigations using different measures (e.g., categorical selection task, cued definition task) will be critical to further probe their encoding and retention of semantic knowledge of new words. The effect of semantic head was not observed in verbal children with ASD at either the encoding or retention stage, a possible consequence of their deficits in lexical-semantic processing and procedural learning. Phonological memory and compound awareness explained group differences in selecting the correct word forms of new words and predicted learning performance at the encoding stage, with phonological memory being more important for verbal children with ASD in encoding new words.

# Chapter 4 General discussion

## 1 Key findings of this thesis

By teaching Mandarin-speaking children with and without ASD novel words with and without semantic head, and using multiple tasks across several time points to measure their learning outcomes, the current thesis yielded the following key findings. Firstly, the presence of semantic head assisted Mandarin-speaking TD children in forming the phonological representations of new words, while it did not influence the learning performance in Mandarin-speaking children with ASD. These findings highlight the critical role of word characteristics in word learning. In line with previous research (e.g., Hoover et al., 2010), the current findings suggested that as young as preschool age, TD children showed sensitivity to word characteristics in their native language and utilize these characteristics to enhance their word learning. This may help to account for their rapid vocabulary growth. On the other hand, verbal children with ASD were less sensitive to these word characteristics, which may partially contribute to their delayed vocabulary development. This insensitivity may be due to their differences in lexical-semantic processing and procedural learning. These findings support the necessity of taking word characteristics into consideration when constructing a theoretical model that accounts for word learning for children aged over three years.

Second, Study 2 identified a unique profile of word learning skills in verbal children with ASD. The ability to establish the form-referent links and form phonological representations are relatively intact in these children. However, producing the semantic characteristics of the newly learned words is extraordinarily challenging. Further, retention of phonological representations over one-week delay is problematic. These differences may account for the vocabulary delay of many children with ASD.

Third, following the suggestion of moving beyond mapping in the study of word learning (Wojcik et al., 2022), the current studies examined the retention of newly learned word. An overnight decrease in naming performance was observed in both groups of children. These patterns are inconsistent with previous studies that report overnight improvement in school-age children and may suggest developmental changes in retention patterns from preschool age to school age. Future studies should probe deeper into the potential developmental changes across age groups in retention performance, in particular between preschoolers and school-aged children.

Last but not least, the two studies confirmed the predictive role of phonological memory in word learning at the encoding stage for both TD and ASD children, underscoring its importance in this process. The current study was the first to document the predictive role of compound awareness in word learning research. A positive correlation between morphological awareness and vocabulary size has been commonly observed in school-aged children (Michaly & Prior, 2024; Sparks & Deacon, 2013). The current findings have deepened our understanding of the relationship between morphological awareness and vocabulary growth by providing evidence from children's word learning performance. These findings also suggested that in addition to general cognitive skills, meta-linguistic skills play a critical role in children's word learning. In addition to word characteristics, the role of learners' abilities should be considered when constructing a theory of word learning.

## 2 Clinical implications

Delayed vocabulary development is a critical clinical problem for many children with ASD. The findings from the current study yielded several clinical implications that may inform the study of vocabulary intervention for Mandarin-speaking children with ASD. Firstly, as a group, verbal children with ASD demonstrate the ability to establish the form-referent links and form phonological representations of new words, like their TD peers. When designing intervention sessions, clinicians can capitalize on these strengths to build in opportunities for children with ASD to learn phonological forms of new words and establish the mapping between forms and referents.

Secondly, the results of current studies suggest that Mandarin-speaking children with ASD are less sensitive to the role of semantic head in word learning. This difference may impact their learning of compound words. Given that compound words are common in Mandarin (about 65% of the Chinese lexicon), this difference could meaningfully impact the vocabulary development in Mandarin-speaking children with ASD. The ability to apply the semantic head to facilitate word learning was closely related to children's compound awareness. Compound awareness, or more broadly morphological awareness not only plays a role in word learning, but is also a significant predictor of children's literacy skills (Cheng et al., 2017). Therefore, we recommended conducting clinical research to examine whether training that targets the understanding of semantic heads would improve the word learning skills and vocabulary development in children with ASD who have delays in their vocabulary. One plausible approach could be that selecting a group of words that shared the same semantic head as the training materials for one session, such as "桃 花/tao2 hua1/peach flower", "梨 花/li2 hua1/pear flower", "樱 花/ying1 hua2/cherry flower" and so on. This training not only assists children in learning a specific semantic category but may also help them realize the function of semantic head and the lexical structures of Mandarin words. This approach would be similar to the training protocol applied by (Smith et al., 2002), who found that the training that grouped the objects with similar shapes to teach their names accelerated children's vocabulary development.

Thirdly, the definition task is challenging for all children, but extraordinarily so for children with ASD, implying that more emphasis can be placed on assisting children in producing semantic characteristics of new words. This training may help them establish and store detailed semantic representations in a more robust manner. Given the floor effect in the current definition task, many children with ASD may need prompts initially, such as sentences with blanks for colors, categories and other salient features of target words.

Lastly, the predictive role of phonological memory in learning new words was observed, suggesting that attention should be paid in the clinical contexts to phonological memory. Specifically, clinicians could reduce the phonological memory load during the vocabulary intervention, such as controlling the number of target words taught at each session. This approach could be beneficial in improving children's word learning performance.

