Wang, Jiachen (2026) *The role of embedded Artificial Intelligence learning applications in Chinese children's second language acquisition.* PhD thesis.

The role of embedded Artificial Intelligence learning applications in Chinese children's second language acquisition

Jiachen Wang

A Thesis Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy (PhD)

School of Education
College of Social Science
University of Glasgow

May 2026

Abstract

In recent years, the integration of technology-enhanced learning in early childhood education has attracted growing attention, being recognised as a promising approach to improving learning outcomes and promoting educational equity. In this context, researchers have increasingly explored how diverse resources such as multimedia, gamified learning, and artificial intelligence (AI) can be effectively integrated into second language acquisition (SLA) processes, particularly among young learners. Within the context of China's ongoing educational reform, early primary education is experiencing pedagogical innovation, creating heightened demand for AI-powered educational tools. Despite the increasing adoption of AI in educational settings, empirical evidence on its application in early childhood second language (L2) learning remains limited. Prior studies have demonstrated specific advantages of AI tools in providing personalised, interactive learning experiences; however, significant questions exist regarding their effectiveness in enhancing receptive vocabulary development, fostering learner motivation, and supporting enjoyment of learning among young children. Moreover, as central figures in children's early learning environments, parents' and caregivers' perceptions of AI-based applications constitute an essential dimension of educational technology integration that requires further investigation.

To address these gaps, the study adopts a mixed-methods approach to explore the effects of a multimodal AI-based language learning application (Zebra AI), which integrates Automatic Speech Recognition (ASR), adaptive feedback, and learning analytics, on English as a second language development among young Chinese learners aged 5 to 7. The research is theoretically grounded in SLA theories, Multimedia Learning Theory (MLT), Self-Determination Theory (SDT), and the Technology Acceptance Model. A quasi-experimental design was employed, supported by post-intervention questionnaires and thematic analysis of qualitative interviews.

Participants (*N* = 85) were children aged 5 to 7 years, recruited from two primary schools in central China and randomly assigned to either an experimental group (using the Zebra AI application) or a control group (using a non-educational entertainment app). Over a 12-week intervention period, children's receptive vocabulary was assessed using the Peabody Picture Vocabulary Test–Fifth Edition (PPVT-5), while cognitive skills,

including working memory (WM) and theory of mind (ToM), were measured through standardised tasks. Data were analysed using mixed-design ANOVA and regression modelling.

The quantitative findings revealed significant improvements in receptive vocabulary in the experimental group compared to the control group [F (1,83) = 52.42, p <.001, η^2_p = .387], along with significant gains in WM (p < .001). No statistically significant differences were observed in ToM performance. Regression analysis demonstrated that the model significantly explained 44.8% of the variance in receptive vocabulary growth (R^2adj = .448, p < .001). Both the group variable (β = .586, p < .001) and age (β = .240, p = .005) were significant predictors of vocabulary gains. Amongst the core cognitive mechanisms, growth in WM exhibited a marginally significant positive predictive trend (p = .051), whereas the impact of ToM did not reach statistical significance. These findings partially support the hypothesis regarding the relationship between WM and vocabulary acquisition, highlighting the potential contribution of associative learning mechanisms to language development under controlled experimental conditions.

Data from questionnaires measuring children's enjoyment with app-based learning indicated high satisfaction rating for the AI application's usability, visual appeal, and motivational features. Gamified elements, speech recognition, and interactive animation were identified as key factors in sustaining learner engagement.

Thematic analysis of semi-structured interviews with parents and caregivers (N = 8) supported these results, revealing observed improvements in children's vocabulary, motivation, and self-directed learning. Parents and caregivers also reported advantages such as flexibility, reduced homework supervision stress, and cost-effectiveness, while expressing concerns about screen time, device compatibility issues, inappropriate content, and low digital literacy. They suggested future design should support offline access, dialects, peer-support and parent-caregiver training to enhance usability in diverse early childhood learning contexts.

This study contributes to the expanding literature in Intelligent Computer-Assisted Language Learning (ICALL), demonstrating empirical evidence for the effectiveness of AI-enhanced tools in early SLA education through the specific application of Zebra AI.

However, it is important to note that these findings are specific to the algorithmic and pedagogical design of the tested application and may not be directly applied to other AI tools with different functions. It confirms this technology's capacity to enhance both learning outcomes and engagement, offering insights for educators, designers, and policymakers within the context of specific design features, to promote effective, equitable early language learning in digital learning environments.

Table of Contents

List of Figures

List of Tables

Statement on the Use of Artificial Intelligence

I acknowledge the use of ChatGPT 4.0 (Open AI, https://chat.openai.com) as a tool to proofread the final version of this work. Its assistance was limited to the identification of spelling, punctuation, grammatical, and syntactical errors.

Statement on any conflict of interest

I declare that there is no conflict of interest in relation to this research. I have no financial, professional, or personal affiliation with Zebra AI or its parent company. The use of Zebra AI in this research is solely for academic purposes, and I have not received any funding, sponsorship, or other form of support from the company.

Acknowledgement

My appreciation goes to all those who have helped me and supported me throughout this research.

First and foremost, I would like to express my gratitude to my family. My mother, who I loved more than anything in the world. And I am also deeply grateful for my father's financial support over the past four years. I could not have come this far without them.

Secondly, I would like to extend my deepest thanks to my supervisors, Dr Joanna Wincenciak and Dr Christopher Hand. Their outstanding supervision, guidance, and support have been invaluable throughout this journey. I am especially thankful for their belief in my potential, even at times when I doubted myself. It has been an honour to be supervised by them and to have had the opportunity to learn from them.

I would also like to extend my sincere thanks to all those who participated in this study, including the 85 children and the 8 parents and caregivers who kindly took part. Their engagement, enthusiasm, and generous contribution of time were invaluable to the success of this research.

As well, I would like to express my thanks to all my close friends in China and the UK. Thanks for supporting me mentally throughout the journey. I am especially grateful to my dearest friend, Xinxin—it was your patience, your willingness to listen, and your constant companionship that helped me overcome the most challenging times. Your presence reminded me that I was never alone, and for that, I will always be thankful.

Lastly, I would like to thank my beloved cat, Dobby. Thank you for your soft, furry companionship throughout the writing process — whether sitting quietly by my side during long nights or simply being there when I needed comfort. You may not understand my words, but you've always sensed my moods. May you always be happy, in your own mischievous and curious way.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed Name: Jiachen Wang

Signature:

List of Abbreviations

AI	Artificial Intelligence
ASR	Automatic Speech Recognition
ATT	Attitude toward Using
CALL	Computer Assisted Language Learning
CTML	Cognitive Theory of Multimedia Learning
EFL	English as a Foreign Language
HCI	Human-computer Interaction
HQ	Hedonic Quality
ICALL	Intelligent Computer-Assisted Language Learning
ICT	Information Communication Technology
ITS	Intelligent Tutoring System
L1	First Language
L2	Second Language
LAD	Language Acquisition Device
MALL	Mobile Assisted Language Learning
MALL	Mobile-Assisted Language Learning
MLT	Multimedia Learning Theory
NLP	Natural Language Processing
PEoU	Perceived Ease of Use
PPVT	Peabody Picture Vocabulary Test
PQ	Pragmatic Quality
PSTM	Phonological Short-Term Memory
PU	Perceived Usefulness
SDT	Self-Determination Theory
SLA	Second Language Acquisition
SRS	Spaced Repetition System
STEM	Science, Technology, Engineering, and Mathematics
TA	Thematic Analysis
TAM	Technology Acceptance Model
ToM	Theory of Mind
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UX	User experience

VR Virtual Reality

WM Working Memory (operationalised through an associative memory measure in this study)

XR Extended Reality

ZPD Zone of Proximal Development

Chapter 1. INTRODUCTION

1.1 Research Background

With the rapid development of AI across various areas, education has emerged as a key field for its deep integration and transformative potential. As an advanced digital technology, AI offers data-driven decision-making, adaptive learning paths, and intelligent feedback systems that are reshaping educational practice (Bunz & Janciute, 2018). Globally, governments have positioned AI at the core of education reform strategies. Countries such as the United States, the United Kingdom, Japan, and Singapore have systematically integrated AI into their national education blueprints, aiming to enhance accessibility, personalisation, and equity in learning. For instance, the U.S. Department of Education (2023) emphasises learner-centred AI to support inclusive education. Similarly, the UK's "EdTech Strategy" (2019), Japan's "Basic Plan for the Promotion of Education" (Ministry of Education, 2020), and Singapore's "Smart Nation" initiative (Smart Nation Office, 2022) all reflect strong national commitments to AI-enabled learning. International organisations have also expressed strong consistent with these priorities. UNESCO highlights AI's critical role not only in improving learning quality but also in promoting inclusive and equitable education systems (Holmes et al., 2022).