## 3 Limitations and future research

The current two studies contribute to our word learning abilities in children with and without ASD. However, there are several limitations that should be addressed in further research. First of all, future studies are recommended to narrow down the age range of participants. The age range of children in the current study was wide (3 to 7 years), but the majority of children aged between 4 to 6 years (1 child aged 3;10, 2 children aged 3;11 and 1 child aged 7;02) and did not attend primary schools when they participated in the first session of this study. We acknowledge that the age range of 4 to 6 years is still wide, since children from K1 to K3 are included and children's language, cognitive and memory abilities develop rapidly during this stage. As an initial step in examining the effect of semantic head in Mandarin-speaking preschoolers, our results have suggested a facilitative effect for preschoolers as a whole. By examining children from different age groups, future studies can address questions such as when the semantic head starts to influence children's word learning and whether its importance changes as children grow older.

Second, the current findings indicated that the effect of semantic head was only observed at the encoding stage and was absent at the retention stage. Does this finding suggest that the role of semantic head is not so important since its facilitation cannot persist in the retention stage? The current studies examined the effect of semantic head mainly in preschoolers, and evidence from learners of other ages may help to address this question. James et al. (2019) found that the effect of neighborhood density was also absent in school-aged children's retention performance while it was observed in adults' retention performance. Further evidence from older children (i.e. school-aged) and adults is needed to gain a more comprehensive understanding of the semantic head effect.

Another limitation is that the current studies only examined one word characteristic (i.e., the presence of semantic head) in Mandarin-speaking children. The effect of word form characteristics (e.g., neighborhood density) has been well documented in English-speaking individuals, but this effect has not been examined in Mandarin-speaking individuals to the best our knowledge. Future research can investigate the role of word form characteristics in Mandarin-speaking children. Take phonological neighborhood density as an example. Unlike English words, Mandarin words are primarily disyllabic, with one phoneme corresponding to one syllable. Moreover, words are not easy to define in Chinese. Therefore, calculating the phonological neighborhood density for the entire word is challenging, whereas it is more appropriate to calculate the phonological neighborhood density for each syllable that constitutes the word. It will be interesting to explore whether phonological neighborhood density influences word learning performance in Mandarin-speaking children, and whether this effect is influenced by the position of the syllable. Moreover, future research can compare the effect of phonological neighborhood density with the effect of semantic head to examine which word characteristic is more important for Mandarin-speaking individuals.

Fourth, more efforts should be made to explore the mechanism underlying the memory evolution of newly learned words. Results of the current two studies suggest that developmental changes may be observed in overnight consolidation. Further studies should examine the retention of different aspects of word learning across various age groups to provide a clearer picture of the development of children's retention profiles.

It is worthwhile exploring the factors that would influence children's retention performance as well. The current design cannot rule out the influence of repeated tests, therefore, future research can include the time point of delayed tests as a between-subject variable to examine whether the timing of previous retrieval practices would influence the following retention performance. On the other hand, the absence of semantic head effect at the retention stage may also suggest that factors that influence encoding and retention performance may be different. Word characteristics was not the only factor that has been found to play a role only at the encoding stage. Pomper et al. (2022) also found that the effect of learning context (direct instruction vs indirect instruction) was not observed at the retention stage. These findings indicated that we need to consider the influencing factors at different stages of word learning, especially which factors would affect the retention performance of new words (Henderson et al., 2021; Wojcik, 2021).

Fifth, in terms of semantic learning, definitional performance of real words is needed to determine whether the difficulty exhibited by children with ASD in the definition task stems from their poor definition skills or from the failure to form sufficient semantic representations of new words. Future research can also consider using the cued definition task, the drawing task, or receptive measures of semantic knowledge (e.g., category selection task) to examine semantic representations of new words in children with ASD (McGregor et al., 2002; Pomper et al., 2022).

Language or vocabulary ability-matched younger TD controls should be included in future studies to illustrate whether these semantic-related difficulties in word learning exhibited by children with ASD are a developmental delay or a core deficit in their word learning abilities. Considering the task requirement, we recruited children aged around 4 years old at a minimum. Although the age range of the two groups was similar, children with ASD were about six months older than TD children on average and showed delayed vocabulary development. In this case, the language ability-matched TD controls will be younger than 4 years and may not be able to complete the current tasks. In line with the first point of this section, follow-up studies should narrow down the age range of participants and recruit the following three groups of children: 5-6 years children with ASD, age-matched TD children, language ability matched younger TD children (4-5 years old).

At a more general level, it is worthwhile examining the effect of learning context, for example, direct versus indirect instruction (Pomper et al., 2022), and rich versus sparse linguistic context (He et al., 2020) on ASD children's word learning performance. Furthermore, more attention can be paid to the learning mechanisms for other word classes (e.g., verb, spatial and descriptive terms) and multimodal learning (e.g., integration of gestural or tactile cues) in children with ASD.

Last but not least, future research can explore the relationship between other cognitive skills (e.g., sustained attention) and meta-linguistic skills (e.g., phonological awareness) and children's word learning performance. Taking a step further, intervention studies can be conducted to establish a causal relationship between well-documented predictors (e.g., phonological memory) and learning outcomes.

# Reference


Adlof, S. M., & Patten, H. (2017). Nonword repetition and vocabulary knowledge as predictors of children's phonological and semantic word learning. *Journal of speech, language, and hearing research*, *60*(3), 682-693. https://doi.org/10.1044/2016_JSLHR-L-15-0441

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., Welch, K., & Pickles, A. (2007). Patterns of Growth in Verbal Abilities Among Children With Autism Spectrum Disorder. *Journal of consulting and clinical psychology*, *75*(4), 594-604. https://doi.org/10.1037/0022-006X.75.4.594

Archibald, L. M. D. (2018). The reciprocal influences of working memory and linguistic knowledge on language performance: Considerations for the assessment of children with developmental language disorder. *Language, speech & hearing services in schools*, *49*(3), 424-433. https://doi.org/10.1044/2018_LSHSS-17-0094

Arunachalam, S., & Luyster, R. J. (2016). The integrity of lexical acquisition mechanisms in autism spectrum disorders: A research review. *Autism research*, *9*(8), 810-828. https://doi.org/10.1002/aur.1590

Bang, J. Y., & Nadig, A. (2020). An investigation of word learning in the presence of gaze: Evidence from school-age children with typical development or Autism Spectrum Disorder. *Cognitive development*, *54*, 100847. https://doi.org/10.1016/j.cogdev.2020.100847

Bao, H. (2023). *bruceR: Broadly Useful Convenient and Efficient R Functions*. R package version 2023.9. https://CRAN.R-project.org/package=bruceR

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

Bean Ellawadi, A., & McGregor, K. K. (2016). Children with ASD can use gaze to map new words. *International journal of language & communication disorders*, *51*(2), 212-218. https://doi.org/10.1111/1460-6984.12192

Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of child language*, *6*(2), 183-200. https://doi.org/10.1017/S0305000900002245

BjorK, R. A., & Kroll, J. F. (2015). Desirable Difficulties in Vocabulary Learning. *The American journal of psychology*, *128*(2), 241-252. https://doi.org/10.5406/amerjpsyc.128.2.0241

Carter, C. K., & Hartley, C. (2021). Are Children With Autism More Likely to Retain Object Names When Learning From Colour Photographs or Black-and-White Cartoons? *Journal of Autism and Developmental Disorders*, *51*(9), 3050-3062. https://doi.org/10.1007/s10803-020-04771-2

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological medicine*, *41*(3), 619-627. https://doi.org/10.1017/S0033291710000991

Chen, C., & Duan, M.-s. (2016). Part-of-Speech Annotation of Morphemes in Disyllabic Compounds. *Chinese Language Learning*, *01*, 30-41.

Chen, X., Hao, M., Geva, E., Zhu, J., & Shu, H. (2009). The role of compound awareness in Chinese children's vocabulary acquisition and character reading. *Reading & writing*, *22*(5), 615-631. https://doi.org/10.1007/s11145-008-9127-9

Chen, Y., & Liu, H.-M. (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in developmental disabilities*, *35*(1), 10-20. https://doi.org/10.1016/j.ridd.2013.10.010

Cheng, Y., Zhang, J., Li, H., Wu, X., Liu, H., Dong, Q., Li, L., Nguyen, T. P., Zheng, M., Zhao, Y., & Sun, P. (2017). Growth of Compounding Awareness Predicts Reading Comprehension in Young Chinese Students: A Longitudinal Study From Grade 1 to Grade 2. *Reading research quarterly*, *52*(1), 91-104. https://doi.org/10.1002/rrq.155

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the*

*Royal Society B: Biological Sciences*, *364*(1536), 3773-3800. https://doi.org/10.1098/rstb.2009.0111

Dosi, I., & Gavriilidou, Z. (2020). The Role of Cognitive Abilities in the Development of Definitions by Children with and Without Developmental Language Disorder. *Journal of psycholinguistic research*, *49*(5), 761-777. https://doi.org/10.1007/s10936-020-09711-w

Duff, D. (2019). Has Vocabulary Intervention Had an Effect? A Valid, Reliable, and (Fairly) Quick Outcome Measure for Semantic Knowledge. *Language, speech & hearing services in schools*, *50*(4), 506-517. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0134

Ehlen, F., Roepke, S., Klostermann, F., Baskow, I., Geise, P., Belica, C., Tiedt, H. O., & Behnia, B. (2020). Small Semantic Networks in Individuals with Autism Spectrum Disorder Without Intellectual Impairment: A Verbal Fluency Approach. *Journal of Autism and Developmental Disorders*, *50*(11), 3967-3987. https://doi.org/10.1007/s10803-020-04457-9

Ferman, S. (2021). Morpholexical Processes in Children with Autism Spectrum Disorder: Evidence from Artificial Word Learning. *Folia Phoniatrica et Logopaedica*, *73*, 1-15. https://doi.org/10.1159/000512849

Field, C., Lewis, C., & Allen, M. L. (2019). Referent selection in children with Autism Spectrum Condition and intellectual disabilities: Do social cues affect word-to-object or word-to-location mappings? *Research in developmental disabilities*, *91*, 103425-103425. https://doi.org/10.1016/j.ridd.2019.05.004