Within this global trend, China has demonstrated strong policy leadership by incorporating AI into national education strategies, such as the New Generation AI Development Plan (China, 2017), the AI + Internet Action Plan, and Education Modernisation 2035 (China, 2019). These policies advocate for the creation of intelligent learning environments and personalised education through AI integration. As national strategies continue to advance, the impact of AI in education has gradually extended into home-based learning contexts, a trend that has become particularly evident since the introduction of the "Double Reduction" policy. Specifically, due to the intensifying problems of excessive academic pressure, unequal distribution of educational resources, and the commercialisation of private tutoring, China implemented the "Double Reduction" policy in 2021 (http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202107/t20210724_546576.html), targeting primary and lower secondary school students. The policy strictly limits the private tutoring which prompting many parents and students to seek alternative learning tools to meet their need for improved academic performance. In this context, educational applications have rapidly emerged. A 2020 report found that 29 K12 applications collectively released 103 educational apps with over 17.25 billion downloads across major Android markets (Jiemian, 2020). These data reflect an increase in Chinese families' demand for learning apps.

Language learning, particularly English learning for children, has become as one of the key focus areas, as early English proficiency is increasingly considered to be a foundational skill for future academic success and global competitiveness. Several open-market applications have applied AI in the context of children's language learning. Unlike official school applications, these are open to the public and independently chosen by parents and learners. For this study, the focus is specifically on the Zebra AI (https://banmaapp.com), designed for young learners.

For this study, the focus is specifically on Zebra AI, designed for young learners. The Zebra English learning module integrates key AI technologies, including ASR, adaptive algorithms, and learning analytics, to establish a highly interactive and data-driven learning environment. The ASR module is designed to recognise and assess learners' oral production in real-time, providing immediate feedback on pronunciation. The adaptive mechanism dynamically identifies learning gaps to generate personalised review sequences, thereby facilitating personalised learning pathways. Furthermore, learning analytics supports the system's adaptive regulation and pedagogical strategies by collecting and interpreting learning performance data. These AI-driven functionalities collectively support multimodal interactions (integrating animations and human instruction), learner engagement, and sustained language development in an immersive digital environment.

Comparatively, other applications such as Liulishuo (https://www.liulishuo.com) have developed a localised ASR model designed for Chinese learners, enabling the recognition and assessment of English pronunciation with a Chinese accent. This function plays a significant role in pronunciation correction and building children's confidence in spoken English, particularly during the early stages of learning. In addition, Yuanfudao's AI systems (https://m.yuanfudao.com) employ learning analytics to construct personalised learning pathways by analysing children's behavioural data, such as learning pace, response patterns, and interaction habits. These innovations serve as practical foundations for studying AI-assisted L2 learning in early childhood contexts.

In the field of SLA, effective language acquisition depends on high-frequency input and output, meaningful interaction, immediate feedback, and personalised support (Ellis, 2015). These requirements are consistent with the affordances of AI-based learning tools. Current AI-based applications in language education apply natural language processing (NLP) for semantic understanding, ASR for pronunciation feedback, intelligent error correction, virtual tutors for immersive dialogue, and recommendation algorithms for content personalisation (Zawacki-Richter et al., 2019; Zou et al., 2023). These technological affordances are particularly significant in the context of early childhood second language learning. As early childhood is a critical period for language development, consistent language exposure and meaningful interaction are essential for the language acquisition and the development of communication skills. These AI-based language learning applications can provide personalised input, interactive feedback, and multimodal immersive learning environments that meet the developmental learning needs of young children. Consequently, these applications are frequently used by children at home as part of informal learning. Parents increasingly recognise the potential of language learning applications and treat them as supplements to formal education (Vaiopoulou et al., 2021).

1.2 Research Rationale

The inspiration for this research stems from my personal language learning experience and observations of teaching practice. As a Chinese English learner, I experienced directly the challenges caused by a lack of contextual support and personalised feedback in traditional classroom settings. Years later, while pursuing a degree in educational technology, I began to explore AI-driven language learning tools. This prompted me to consider: if children could obtain interactive and adaptive AI applications during early childhood, would this improve the passive experience of traditional learning and enhance motivation and efficiency?

My experience as a part-time English teacher further deepened this reflection. Children aged 5 to 7 are typically curious and eager to learn, yet their developing cognitive abilities limit their capacity for sustained attention, and they often show significant individual differences. Conventional teaching methods often lack in addressing their diverse needs at both cognitive and emotional levels. This led me to ask whether AI-based language learning applications could function not only as digital toys, but as serious educational tools which can offer interactive engagement, adaptive feedback and learning analytics to support early language development. Over time, this question became central to my research.

When I began exploring this topic in 2021, the majority of research was focused on online education, gamified learning and various other forms of technology-assisted learning (Chapman & Rich, 2018; T. Chen et al., 2020; Palvia et al., 2018). In contrast, studies specifically targeting AI-based applications were relatively scarce, and the existing literature more focused on teenagers or undergraduate students (Arifiati et al., 2020; Shortt et al., 2023; Zawacki-Richter et al., 2019), resulting in limited empirical evidence concerning young children.

Building on this gap, previous research has also paid limited attention to three critical aspects. First, empirical research is limited on whether AI tools can effectively support receptive vocabulary acquisition in children aged 5 to 7. Second, few studies explore whether young learners experience enjoyment and sustained engagement while using these tools. Third, parent-caregivers' attitudes, as learning decision-makers are rarely considered, despite their important influence on children's learning environments.

Besides these design and experiential considerations, the geographical and cultural contexts of existing studies also present limitations. Most research has been conducted in Western educational settings. Despite the widespread availability of applications and learning opportunities, little research has examined their impact on early years learning in China. Especially in the context of the "Double Reduction" policy, where offline tutoring is under strict regulation, the trend of home-based learning with the help of AI-based language learning applications has grown rapidly. Children are using these applications more frequently, and this has led parents to pay more attention to their educational value, ease of use, and the quality of interactive experience. As a result, under the combined influence of policy and technology, it is both timely and meaningful to systematically investigate the actual impact of AI-based language learning tools on children’s SLA, as well as to explore parent-caregivers’ attitudes and levels of acceptance. Such research not only reflects real-world educational needs but also offers valuable insights for improving the design and promotion of language learning tools for young learners in the future.

In conclusion, this research aims to address both theoretical and practical gaps in current research on AI-supported language learning for children, with a particular focus on Chinese learners aged 5 to 7. It systematically examines a specific AI-based language learning application (Zebra AI) from three key aspects: learning outcomes, hedonic experience, and parent-caregivers ’perspectives. This not only offers empirical support for the optimal design of this type of AI-based educational tool but also provides theoretical insights and practical guidance for understanding how technology can more effectively support children’s language development.

1.3 Research Aims and Questions

While the application of AI technology in L2 learning has received increasing academic attention, at the time this research was initiated in 2021, systematic research evaluating its efficacy in supporting very young children (aged 5 to 7) remained generally scarce. The existing literature has largely concentrated on adolescents and undergraduate students, resulting in a relatively limited focus on supporting the language development of younger children.

Specifically, academic consensus remains inconclusive regarding the extent to which individual differences in cognitive and social-cognitive abilities (such as WM and ToM) predict or moderate the effectiveness of assisted learning. To ensure consistency with established literature, this thesis adopts the overarching term WM. However, it must be explicitly noted that within this study, WM is operationalised as a narrower aspect of this construct, specifically associative memory. Investigating this associative function is paramount, as it plays a central role in cross-modal binding. It may therefore influence a child's ability to map auditory labels onto the visual animations presented by the AI-based application. ToM may simultaneously predict their learning engagement during interactions with AI characters or simulated environments.

Consequently, these two variables are explicitly integrated into our analytical framework as core moderating factors, used to account for the individual variance observed in AI application learning outcomes.

At the same time, systematic research is still lacking concerning children's subjective experience (e.g., hedonics and sustained motivation) and the roles and attitudes of parents/carers. Given the near paucity of empirical research on these specific variables within the Chinese cultural context, this research aims to address both this theoretical and empirical void.

To deal with this research gap, the aim of this research is to systematically investigate the potential impact of AI-based language learning applications on receptive vocabulary acquisition among Chinese children aged 5 to 7. The research focuses on three key aspects: (1) learning outcomes, (2) children's hedonic experience, and (3) the perspectives of parents and caregivers.

To achieve this aim, an approach with mixed methods was employed, with both quantitative and qualitative data. They include pre- and post-intervention assessments of children's receptive vocabulary, WM, and ToM; hedonic experience questionnaire responses; and semi-structured interview data collected from parents and caregivers.

Accordingly, the research deals with the following three research questions:

RQ 1: To what extent does the AI-based app improve children's receptive vocabulary, and how are these outcomes predicted by their WM and ToM?

RQ 2: Which functions and design interface influence children's hedonic experience with the AI-based language learning app?

RQ 3: What are parent-caregivers' views regarding the effectiveness of using AI-based learning apps for learning a second language?

1.4 Research Setting

To deal with the three research questions, this research adopted a mixed-method experimental design (Creswell & Creswell, 2017). Given that the research focuses on both the objective effects and subjective experiences of language learning, as well as the role of parent-caregivers in the learning process, a single research method may be insufficient to fully capture how AI-based language learning applications influence young children's L2 development. The use of mixed-method approach allowed for the integration of quantitative and qualitative data, thereby enhancing the explanatory power of the findings.

Atotal of 85 Chinese children aged 5 to 7 were recruited and assigned to either an experimental group ($N = 43$) or a control group ($N = 42$). The intervention lasted for 12 weeks. Children in the experimental group engaged in weekly learning tasks by using a designated AI-based language learning application, while those in the control group used non-educational entertainment apps, such as game platforms.

To assess the effects of the intervention, participants were tested before and after the intervention on English receptive vocabulary, WM, and ToM tasks. A variety of standardised instruments were employed to ensure the validity and reliability of the measurements, including the PPVT-5 (Dunn, 2019), WM task, and ToM tasks. The resulting quantitative data were analysed through pre- and post-test comparisons both within and between groups to evaluate the impact of the AI intervention on language and related cognitive abilities.

At the end of the intervention, children's hedonic experience with the AI-based application was explored through a child-friendly open-ended questionnaire, and parent-caregivers' perspectives were collected through semi-structured interviews. By integrating qualitative and quantitative findings, the study offered a comprehensive analysis of the potential impact of AI-based language learning tools on children's SLA from three perspectives: vocabulary development, hedonic experience, and parent-caregivers'involvement.

1.5 Research Significance

This research has significant value in theoretical, methodological, and practical terms.

At the theoretical level, it deepens our understanding of how specific AI-based language learning application can support young children's L2 learning. Although AI has received considerable attention in the field of language education, most existing studies have focused on young people or undergraduate students, and few have systematically investigated its applicability and effectiveness during the critical developmental stage of children aged 5 to 7. To address this gap, the present research provides empirical evidence for the effectiveness of the tested AI applicaton in early language learning and contributes to a paradigm shift in AI and SLA research by extending its scope from older learners to younger children.

In addition, the research introduces children's hedonic experience as a key yet underexplored variable in AI-assisted language learning. For learners with still-developing cognitive capacities, motivational, emotional, and experiential factors play a crucial role in language acquisition. By focusing on the child's view, this research highlights the emotional aspect oftechnology use and broadens the SLA field from a knowledge-centred approach to a more holistic, experience-driven model.

In practical terms, the research also emphasises the essential role of families in technology-supported learning. Through interviews with parents, it explores their acceptance of AI tools, their observations of learning outcomes, and the ways they support their children's use of technology in home settings. This not only underscores the family's dynamic role in the learning ecosystem but also provides both theoretical insights and practical references for integrating AI-based education tools into home learning environments.

From a methodological perspective, the research adopts a mixed-methods design, combining receptive vocabulary testing, hedonic experience questionnaires, and interviews with parents and caregivers. It offers a comprehensive analysis of AI learning applications by examining cognitive processing, emotional engagement, and social interaction. Moreover, it explores tools for measuring hedonic experience in young learners, establishing a foundation for developing child-centred approaches to user experience evaluation.

In summary, this research fills a gap in empirical research on the use of multimodal AI-enhanced tools in young children's language learning. It also contributes to the scientific application of AI in early childhood education by offering new theoretical perspectives, research angles, and methodological innovations.

1.6 Outline of the Thesis

This thesis is structured into seven chapters.

Chapter 1 (this chapter) introduces the research background, personal motivation, and rationale behind the study. It clearly presents the research aims and questions, and briefly outlines the research design, methodology, as well as the significance of the research.

Chapter 2 presents a comprehensive literature review. It begins by examining key characteristics of early SLA in children, with a focus on the importance of receptive vocabulary and the relationship between language and cognitive development, and the specific mechanisms of early childhood vocabulary acquisition.The chapter then reviews the development and current application of AI language learning tools in children's language education. It outlines key design principles for AI-driven vocabulary apps, before concluding with a discussion of their pedagogical potential and possible limitations.

Importantly, this chapter points out two underexplored areas in existing research, children's hedonic experience and parent-caregivers' perspectives to explore their impact on learning motivation and technology adoption. Based on the prior research, the chapter identifies theoretical and practical gaps that this study seeks to address, thereby providing a basis for the following chapters.

Chapter 3 constructs the theoretical framework of the study by integrating multiple relevant theories, including SLA theories (e.g., the Input Hypothesis and Interaction Hypothesis), MLT, and SDT. It also incorporates user experience model (UX model) and the TAM model. This integration provides a well-supported theoretical foundation for understanding the mechanisms through which AI-based language learning tools may influence children's language development and informs the selection of research variables and design.

Chapter 4 outlines the research methodology in detail, including the study design, participant recruitment, intervention procedures, and data collection methods. It explains the rationale for adopting a mixed-methods approach and introduces both quantitative instruments (e.g., PPVT-5, WM task, ToM tasks) and qualitative tools (e.g., hedonic experience questionnaires, parent- caregivers' interviews). The chapter also discusses data analysis strategies and ethical considerations.

Chapter 5 presents the results of the study, including quantitative analyses of the effects of AI- based language learning applications on children's receptive vocabulary, WM, and ToM. It also reports findings from the hedonic experience questionnaires and summarises parent-caregivers' observations and feedback regarding their children's use of the application. All data are presented systematically through statistical analysis and thematic interpretation.

Chapter 6 provides an in-depth discussion of the findings, comparing quantitative and qualitative results with existing literature. It explains how AI language learning tools affect children's vocabulary development, hedonic experience, and the role of parent-caregivers' influence. This chapter also explores the theoretical and practical implications of the study, reflects on its limitations, and offers directions for future research.

Chapter 7 concludes the thesis by summarising the main findings and addressing the research questions. It highlights the theoretical contributions and practical recommendations based on the study, acknowledges its limitations, and proposes future directions, particularly in improving AI language learning applications and enhancing learning support strategies for young children.

1.7 Glossary

Artificial intelligence (AI): The capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making (Russell, 2021).

Second language（L2): L2 refers to any language learned in addition to a person's first language (Gass et al., 2020).