Fletcher, F. E., Knowland, V., Walker, S., Gaskell, M. G., Norbury, C., & Henderson, L. M. (2020). Atypicalities in sleep and semantic consolidation in autism. *Developmental science*, *23*(3), e12906-n/a. https://doi.org/10.1111/desc.12906

Floyd, S., Jeppsen, C., & Goldberg, A. E. (2021). Brief Report: Children on the Autism Spectrum are Challenged by Complex Word Meanings. *Journal of Autism and Developmental Disorders*, *51*(7), 2543-2549. https://doi.org/10.1007/s10803-020-04687-x

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of child language*, *44*(3), 677-694. https://doi.org/10.1017/S0305000916000209

Gallucci, M. (2023). *GAMLj-General Analyses for the Linear Model in Jamovi*. https://gamlj.github.io/index.html

Gavriilidou, Z., Markos, A., & Dourou, C. (2022). The Interaction of Variables Affecting Definitional Skills: Extending Previous Research on Word Definitions. *Journal of Language and Education*, *8*(3), 65-80. https://doi.org/10.17323/jle.2022.12901

Gladfelter, A., & Barron, K. L. (2020). How children with autism spectrum disorder, developmental language disorder, and typical language learn to produce global and local semantic features. *Brain sciences*, *10*(4), 231. https://doi.org/10.3390/brainsci10040231

Gladfelter, A., & Goffman, L. (2018). Semantic richness and word learning in children with autism spectrum disorder. *Developmental science*, *21*(2), e12543-n/a. https://doi.org/10.1111/desc.12543

Gómez, R. L., & Edgin, J. O. (2015). Sleep as a Window Into Early Neural Development: Shifts in Sleep-Dependent Learning Effects Across Early Childhood. *Child development perspectives*, *9*(3), 183-189. https://doi.org/10.1111/cdep.12130

Gordon, K. R., Lowry, S. L., Ohlmann, N. B., & Fitzpatrick, D. (2022). Word Learning by Preschool-Age Children: Differences in Encoding, Re-Encoding, and Consolidation across Learners during Slow Mapping. *Journal of speech, language, and hearing research*, *65*(5), 1956-1977. https://doi.org/10.1044/2022_JSLHR-21-00530

Gordon, K. R., McGregor, K. K., Waldier, B., Curran, M. K., Gomez, R. L., & Samuelson, L. K. (2016). Preschool Children's Memory for Word Forms Remains Stable Over Several Days, but Gradually Decreases after 6 Months. https://doi.org/10.3389/fpsyg.2016.01439

Gordon, K. R., Storkel, H. L., Lowry, S. L., & Ohlmann, N. B. (2021). Word Learning by Preschool-Age Children With Developmental Language Disorder: Impaired

Encoding and Robust Consolidation During Slow Mapping. *Journal of speech, language, and hearing research*, *64*(11), 4250-4270. https://doi.org/10.1044/2021_JSLHR-21-00046

Gray, S., Lancaster, H., Alt, M., Hogan, T. P., Green, S., Levy, R., & Cowan, N. (2020). The structure of word learning in young school-age children. *Journal of speech, language, and hearing research*, *63*(5), 1446-1466. https://doi.org/10.1044/2020_JSLHR-19-00186

Gray, S., Pittman, A., & Weinhold, J. (2014). Effect of phonotactic probability and neighborhood density on word-learning configuration by preschoolers with typical development and specific language impairment. *Journal of speech, language, and hearing research*, *57*(3), 1011-1025. https://doi.org/10.1044/2014_JSLHR-L-12-0282

Gray, S. I., Levy, R., Alt, M., Hogan, T. P., & Cowan, N. (2022). Working Memory Predicts New Word Learning Over and Above Existing Vocabulary and Nonverbal IQ. *Journal of speech, language, and hearing research*, *65*(3), 1044-1069. https://doi.org/10.1044/2021_JSLHR-21-00397

Gupta, P., & Tisdale, J. (2009). Word learning, phonological short-term memory, phonotactic probability and long-term memory: towards an integrated framework. *Philosophical transactions of the Royal Society of London. Series B. Biological sciences*, *364*(1536), 3755-3771. https://doi.org/10.1098/rstb.2009.0132

Hani, H. B., Gonzalez-Barrero, A. M., & Nadig, A. S. (2013). Children's referential understanding of novel words and parent labeling behaviors: similarities across children with and without autism spectrum disorders. *Journal of child language*, *40*(5), 971-1002. https://doi.org/10.1017/S0305000912000426

Hartley, C., Bird, L.-A., & Monaghan, P. (2019). Investigating the relationship between fast mapping, retention, and generalisation of words in children with autism spectrum disorder and typical development. *Cognition*, *187*, 126-138. https://doi.org/10.1016/j.cognition.2019.03.001

Hartley, C., Bird, L.-A., & Monaghan, P. (2020). Comparing cross-situational word learning, retention, and generalisation in children with autism and typical development. *Cognition*, *200*, 104265-104265. https://doi.org/10.1016/j.cognition.2020.104265

He, A. X., & Arunachalam, S. (2017). Word learning mechanisms. *Wiley interdisciplinary reviews. Cognitive science*, *8*(4), e1435-n/a. https://doi.org/10.1002/wcs.1435

He, A. X., Kon, M., & Arunachalam, S. (2020). Linguistic Context in Verb Learning: Less is Sometimes More. *Language Learning and Development*, *16*(1), 22-42. https://doi.org/10.1080/15475441.2019.1676751

Henderson, L., Powell, A., Gareth Gaskell, M., & Norbury, C. (2014). Learning and consolidation of new spoken words in autism spectrum disorder. *Developmental science*, *17*(6), 858-871. https://doi.org/10.1111/desc.12169

Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of experimental child psychology*, *116*(3), 572-592. https://doi.org/10.1016/j.jecp.2013.07.004

Henderson, L. M., van Rijn, E., James, E., Walker, S., Knowland, V. C. P., & Gareth Gaskell, M. (2021). Timing storytime to maximize children's ability to retain new vocabulary. *Journal of experimental child psychology*, *210*, 105207-105207. https://doi.org/10.1016/j.jecp.2021.105207

Henderson, L. M., Weighall, A. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children: Sleep-associated consolidation of vocabulary. *Developmental science*, *15*(5), 674-687. https://doi.org/10.1111/j.1467-7687.2012.01172.x

Hoff-Ginsberg, E. (2014). *Language development* (5th ed.). Wadsworth Cengage Learning.

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of memory and language*, *63*(1), 100-116. https://doi.org/10.1016/j.jml.2010.02.003

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

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of sleep research*, *24*(5), 503-509. https://doi.org/10.1111/jsr.12306

Hou, W., & Su, Y. E. (2022). The influence of atypical attention and memory on vocabulary delay in children with autism spectrum disorder. *Advances in Psychological Science*, *30*(11), 2558-2569. https://doi.org/10.3724/SP.J.1042.2022.02558

Inokuchi, E., & Kamio, Y. (2013). Qualitative analyses of verbal fluency in adolescents and young adults with high-functioning autism spectrum disorder. *Research in autism spectrum disorders*, *7*(11), 1403-1410. https://doi.org/10.1016/j.rasd.2013.08.010

James, E., Gaskell, M. G., & Henderson, L. M. (2019). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental science*, *22*(3), e12776-n/a. https://doi.org/10.1111/desc.12776

James, E., Gaskell, M. G., Murphy, G., Tulip, J., & Henderson, L. M. (2023). Word learning in the context of semantic prior knowledge: evidence of interference from feature-based neighbours in children and adults. *Language, Cognition and Neuroscience*, *38*(2), 157-174. https://doi.org/10.1080/23273798.2022.2102198

James, E., Gaskell, M. G., Weighall, A., & Henderson, L. (2017). Consolidation of vocabulary during sleep: The rich get richer? *Neuroscience and biobehavioral reviews*, *77*, 1-13. https://doi.org/10.1016/j.neubiorev.2017.01.054

Jiang, Z., Sun, J., & Wei, X. (2011). The effects of semantic relation between morphemes in compounds on preschool Chinese children's compound understanding. *Applied Linguistics*, *3*, 87-95.

Jing, W., & Fang, J. (2014). Brief report: Do children with autism gather information from social contexts to aid their word learning? *Journal of Autism and*

*Developmental Disorders*, *44*(6), 1478-1482. https://doi.org/10.1007/s10803-013-1994-5

Jing, W., Fang, J., & Zhao, W. (2014). Word Learning with Multiple Cues in Children with Autistic Spectrum Disorder: Evidence from Eye Movements. *Acta Psychologica Sinica,*, *43*, *385-395*.

Jing, W., Tian, Q., Guo, W., & Fang, J. (2013). Word Learning with Social Attention Cues in Children with Autism: Evidence from Eye Movements. *Journal of Psychological Science*, *36*(6), 1421-1127. https://doi.org/10.16719/j.cnki.1671-6981.2013.06.003.

Kang, D., Li, J., & Cai, S. (2022). Preschool children's sleep problem and language disorder. *Advances in Psychological Science*, *30*(6), 1270-1281. https://doi.org/10.3724/SP.J.1042.2022.01270

Knowland, V. C. P., Fletcher, F., Henderson, L.-M., Walker, S., Norbury, C. F., & Gaskell, M. G. (2019). Sleep Promotes Phonological Learning in Children across Language and Autism Spectra. *Journal of speech, language, and hearing research*, *62*(12), 4235-4255. https://doi.org/10.1044/2019_JSLHR-S-19-0098

Lam, B. P.-W., & Sheng, L. (2016). The development of morphological awareness in young bilinguals: Effects of age and L1 background. *Journal of speech, language, and hearing research*, *59*(4), 732-744. https://doi.org/10.1044/2015_JSLHR-L-14-0171

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

Leonard, L. B., Kueser, J. B., Deevy, P., Haebig, E., Karpicke, J. D., & Weber, C. (2022). The contributions of immediate retrieval and spaced retrieval to word learning in preschoolers with developmental language disorder. *Autism & developmental language impairments*, *7*, 239694152210776. https://doi.org/10.1177/23969415221077652

Li, Y. (2004). *儿童语言的发展(第 2 版) [The development of child langauge 2nd]*. Central China Normal University Press.