Second Language Acquisition (SLA)：The learning of a nonnative language after the first language has been learned, either in a naturalistic setting or in a formal classroom setting (Spada & Lightbown, 2019).

Natural language processing (NLP): A subfield of computer science and AI that uses machine learning to enable computers to understand and communicate with human language.

Chapter 2. LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to review the theoretical and empirical research findings related to this thesis, and to lay a theoretical and methodological foundation for exploring the impact of AI-based learning applications on children's L2 receptive vocabulary acquisition. The contents cover the key mechanisms of children's L2 learning, the evolution of digital technology in language learning, the role and influence of AI applications, the influence of design features on user experience, and parent-caregivers' roles and attitudes in children's learning process.

First, Section 2.2 reviews the unique aspects of children's SLA, emphasising their developmental characteristics in language input processing, meaning construction, and social interaction, while highlighting the significant differences between children and adults in their learning paths. Then, this section discusses the fundamental role of receptive vocabulary in second language development, especially in input-driven language learning. Building on this, Section 2.2.3 provides a detailed examination of the mechanisms underlying early childhood vocabulary acquisition, outlining the developmental trajectory from phonological segmentation and short-term maintenance to meaning construction and long-term consolidation. In addition, it also examines two key variables from the perspective of cognitive psychology, namely WM and ToM, and discusses how they affect children's language reception and comprehension. These cognitive factors also provide foundational support for understanding the underlying mechanisms of children's L2 learning and informing the subsequent experimental design.

Section 2.3 provides a comprehensive review of the evolution, theoretical foundations, and practical implementation of AI in L2 learning, tracing the evolution from CALL to AI-driven ICALL. Regarding pedagogical design, the review integrates key principles governing contemporary AI-driven vocabulary applications. Grounded in theories such as Dual Coding and the 'Testing Effect', these applications prioritise multimodal input integration, active retrieval practice, spaced repetition systems, and adaptive feedback mechanisms to facilitate memory consolidation. Crucially, the review of the Chinese context reveals a significant research gap: despite the widespread commercial adoption of these tools by children, empirical studies remain disproportionately focused on adult learners, underscoring the urgent need to investigate their efficacy in early childhood.

Section 2.4 explores how design features in AI learning applications influence children's hedonic experience. It first introduces the conceptual framework and measurement methods of hedonic experience before analysing the potential impact of design elements in four dimensions: interface design and usability; gamification elements and motivational support; content presentation and multimedia integration; and feedback and reward strategies. These aspects help explain children's sustained engagement in learning activities from a user experience perspective.

Finally, Section 2.5 focuses on parent-caregivers' roles and attitudes in children's language learning. It discusses how parent-caregivers serve as sources of guidance, monitoring, and support in children's learning activities. It then reviews current research on parent-caregivers' acceptance of educational technology and willingness to use it, and further analyse key influencing factors, such as educational beliefs, digital literacy, and cultural background, thereby providing sociocultural support for subsequent intervention recommendations.

In summary, this chapter establishes an interdisciplinary theoretical foundation for the research by integrating language learning theories, cognitive psychological processes, educational technology development trends, user experience research, and family support factors, while also providing a reference for the subsequent research design and empirical analysis.

2.2 Early L2 Learning in Children

Early childhood is a critical stage of language development and cognitive growth, with unique advantages for SLA (Flege et al., 1999; Lenneberg, 1967). Compared with adults, children show greater plasticity in speech imitation, language input processing, and motivation for social interaction, which places them in a stronger position to internalise language through natural input and interaction (Birdsong, 2006; Oliver & Azkarai, 2017). However, their language development is also significantly influenced by vocabulary growth and cognitive abilities.

2.2.1 Developmental Features of Children's SLA

A second language typically refers to any language learned after the acquisition of a first language. It may be learned as a foreign language in a native-language environment, or acquired for communication, survival, or education in a second-language context (Ellis, 2008). First language (L1) acquisition is generally regarded as a natural, unconscious process that relies on immersion and social interaction within a language-rich environment, rather than formal instruction (Chomsky, 1986). Chomsky's Language Acquisition Device (LAD) hypothesis proposes that humans are biologically predisposed to acquire language, which enables children to grasp complex grammatical structures without explicit instruction. Although this hypothesis emphasises the implicit and rapid nature of language acquisition, providing an important theoretical foundation for related research, it has been criticised for neglecting these environmental and social influences (Tomasello, 2003). To fully understand lexical development, one must consider that vocabulary is not merely "triggered" by innate mechanisms but is built through social activities. Specifically, children learn words through joint attention and intention-reading during meaningful dialogues with others (Tomasello, 2003) , where the quality of linguistic input and the social context provide the necessary scaffolding for a child's expanding lexicon.

Building on theories of L1 acquisition, scholars have also begun to explore the unique characteristics of second language learning. In comparison, SLA is typically more intentional, supported by structured input, teacher guidance, and explicit rule learning (Brown, 2000). Krashen (1982) in their Acquisition Learning Hypothesis proposed that L1 acquisition happens through implicit absorption, whereas L2 learning often involves the conscious construction of language rules. They also introduced the Affective Filter Hypothesis (Krashen & Terrell, 1983), which highlights the significant influence of motivation, self-confidence, and levels of anxiety on language input processing. This theory is particularly relevant to children's L2 learning.

Numerous studies have demonstrated that children possess several unique advantages in SLA, particularly in phonological development, language imitation, and expressive fluency (Birdsong, 2006; Doughty & Long, 2008). The Critical Period Hypothesis, proposed by Lenneberg (1967) suggests that there is a biologically optimal window for language acquisition, with childhood considered an especially advantageous period for acquiring complex language structures. Although its applicability to vocabulary and grammar acquisition remains debated (Gass & Selinker, 2008), it has received substantial empirical support in the phonological aspect.

To further examine this advantage, researchers often compare child and adult language learners to observe differences in acquisition when starting from the same level. In other words, although both are beginners at the start, children usually learn language more naturally and efficiently and are more likely to pick up language structures without needing clear instruction (Birdsong, 2016). For example, Flege et al. (1999) found that children performed significantly better than adult learners in imitating speech and approximating native pronunciation. Birdsong (2016) also noted that child L2 learners generally outperform adults in terms of language processing speed, pronunciation naturalness, and communicative fluency.

In addition to their cognitive advantages, children also show notable characteristics in terms of emotional states and learning motivation. Compared to adults, children typically demonstrate a greater willingness to express themselves and experience less affective pressure when learning a language (Oliver & Azkarai, 2017; Wells, 1985). They accept mistakes more easily and tend to express themselves with less anxiety, which allows them to experiment with language output more freely (Oliver & Azkarai, 2017). In contrast, adult learners often experience anxiety and self-consciousness due to fear of making mistakes, which can negatively impact their fluency and self-confidence (Krashen, 1982).

Children and adults also differ significantly in how they process language input and select learning strategies. As children's cognitive control systems are not yet fully developed, they have shorter attention spans and are still developing metacognitive strategies (Birdsong, 2016; Spada & Lightbown, 2019). This could make it more challenging for them to maintain consistent progress in mastering language forms (Gathercole & Baddeley, 1993). However, children are more effective at acquiring language through immersion and meaningful experiences in real- world contexts rather than relying on explicit rule-based instruction (Birdsong, 2016). A longitudinal study of children in bilingual families by Hoff et al. (2014) shows that children tend to acquire language use gradually and naturally through activities such as playing games, talking with peers, listening to stories, or watching animated content. These rich and authentic language environments provide abundant input, allowing children to refine and construct their linguistic knowledge through repeated social interactions (Tomasello, 2003). In contrast, adults often rely on structured methods, such as grammar instruction and vocabulary memorisation. Although their analytical skills are strong, their learning paths tend to be more structured and less flexible (Lenneberg, 1967).

Furthermore, children are generally less affected by exam pressure and fear of error correction, which encourages them to express themselves more actively in relaxed settings (Krashen, 1985). This willingness to express themselves provides a foundation for the effectiveness of feedback. With appropriate guidance from teachers or parents, children can receive timely feedback and support through scaffolding, helping them gradually develop language proficiency, self- confidence, and cross-linguistic transfer abilities (Vygotsky, 1978). It is worth noting that parents tend to be relatively tolerant when correcting children's grammatical errors. As a result, children are often able to develop fluency first before gradually refining their linguistic accuracy. Such feedback is typically more positive and supportive rather than strictly corrective, which plays a key role in sustaining learning motivation and emotional comfort (Indriati, 2016).

The reviewed literature indicates that children show significant strengths in SLA, particularly in speech imitation, language processing, emotional engagement, and learning motivation. Therefore, based on children's diverse strengths in language learning, this study focuses on them as the primary research participants to examine the effects of AI-based learning applications on their SLA. In the process of children's language development, receptive vocabulary growth is considered a key indicator of language comprehension and input processing ability. The following section will focus on the role of receptive vocabulary within the context of SLA and further explain its theoretical and practical significance in research on children's language learning.

2.2.2 The Importance of Receptive Vocabulary in Children’s SLA

As a fundamental element of language, vocabulary plays a crucial role in all stages of language learning (Schmitt & Schmitt, 2020; Susanto, 2017). Numerous studies have shown that vocabulary knowledge is not only a key indicator of language proficiency but also a significant predictor of learners' academic achievement and literacy (Alqahtani, 2015; Qian, 2002). Vocabulary development is essential in both first and SLA. In L2 learning, vocabulary knowledge forms the foundation for language comprehension and production. It also serves as a prerequisite for effective acquisition and use of the target language (Nassaji, 2004). According to Input Hypothesis (Krashen, 1985), vocabulary is considered a core element of comprehensible input and an important way to support language development.

In vocabulary acquisition theory, Nation (2001) classifies vocabulary into receptive and productive types. Receptive vocabulary refers to words that learners can recognise and understand through listening or reading, even if they do not actively use them in speaking or writing (Bialystok et al., 2010). It typically precedes productive vocabulary development and is crucial in the early stages of language acquisition. Webb and Nation (2017) pointed out that a key aspect of children's language development is the steady growth of receptive vocabulary, which reflects improvements in their language comprehension skills. Therefore, given the central role of receptive vocabulary in language input comprehension, within the current thesis it is a key variable to evaluate children's L2 comprehension and processing abilities, and measures vocabulary recognition ability to reveal the developmental stages and potential of their language acquisition (Bialystok et al., 2010).

In empirical studies, receptive vocabulary is often assessed using standardised tests (such as the Peabody Picture Vocabulary Test, PPVT; Dunn , 2019) or computer-assisted recognition tasks (Vatalaro et al., 2018). By requiring learners to match images with spoken words, these tools indirectly reflect their level of understanding of language input and offer a quantitative measure for the objective assessment of L2 proficiency. However, other studies have noted that such tools present limitations in capturing vocabulary depth, ensuring cross-cultural validity, and representing authentic language contexts (Mancilla-Martinez & Lesaux, 2010). Therefore, it is important to take potential limitations into full consideration when interpreting the relevant data.

From the perspective of developmental psycholinguistics, children in the early stages of SLA typically show a developmental pattern in which comprehension comes before expression. Lesaux et al. (2006) and Oliver and Azkarai (2017) noted that children's language comprehension often precedes their productive abilities, a developmental pattern that follows the natural path of language learning and lays the foundation for the emergence of productive language. From a cognitive standpoint, this phenomenon can be explained by the fact that language comprehension demands fewer cognitive resources, whereas language production involves greater information integration and expressive control (Baddeley, 2003). Studies have shown that children's limited WM and underdeveloped attentional control may limit the fluency and complexity of their language production (Gathercole & Baddeley, 1993).

At this stage, receptive vocabulary plays an essential role in evaluating children's early understanding of language, as it forms a core part of comprehensible input. Sun and Yin (2022) suggested that for young learners, the development of receptive vocabulary not only reflects their ability to process input but also serves as an observable indicator of their ongoing language development. Empirical evidence supports this perspective (Paribakht & Wesche, 1997). Furthermore, their research found that children typically follow a developmental sequence in which vocabulary reception precedes production. In this sequence, learners first acquire the ability to recognise and understand new words within context, and only later are they able to use these words in speaking or writing. This progression highlights the essential role that receptive vocabulary plays in the early stages of children's SLA.

In summary, receptive vocabulary plays a central role in children's SLA, and its development is closely linked to language comprehension, internalisation, and subsequent productive language skills. Therefore, this thesis considers receptive vocabulary as a key indicator to examine the developmental features and pathways of children's early SLA. The section 2.3 will further explore how different instructional interventions, particularly AI-based learning tools, influence the acquisition and development of children's receptive vocabulary.

2.2.3 Early childhood vocabulary acquisition

Early childhood vocabulary acquisition is not simply about memorising words. Rather, it is a developmental process supported by neural, cognitive, and social–linguistic mechanisms (Kuhl, 2004; Tomasello, 2003). To understand how technology-mediated interventions such as Zebra AI may support children's receptive vocabulary learning, it is important to examine how children process speech sounds, construct meaning, and integrate new words into the mental lexicon.

Previous research broadly agrees that vocabulary learning involves several related but distinct mechanisms that develop over time (Bloom, 2000; Hirsh-Pasek & Golinkoff, 2006; Hoff, 2006). This process can be described as progressing through a series of stages (Werker & Curtin, 2005), including phonological perception and segmentation, short-term maintenance and encoding (Gathercole & Baddeley, 1989), meaning construction through referential inference (Bloom, 2000), and long-term consolidation in the mental lexicon (Davis & Gaskell, 2009).

Vocabulary acquisition begins with children's ability to process continuous speech. Before understanding meaning, children must first identify individual words within the speech stream, a challenge known as the segmentation problem. Research suggests that the learning brain is not a passive listener but uses rhythmic and prosodic cues, such as stress patterns and syllable boundaries, to support speech segmentation (Goswami, 2019). This phonological sensitivity provides a foundation for the initial encoding of new word forms. Once word forms have been segmented, the ability to temporarily maintain these unfamiliar sound patterns becomes crucial. A substantial body of research has shown that phonological short-term memory (PSTM) is a strong predictor of vocabulary development (Gathercole, 2006; Gathercole & Adams, 1993; Service, 1992). For example, Chrysochoou et al. (2013) found that vocabulary size in Greek-speaking children was significantly associated with PSTM at ages 7.5 and 8.5, highlighting the importance of PSTM in early vocabulary growth.

Taken together, these findings indicate that PSTM acts as a key constraint in vocabulary acquisition. If children cannot temporarily rehearse and retain new phonological forms, their consolidation into long-term memory is likely to be limited.

Once phonological forms have been segmented and temporarily maintained, children face a further challenge in vocabulary learning: determining what a new word refers to. This issue is often illustrated by Quine's "Gavagai" problem, which highlights the difficulty of deciding whether a novel word refers to a whole object, its parts, or its properties (Quine, 1960).

Research suggests that children resolve this referential uncertainty through a combination of cognitive constraints and social–pragmatic inference. One well-established mechanism is fast mapping, which refers to children's ability to form an initial guess about a word's meaning after minimal exposure (Carey & Bartlett, 1978). This process is guided by cognitive biases rather than chance. In particular, the whole-object assumption leads children to interpret a new label as referring to an entire object, rather than to one of its features (Markman, 1990; Saxton, 2017).

Together, these constraints allow children to establish early links between word forms and likely referents, even in highly ambiguous learning situations.

However, research consistently shows that word knowledge formed through fast mapping is often unstable (Horst & Samuelson, 2008). Biemiller and Boote (2006) found that for children aged 5–8, a single exposure to new words through story reading is not sufficient for durable vocabulary learning. Instead, stable lexical knowledge develops through slow mapping, a process in which repeated exposure to the same word across different contexts allows children to confirm and refine their initial interpretations (Henderson et al., 2013). This distinction highlights the difference between early encoding and later consolidation in vocabulary acquisition. While cognitive constraints help initiate learning, long-term retention depends largely on repeated and varied input.