Li, Y., & Zhu, J. (2012). *Wechsler Intelligence Scale for Children Fourth Edition (Chinese version)*. King May Psychology Assessment, Ltd. with the license by NSC Person, lnc.

Liang, D. (2017). *An Introduction to Children's Language Disorders*. The Commercial Press.

Liu, S. Y., Wang, X., Chen, Q., Chen, J. J., Jin, C. K., Zhan, X. L., Guo, C. H., Li, X. H., Lin, L. Z., & Jing, J. (2022). The validity and reliability of the simplified Chinese version of the Social Communication Questionnaire. *Autism research*, *15*(9), 1732-1741. https://doi.org/10.1002/aur.2726

Liu, Z., Wang, G., Tang, H., Wen, F., & Li, N. (2014). Reliability and validity of the Children's Sleep Habits Questionnaire in preschool-aged Chinese children. *Sleep and biological rhythms*, *12*(3), 187-193. https://doi.org/10.1111/sbr.12061

Lucas, R., & Norbury, C. F. (2014). Orthography facilitates vocabulary learning for children with autism spectrum disorders (ASD). *Quarterly journal of experimental psychology (2006)*, *67*(7), 1317-1334. https://doi.org/10.1080/17470218.2013.859714

Lucas, R., Thomas, L., & Norbury, C. F. (2017). Can Children with Autism Spectrum Disorders Learn New Vocabulary From Linguistic Context? *Journal of Autism and Developmental Disorders*, *47*(7), 2205-2216. https://doi.org/10.1007/s10803-017-3151-z

Luyster, R., & Lord, C. (2009). Word Learning in Children With Autism Spectrum Disorders. *Developmental psychology*, *45*(6), 1774-1786. https://doi.org/10.1037/a0016223

Ma, X., & Liang, D. (2019). The Reasons for the Deficiencies of Chinese Children with Specific Language Impairment in Fast Mapping Word Learning: Phonological Working Memory or Semantic Representation. *Contemporary Linguistics*, *21*(1), 104-118.

Mason, G. M., Lokhandwala, S., Riggins, T., & Spencer, R. M. C. (2021). Sleep and human cognitive development. *Sleep medicine reviews*, *57*, 101472-101472. https://doi.org/10.1016/j.smrv.2021.101472

McGhee, R., & Ehrler, D. (2008). *Primary Test of Nonverbal Intelligence Technical Manual*.

McGregor, K. K. (2014). What a Difference a Day Makes: Change in Memory for Newly Learned Word Forms over 24 Hours. *Journal of speech, language, and hearing research*, *57*(5), 1842-1850. https://doi.org/10.1044/2014_JSLHR-L-13-0273

McGregor, K. K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2020). The fast-mapping abilities of adults with developmental language disorder. *Journal of speech, language, and hearing research*, *63*(9), 3117-3129. https://doi.org/10.1044/2020_JSLHR-19-00418

McGregor, K. K., Friedman, R. M., Reilly, R. M., & Newman, R. M. (2002). Semantic Representation and Naming in Young Children. *Journal of speech, language, and hearing research*, *45*(2), 332-346. https://doi.org/10.1044/1092-4388(2002/026)

McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of speech, language, and hearing research*, *60*(10), 2891-2905. https://doi.org/10.1044/2017_JSLHR-L-17-0031

McGregor, K. K., Rost, G., Arenas, R., Farris-Trimble, A., & Stiles, D. (2013). Children with ASD can use gaze in support of word recognition and learning. *Journal of child psychology and psychiatry*, *54*(7), 745-753. https://doi.org/10.1111/jcpp.12073

McGregor, K. K., Sheng, L., & Ball, T. (2007). Complexities of Expressive Word Learning Over Time. *Language, speech & hearing services in schools*, *38*(4), 353-364. https://doi.org/10.1044/0161-1461(2007/037)

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*(4), 831-877. https://doi.org/10.1037/a0029872

Michaly, T., & Prior, A. (2024). Spelling morphology in Hebrew: Comparing monolingual and bilingual children. *Reading & writing*.

https://doi.org/10.1007/s11145-024-10582-z

Montgomery-Downs, H. E., O'Brien, L. M., Gulliver, T. E., & Gozal, D. (2006). Polysomnographic characteristics in normal preschool and early school-aged children. *Pediatrics*, *117*(3), 741-753.

Naigles, L. R. (2013). Input and language development in children with autism. *Seminars in speech and language*, *34*(4), 237-248. https://doi.org/10.1055/s-0033-1353446

Naigles, L. R., & Tek, S. (2017). 'Form is easy, meaning is hard' revisited: (re) characterizing the strengths and weaknesses of language in children with autism spectrum disorder. *Wiley interdisciplinary reviews. Cognitive science*, *8*(4), e1438-n/a. https://doi.org/10.1002/wcs.1438

Neergaard, K. D., Huang, C.-R., & Siew, C. (2019). Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation. *Complexity*, *2019*(1). https://doi.org/10.1155/2019/6979830

Neergaard, K. D., Xu, H., German, J. S., & Huang, C.-R. (2022). Database of word-level statistics for Mandarin Chinese (DoWLS-MAN). *Behavior research methods*, *54*(2), 987-1009. https://doi.org/10.3758/s13428-021-01620-7

Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: Word learning in autistic disorders. *Neuropsychologia*, *48*(14), 4012-4019. https://doi.org/10.1016/j.neuropsychologia.2010.10.015

Nowosad, J. (2018). *'CARTOColors' Palettes*. R package version 1.0.0. https://jakubnowosad.com/rcartocolor/

OpenAI. (2023). *GPT-4o [Large language model]*. In OpenAI. https://www.openai.com/

Owens, R. E. (2020). *Language development : an introduction* (Tenth edition. ed.). Pearson Education, Inc.

Pastor-Cerezuela, G., Fernández-Andrés, M.-I., Feo-álvarez, M., & González-Sala, F. (2016). Semantic verbal fluency in children with and without Autism Spectrum Disorder: Relationship with chronological age and IQ. *Frontiers in psychology*, *7*, 921-921. https://doi.org/10.3389/fpsyg.2016.00921

Peng, G., & Chen, F. (2020). Speech Development in Mandarin-Speaking Children. In H. M. Liu, F. M. Tsao, & P. Li (Eds.), *Speech Perception, Production and Acquisition: Multidisciplinary approaches in Chinese languages* (pp. 219-242). Springer Singapore. https://doi.org/10.1007/978-981-15-7606-5_12

Pomper, R., McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Ohlmann, N. (2022). Direct Instruction Improves Word Learning for Children With Developmental Language Disorder. *Journal of speech, language, and hearing research*, *65*(11), 4228-4249. https://doi.org/10.1044/2022_JSLHR-22-00300

Posit team. (2024). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. http://www.posit.co/

Pyke, W., Lunau, J., & Javadi, A.-H. (2024). Does difficulty moderate learning? A comparative analysis of the desirable difficulties framework and cognitive load theory. *Quarterly journal of experimental psychology (2006)*, 17470218241308143. https://doi.org/10.1177/17470218241308143

Quine, W. V. (1960). *Word and object*. Technology Press of the Massachusetts Institute of Technology.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing,Vienna, Austria. https://www.R-project.org/

Rothwell, C., Westermann, G., & Hartley, C. (2023). How do Autistic and Neurotypical Children's Interests Influence their Accuracy During Novel Word Learning? *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-023-06066-8

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science (American Association for the Advancement of Science)*, *274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926

Samuelson, L. K. (2021). Toward a Precision Science of Word Learning: Understanding Individual Vocabulary Pathways. *Child development perspectives*, *15*(2), 117-124. https://doi.org/10.1111/cdep.12408

Schimke, E. A. E., Angwin, A. J., Gomersall, S. R., & Copland, D. A. (2023). The effect of sleep and semantic information on associative novel word learning. *Memory (Hove)*, *31*(2), 234-246. https://doi.org/10.1080/09658211.2022.2142243

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What Is Learned under Difficult Conditions Is Hard to Forget: Contextual Interference Effects in Foreign Vocabulary Acquisition, Retention, and Transfer. *Journal of memory and language*, *46*(2), 419-440. https://doi.org/10.1006/jmla.2001.2813

Schreck, K. A., & Richdale, A. L. (2020). Sleep problems, behavior, and psychopathology in autism: inter-relationships across the lifespan. *Current Opinion in Psychology*, *34*, 105-111. https://doi.org/10.1016/j.copsyc.2019.12.003

Sheng, L., Zheng, L., Chen, S., Wang, D., & Lam, B. P. W. (in development). *Development of a Quick Vocabulary Assessment Tool for Mandarin Chinese: The Mandarin Online Oral Vocabulary Screener*.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object Name Learning Provides On-the-Job Training for Attention. *Psychological science*, *13*(1), 13-19. https://doi.org/10.1111/1467-9280.00403

Spanò, G., Gómez, R. L., Demara, B. I., Alt, M., Cowen, S. L., & Edgin, J. O. (2018). REM sleep in naps differentially relates to memory consolidation in typical preschoolers and children with Down syndrome. *Proceedings of the National Academy of Sciences - PNAS*, *115*(46), 11844-11849. https://doi.org/10.1073/pnas.1811488115

Sparks, E., & Deacon, S. H. (2013). Morphological awareness and vocabulary acquisition: A longitudinal examination of their relationship in English-speaking children. *Applied Psycholinguistics*, *36*(2), 299-321. https://doi.org/10.1017/s0142716413000246

Spek, A., Schatorjé, T., Scholte, E., & van Berckelaer-Onnes, I. (2009). Verbal fluency in adults with high functioning autism or Asperger syndrome. *Neuropsychologia*, *47*(3), 652-656. https://doi.org/10.1016/j.neuropsychologia.2008.11.015

Storkel, H. L. (2001). Learning New Words: Phonotactic Probability in Language

Development. *Journal of speech, language, and hearing research*, *44*(6), 1321-1337. https://doi.org/10.1044/1092-4388(2001/103)