Beyond consolidation over time, the quality of early word learning also depends on the learning context. Meaning construction is not an isolated cognitive process but is embedded in social interaction. Research shows that passive language exposure, such as television viewing, is insufficient for typical language development, whereas interactive contexts support learning through social contingency (Roseberry et al., 2014; Saxton, 2017).

Vocabulary learning therefore relies heavily on interactive settings involving joint attention, in which children infer word meanings by interpreting speakers' communicative intentions (Tomasello, 2003). Clark and Casillas (2015) further emphasise that physical and conversational co-presence plays a central role in meaning construction, with everyday feedback and shared activities supporting language development. Gestures, including pointing and showing, are particularly important in this process. Studies indicate that gesture–word combinations facilitate early word learning and predict later language outcomes (Capirci et al., 1996).

Moreover, gestural cues remain important even in digital or non-present contexts, as they reduce working memory demands and provide cognitive scaffolding for novice learners (Ping & Goldin-Meadow, 2010). Together, these findings suggest that children integrate auditory input with visual and gestural information to form multimodal social–pragmatic inferences during vocabulary learning (Kelly et al., 2010).

As children gain more lexical experience, newly learned words are gradually integrated into the mental lexicon. The mental lexicon is not a simple list of words, but a dynamic system that organises and retrieves lexical knowledge (Levelt, 1989). Saxton (2017) suggests that children's lexical knowledge is structured mainly through semantic networks, where words are linked by meaning and sound similarity. For example, the word dog may activate related concepts such as cat or bone. During early development, the well-known vocabulary spurt is thought to occur when new labels are mapped onto existing concepts, leading to the expansion and refinement of these semantic networks (Aitchison, 2012).

For L2 learners, the integration of new words differs from L1 acquisition. Research suggests that early L2 lexical representations rely more on phonological form than on semantic organisation (Jiang, 2000; Singleton, 1999). Beginner learners often link new L2 words directly to existing L1 concepts, rather than forming new conceptual categories. This reliance on L1 mediation is consistent with the Revised Hierarchical Model (Kroll & Stewart, 1994).

For young L2 learners, such as children aged five to seven, vocabulary learning therefore involves two closely related processes: establishing stable phonological forms and linking these forms to existing conceptual knowledge. Clark (1993, 2016) distinguishes between two stages of vocabulary development: encoding and consolidation. Encoding corresponds to fast mapping, while consolidation depends on repeated exposure, use across contexts, and input from multiple speakers. Through consolidation, connections within semantic networks are strengthened, allowing word forms to become stable and easily retrievable (Clark, 2017). In SLA, this process is especially important, as learners require stronger and more sustained input to convert fragile phonological representations into long-term lexical knowledge (Jiang, 2000).

In summary, research suggests that children's vocabulary learning depends on two closely linked processes. First, the development of the mental lexicon requires repeated and varied language input that supports consolidation. Second, meaning construction depends on the integration of cognitive constraints and social–pragmatic cues. Together, these processes form the cognitive basis of vocabulary acquisition, while also placing demands on children's limited cognitive resources.

From this perspective, the key question in technology-mediated learning is not simply whether digital tools can present new words, but whether they can support phonological encoding, referential inference, and long-term consolidation. Building on this framework, the next section examines the role of cognitive resources, such as working memory, in vocabulary learning and explains how these factors are operationalised in the present study.

2.2.4 Cognitive Factors in Children's SLA

In recent years, children's cognitive development and its influence on SLA have received increasing attention. According to cognitive psychology, children aged 5 to 7 are in a critical developmental transition from the pre-operational to the concrete operational stage (Piaget & Cook, 1952). During this period, the cognitive focus moves from simply retaining separate items to actively connecting multiple pieces of information (Cowan et al., 2006; Gathercole, 1998). At this stage, foundational cognitive abilities fundamentally shape how children receive and internalise language input (Paradis, 2011; Unsworth, 2016).

As a core cognitive system, WM provides the essential mental workspace for language processing. While Baddeley (2000) model identifies the phonological loop and visuospatial sketchpad as key subsystems, their role in SLA extends beyond mere storage. Specifically, phonological short-term memory is widely posited to play a crucial role in the initial encoding of unfamiliar L2 words. Research by Service (1992) and Cheung (1996) established that the capacity of the phonological loop to temporarily hold unfamiliar sound traces significantly predicts long-term vocabulary learning success. Furthermore, Engel de Abreu (2011) confirmed showed that WM capacity limits children's ability to process unfamiliar phonological inputs. These established findings provide the general cognitive framework for understanding how children process linguistic information.

However, it remains debated whether storing and connecting information are the same process. While some theoretical models suggest associative processes as embedded within the broader WM system (Oberauer, 2002), other research argues for the functional separability of the binding mechanism from simple maintenance capacity (Page & Norris, 2009). Adopting this perspective, throughout this thesis, the term WM will be adopted to maintain theoretical consistency. However, it must be explicitly stated that within the context of this study, WM refers to a narrower aspect of this construct, which is responsible for relational binding and distinct from mere temporary maintenance.

In the specific context of L2 vocabulary acquisition, WM plays a more important role. Unlike L1 acquisition, which is typically supported by rich contextual input, L2 learning requires learners to link unfamiliar phonological forms to meanings which are not naturally connected (de Saussure, 1959). WM is the primary cognitive tool available to children to bridge this arbitrary gap. Furthermore, receptive vocabulary performance largely reflects retrieval efficiency (Nation, 2001). Stronger WM supports clearer sound–meaning links, allowing the appropriate concept to be activated rapidly when a word is heard.

This mechanism is particularly important for children aged 5–7. At this stage, logical analytical abilities are still developing and cognitive resources are limited (Gathercole, 1999; Piaget, 1964). Efficient WM enables young learners to form sound–meaning bindings with fewer cognitive resources, thereby supporting learning (Verhagen & Leseman, 2016). Accordingly, given that Zebra AI adopts a Visual–Auditory Paired-Associate Learning paradigm (Hulme et al., 2007), WM was selected as the primary cognitive predictor rather than general storage capacity. The AI's multimedia input, consistent with the Cognitive Theory of Multimedia Learning (Mayer, 2005), places direct demands on this binding process, making WM a more essential and task-relevant predictor of intervention outcomes.

In addition to core cognitive factors, social cognition also plays a significant role in language acquisition. ToM as an important part of social cognition, refers to the ability to understand that others hold beliefs, intentions, knowledge, and emotions that differ from one's own (Wellman, 2018). According to the developmental stage model by Wellman and Liu (2004), children aged 5 to 7 are in a critical period for the development of ToM abilities. At this stage, children not only begin to recognise that others may have different perspectives, but they are also able to predict behaviour based on false beliefs. Furthermore, they can form second-order mental representations, such as "I know that he knows something" (Wellman & Liu, 2004).

As ToM develops, children demonstrate enhanced social reasoning skills during language acquisition. Instead of depending only on surface-level linguistic signals, they can interpret the speaker's communicative intentions and identify potential pragmatic meanings. This allows them to acquire word meanings more accurately from complex contexts (De Villiers, 2014; Milligan, 2007). This ability is particularly critical in SLA, especially when dealing with polysemy, contextual ambiguity, and indirect referential expressions (Matthews et al., 2006). Research has shown that the development of ToM is closely linked to children's sensitivity to pragmatic cues, which helps them to grasp the speaker's intended meaning and effectively decode implicit information in language (De Villiers, 2007). Therefore, in L2 learning, ToM not only facilitates vocabulary comprehension but also enables children to use language more flexibly in real-life communication, thereby enhancing their communicative competence and contextual adaptability. Building on this connection between ToM and language development, Bloom (2002) further pointed out that vocabulary acquisition is, at its core, a process of "mind reading" : learners must understand the speaker's intention to correctly associate new words with specific referents or concepts. Studies suggest that bilingual children may outperform monolingual children on ToM tasks, possibly due to their increased need to monitor interlocutors' perspectives across languages (Berguno & Bowler, 2004; Javor, 2016). This relationship suggests that L2 learning may also contribute to the refinement of ToM abilities.

Empirical studies further support the close relationship between ToM and language comprehension. Nguyen and Astington (2014) assessed English monolingual, French monolingual, and French-English bilingual children using tasks measuring false belief understanding, WM, and inhibitory control. Their findings showed that bilingual children outperformed both monolingual groups on the false belief and WM tasks, though no significant advantage was observed in the inhibition task. Milligan et al. (2007) conducted a meta-analysis of 104 studies involving 8,891 children under the age of seven and found a significant positive correlation between children's performance on false-belief tasks and their language ability. This association showed a moderate to large overall effect size, which remained significant even after controlling for age. This suggests that ToM serves as a central cognitive component supporting verbal reasoning and word comprehension.

These findings further support the close connection between both first and L2 experience and the development of ToM. Although most studies support the positive influence of ToM on language acquisition, some researchers have suggested that the relationship may be bidirectional, with language ability also contributing to the development of ToM (Astington & Jenkins, 1999).

In summary, WM and ToM represent core components of information processing and social reasoning in language learning. Given the potential influence of WM and ToM on children's L2 receptive vocabulary acquisition, this study included WM, ToM, and receptive vocabulary in the research model as primary measurement variables. This integration not only helps to better understand the AI-supported development process of children's language learning but also provides theoretical and cognitive support for evaluating the effectiveness of technology-based interventions.

2.3 Artificial Intelligence in L2 Learning

This section reviews the evolution of language learning technology from traditional CALL to Intelligent ICALL, highlighting the limitations of traditional CALL systems in providing personalised support and effective human-computer interaction (HCI), which hinders their ability to meet the increasingly diverse and contextualised needs of today's learners. To cope with these limitations, researchers and developers have gradually integrated AI technologies, such as natural language processing, speech recognition, and adaptive recommendation systems into language learning environments, facilitating the development of more intelligent, interactive, and adaptive tools, and thus contributing to the emergence of ICALL. This section also examines the broader application of AI in education, particularly its innovative uses in language learning, with a focus on the adoption, key features, and potential challenges of AI-based applications in the context of basic education in China.

2.3.1 From CALL to ICALL

With the rapid development of information technology, CALL has gradually become an important means of associating language education with digital technology, and continues to generate widespread interest in the field of SLA (Rasekh Eslami & Zohoor, 2023; Zhang, 2021). Since its emergence in the 1960s, CALL has undergone significant theoretical and technological evolution and is generally divided into three stages: behaviourist CALL, communicative CALL, and integrated CALL (Bax, 2003; Warschauer & Healey, 1998).

Rooted in behaviourist theory, early CALL emphasised repetitive drills and programmed instruction, such as grammar gap-fill and vocabulary matching exercises (Gündüz, 2005). These approaches helped beginners practice language forms but lacked contextual support and interactivity (Gündüz, 2005; Skinner, 1957). The communicative CALL of the 1980s shifted the focus to functional language use, emphasising role-play activities, task-based learning, and multimedia integration (Chapelle, 2001). These methods significantly improved learners' fluency and expressive abilities (Chapelle, 2001). However, the interaction remained largely predetermined and lacked dynamic responsiveness to learner input (Zhang, 2021).

With the emergence of integrated CALL, computer-mediated communication promoted cross-cultural collaboration and social interaction, further enhancing the social and contextual dimensions of language learning (Warschauer et al., 1996). However, systems at this stage were often text-dominated and lacked effective support for spoken interaction. This text-oriented nature limited its potential to integrate multimodal language input, especially in task-based instructional settings that require authentic communication and contextual support (Blake, 2016). In addition, integrated CALL systems often required advanced technological skills from educators, creating a high entry threshold for practical implementation (Hampel & Stickler, 2005).

Although CALL continues to incorporate emerging technologies, it still faces challenges in offering personalised support, interactive negotiation, and age-appropriate adaptability (Bax, 2003; Smith & Mccurrach, 2021). These limitations are particularly evident in child-centred learning contexts, manifesting in limited individualised feedback, restricted interaction types, and a lack of age-appropriate content. These issues hinder both the comprehensibility of language input and the motivation to learn (Levy, 2006).

In response to these persistent limitations of conventional CALL, researchers have increasingly turned to ICALL as a more sophisticated paradigm for supporting language learning (Heift & Schulze, 2007) In academic discourse, ICALL represents a significant conceptual and technological advancement within the broader field of CALL. While early generations of traditional CALL systems predominantly relied on drill-and-practice formats based on pre-programmed responses, ICALL is distinguished by its explicit integration of AI techniques. Scholars such as Heift and Schulze (2007) characterise ICALL as the incorporation of NLP technologies into the architecture of Intelligent Tutoring Systems (ITS). Amaral and Meurers (2011) further refine this definition, arguing that ICALL systems combine linguistic intelligence (e.g., semantic analysis of learner output) with pedagogical intelligence (e.g., feedback generation and path adaptation). This integration enables systems to analyse learner input in a principled manner and to provide personalised support, rather than relying on simple string matching. The operational framework of ICALL can be examined across three interrelated dimensions: core technologies, system architecture, and pedagogical capabilities.

The foundational distinction of ICALL lies in its capacity to process natural language in ways that approximate human-like understanding. NLP constitutes a core component, enabling systems to perform sophisticated operations such as automated error diagnosis (Heift, 2022). In the domain of oral language acquisition, ASR plays a critical role by converting spoken learner input into representations that can be evaluated for pronunciation accuracy and fluency (Ehsani & Knodt, 1998).

In recent years, the scope of these technologies has expanded significantly. The integration of Generative AI, such as GPT-based language models, has enhanced the adaptability of ICALL systems, enabling them to simulate human-like communication patterns and create more immersive environments (Law, 2024). For instance, adaptive writing systems powered by Large Language Models can now provide detailed feedback on semantic coherence and discourse structure, fostering metacognitive reflection (Su et al., 2023; Yan, 2023). Alongside these linguistic technologies, Machine Learning algorithms enable the analysis of large-scale interaction data, supporting the prediction of learning trajectories.

At a theoretical level, ICALL systems are commonly grounded in the canonical triadic architecture of Intelligent Tutoring System（ITS). This framework typically comprises three interacting components. First, the Expert Model encapsulates expert knowledge of the target language, including lexical items, grammatical rules, and permissible variations in learner responses. Complementing this, the Learner Model is widely regarded as the locus of system intelligence. This component dynamically represents the learner’s current state by tracking proficiency levels, identifying recurring error patterns, and, in some systems, incorporating cognitive variables such as learning trajectories or forgetting-related processes to personalise instructional decisions (Bull & Kay, 2010). Finally, the Tutoring Model draws upon the Learner Model to make instructional decisions. The development of this component relies not only on technological integration but also on insights from educational psychology. Vygotsky’s (1978) sociocultural theory, which emphasises that language ability is constructed within the Zone of Proximal Development (ZPD) through scaffolding, provides a theoretical foundation for the Tutoring Model (Alkhudiry, 2022). Systems such as AutoTutor (Graesser, 2016) exemplify this by using NLP to simulate educational dialogues that scaffold learners’ ability to negotiate meaning.

The integration of advanced technologies gives rise to distinct pedagogical capabilities, primarily Intelligent Feedback and Adaptivity. From a SLA perspective, when these functions are consistent with concepts like ‘comprehensible input’ (Krashen, 1985), ‘negotiated interaction’ (Long, 1996), and ‘noticing’ (Schmidt, 1990), they may significantly enhance the depth of language learning (Amaral et al., 2011). However, researchers caution that current implementations of these mechanisms remain preliminary and require further evaluation (Felix, 2008).