Storkel, H. L. (2015). Learning from input and memory evolution: Points of vulnerability on a pathway to mastery in word learning. *International journal of speech language pathology*, *17*(1), 1-12. https://doi.org/10.3109/17549507.2014.987818

Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S.-Y. (2013). The Effect of Incremental Changes in Phonotactic Probability and Neighborhood Density on Word Learning by Preschool Children. *Journal of speech, language, and hearing research*, *56*(5), 1689-1700. https://doi.org/10.1044/1092-4388(2013/12-0245)

Storkel, H. L., & Hoover, J. R. (2011). The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *Journal of child language*, *38*(3), 628-643. https://doi.org/10.1017/S0305000910000176

Su, Y., Naigles, L. R., & Su, L.-Y. (2018). Uneven Expressive Language Development in Mandarin-Exposed Preschool Children with ASD: Comparing Vocabulary, Grammar, and the Decontextualized Use of Language via the PCDI-Toddler Form. *Journal of Autism and Developmental Disorders*, *48*(10), 3432-3448. https://doi.org/10.1007/s10803-018-3614-x

Sukenik, N., & Tuller, L. (2023). Lexical Semantic Knowledge of Children with ASD—a Review Study. *Review Journal of Autism and Developmental Disorders*, *10*(1), 130-143. https://doi.org/10.1007/s40489-021-00272-9

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Tardif, T., Fletcher, P., Zhang, Z., & Weilan, L. (2008). *汉语沟通发展使用手册 [The Manual of Chinese Communicative Development Inventory]* Peking University Medical Press.

Taylor, I., & Taylor, M. M. (2014). *Writing and literacy in Chinese, Korean and Japanese* (Revised edition. ed.). John Benjamins Publishing Company.

The jamovi project. (2024). *jamovi (Version 2.4.14) [Computer Software]*. Retrieved from https://www.jamovi.org.

To, C. K. S., Stokes, S., Man, Y., & T'sou, B. (2013). An Analysis of Noun Definition in Cantonese. *Language and speech*, *56*(1), 105-124. https://doi.org/10.1177/0023830912440794

Ullman, M. T. (2016). The Declarative/Procedural Model. In *Neurobiology of Language* (pp. 953-968). https://doi.org/10.1016/b978-0-12-407794-2.00076-6

Vlach, H. A., & Sandhofer, C. M. (2014). Retrieval Dynamics and Retention in Cross‐Situational Statistical Word Learning. *Cognitive science*, *38*(4), 757-774. https://doi.org/10.1111/cogs.12092

Walenski, M., Mostofsky, S. H., & Ullman, M. T. (2014). Inflectional morphology in high-functioning autism: Evidence for speeded grammatical processing. *Research in autism spectrum disorders*, *8*(11), 1607-1621. https://doi.org/10.1016/j.rasd.2014.08.009

Walton, K. M., & Ingersoll, B. R. (2013). Expressive and receptive fast-mapping in children with autism spectrum disorders and typical development: The influence of orienting cues. *Research in autism spectrum disorders*, *7*(6), 687-698. https://doi.org/10.1016/j.rasd.2013.02.012

Wang, Y., Zhang, Y.-b., Liu, L.-l., Cui, J.-f., Wang, J., Shum, D. H. K., van Amelsvoort, T., & Chan, R. C. K. (2017). A Meta-Analysis of Working Memory Impairments in Autism Spectrum Disorders. *Neuropsychology Review*, *27*(1), 46-61. https://doi.org/10.1007/s11065-016-9336-y

Wehren, A., Lisi, R. D., & Arnold, M. (1981). The development of noun definition. *Journal of child language*, *8*(1), 165-175. https://doi.org/10.1017/S0305000900003081

Wek, S. R., & Husak, W. S. (1989). Distributed and Massed Practice Effects on Motor Performance and Learning of Autistic Children. *Perceptual and motor skills*, *68*(1), 107-113. https://doi.org/10.2466/pms.1989.68.1.107

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wojcik, E. H. (2013). Remembering new words: integrating early memory development into word learning. *Frontiers in psychology*, *4*, 151-151. https://doi.org/10.3389/fpsyg.2013.00151

Wojcik, E. H. (2021). The Effect of the Number of Labeled Objects on Novel Referent Selection Across Short and Long Time Delays. *Language Learning and Development*, *17*(4), 411-424. https://doi.org/10.1080/15475441.2021.1909483

Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word learning research should move beyond mapping. *Wiley interdisciplinary reviews. Cognitive science*, *13*(4), e1596-n/a. https://doi.org/10.1002/wcs.1596

Wucailu Autism Research Centre. (2024). *中国孤独症教育康复行业发展状况报告（Ⅴ）[China Autism Education and Rehabilitation Status Report-Ⅴ]*. Guangming Daily Publishing House.

Yang, J., Zheng, L., Gu, W., Hu, H., Feng, C., & Sheng, L. (2021). Research on the Development Characteristics of Chinese Nouns in Children with ASD. *Chinese Journal of Special Eduction*, *250*(4), 42-49.

Yu, J., & Sheng, L. (in preparation). The effect of word characteristics on word learning: Does the presence of semantic head facilitate novel word learning in Mandarin-speaking children?