Furthermore, there is a growing emphasis on the affective dimension of pedagogy. The integration of emotion-aware computing has opened new avenues for support, particularly for young learners. Applications embedded with emotion recognition algorithms, trained on datasets like EmoReact (Nojavanasghari et al., 2016), can dynamically adjust feedback strategies based on voice tone or facial expressions. This affective adaptation involves modifying the tone or pacing of feedback to reduce anxiety and maintain motivation (Valagkouti et al., 2022). Such features are especially critical in early childhood education, where emotional regulation is key to sustained participation.

Having established the theoretical foundations and technical structure of ICALL, it is essential to examine how these mechanisms are implemented in specific learning domains. Among the various sub-skills of language acquisition, vocabulary learning has emerged as a major field for AI integration, particularly within the context of Mobile-Assisted Language Learning (MALL). Consequently, to understand the pedagogical potential of these tools, we must examine the frameworks guiding their development. The following section reviews the core design principles behind state-of-the-art AI vocabulary applications, analysing how theoretical principles are translated into specific functional features and practical applications.

2.3.2 Design Principles for AI-Driven Vocabulary Apps

Vocabulary learning begins with the formation of stable form–meaning links, and Multimodal Input Integration is widely recognised as a core principle in the design of contemporary vocabulary learning applications (Lin & Lin, 2019). Rather than relying on text alone, this approach combines orthographic input, auditory input, and visual representations to create a multi-channel learning environment. This design principle draws on Dual Coding Theory (Paivio, 1986) and the Cognitive Theory of Multimedia Learning (Mayer, 2020), both of which propose that learning is enhanced when verbal and non-verbal information are processed in parallel.

The central assumption is that coordinated verbal and visual input supports stronger associative mappings than the mere addition of multiple information sources. Empirical research suggests that multimodal input can facilitate semantic processing by providing multiple retrieval pathways and by supporting the associative binding of new lexical items (Hulstijn, 2001; Montero Perez et al., 2014; Plass et al., 1998). This mechanism is particularly important for young learners aged 5–7 with limited L2 proficiency. For this age group, studies show that pedagogically designed dynamic animations offer effective scaffolding for form–meaning mapping and are more beneficial than static input alone (Uchikoshi, 2006; Verhallen et al., 2006).

Taking Duolingo as an example, the application applies the theoretical principles outlined above through a highly synchronised interactive design(Moreno & Mayer, 2007). When a new word is introduced, the system usually presents the written form, a related image, and standard target-language pronunciation at the same time (Chun & Plass, 1996; Mayer, 2020; Sweller, 2011). Rather than simply combining multiple media, this design aims to create a perceptual basis for associative encoding by aligning information across different channels(Clark & Paivio, 1991). It should be emphasised that such examples are used here to illustrate how design principles may be implemented, rather than to serve as direct evidence of learning effectiveness.

However, input alone is not sufficient for durable memory retention. A large body of research shows that Retrieval Practice, often referred to as the Testing Effect, plays a central role in long-term memory consolidation. This principle suggests that learning should go beyond repeated exposure and require learners to actively retrieve information from long-term memory. Compared with passive reading or repeated presentation, such active retrieval has been shown to produce stronger retention and better transfer effects (Barcroft, 2007; Roediger & Karpicke, 2006). From an associative learning perspective, retrieval strengthens the links between word forms and meanings by repeatedly reactivating these associations. As highlighted by Nation (2022) and formalised in the

Involvement Load Hypothesis (Laufer & Hulstijn, 2001), this effortful retrieval process is critical for stabilising associative bonds and reducing forgetting.

In practical terms, many AI-based vocabulary learning applications organise retrieval activities along a hierarchy from low to high difficulty, reflecting different depths of vocabulary knowledge (Henriksen, 1999; Laufer & Goldstein, 2004). Recognition-based retrieval typically functions as an entry-level activity, such as multiple-choice or image–text matching tasks, where learners identify target items among distractors. These tasks mainly involve associative recognition and relatively weak retrieval demands (Joe, 1998; Nation, 2022), supporting beginners in forming initial form–meaning links. The image–text matching feature in Baicizhan provides a clear example of this approach.

By contrast, production-based retrieval places greater cognitive demands on learners, as it requires the reconstruction of form–meaning associations without direct cues. The spelling mode in Shanbay illustrates this type of retrieval: learners must produce the full word based on L1 definitions, which encourages deeper associative consolidation. This process reduces the illusion of competence (Bjork, 1999) and has been shown to improve retention, particularly among young learners (Goossens et al., 2014; Karpicke et al., 2014).

Since retrieval is central to memory consolidation, its timing and frequency are key factors in strengthening memory traces. Drawing on the spacing effect, contemporary vocabulary applications commonly adopt spaced and distributed practice by spreading learning activities over time to reduce forgetting and support long-term retention (Cepeda et al., 2006; Rohrer & Pashler, 2007).

In L2 vocabulary learning, this principle is typically implemented through algorithm-driven spaced repetition systems (SRS). Research consistently shows that SRS is more effective than massed learning in supporting vocabulary acquisition, for both adults and young learners (Nakata, 2011; Vlach & Sandhofer, 2012). At a basic level, SRS adjusts review intervals based on learner performance, so that retrieval occurs when form–meaning associations are close to decay, thereby maximising consolidation (Settles & Meeder, 2016).

Some AI applications, such as Momo, further use learning data to model individual forgetting patterns and schedule reviews more efficiently. It should be noted that these systems do not directly predict memory decay, but instead apply heuristic optimisation of review timing based on learners' performance data.

Even so, research broadly agrees that adaptive scheduling can improve learning efficiency by targeting weaker associative links, illustrating an effective integration of cognitive psychology principles with educational technology (Godwin-Jones, 2011; VanLehn, 2011).

To improve the effectiveness of the practice and scheduling mechanisms described above, many systems also incorporate dynamic feedback and adaptive designs. Feedback is carefully structured in both timing and form to support error correction while sustaining learning motivation.

Research shows that immediate corrective feedback helps learners refine form–meaning mappings at early learning stages (Long, 1996). In addition, affective feedback helps reduce anxiety and maintain engagement, aligning with Krashen (1982) Affective Filter hypothesis.

For example, Duolingo creates a positive reinforcement environment through gamified features such as instant audio-visual rewards and progress feedback. This provides affective scaffolding for young learners and supports sustained attention and continued use over time (Reinders, 2012).

Taken together, these elements suggest that adaptive learning systems function as an overarching design framework that integrates the mechanisms described above. Unlike SRS, which mainly adjusts review timing, adaptive systems also modify task difficulty and activity type in response to learners' ongoing performance (Vandewaetere et al., 2011).

This approach is grounded in Vygotsky's concept of the ZPD (Vygotsky, 1978), which has been applied in educational technology to provide context-sensitive scaffolding

(Luckin, 2010). Learning tasks are therefore designed to be slightly beyond the learner’s current level but still achievable.

Research indicates that matching task difficulty to learner ability supports associative encoding and helps maintain engagement (Dede, 2014). For example, Lingvist adapts the order of vocabulary presentation by monitoring form–meaning association strength and retrieval errors, reducing the risk of disengagement caused by tasks that are too easy or too difficult (Kickmeier-Rust & Albert, 2010).

2.3.3 AI Apps in China: Growth and Implementation

In recent years, the development of mobile learning has promoted a transformation in language learning methods. As an important subfield, MALL integrates mobile devices with language teaching resources, focusing on learners’ autonomy over time and location. It supports flexible language input and output in authentic and varied contexts (Kukulska-Hulme & Shield, 2008). With portability, personalisation, and interactivity, MALL offers an innovative learning experience for second language learners, especially in an individualised and fragmented learning scenario.

With the rapid evolution of artificial intelligence technology, language learning tools are gradually transitioning from traditional ICALL systems to more lightweight, portable, and interactive AI-based learning apps. These applications integrate core functions such as NLP, ASR, and adaptive recommendation algorithms. Powered by mobile platforms, they provide immersive, real-time feedback and a highly responsive learning experience, gradually becoming an important medium for L2 learning (Fitria, 2021; Woo & Choi, 2021). From the perspective of sociocultural theory (Vygotsky, 1978) and immersive learning theory, this technological integration enhances contextual engagement and supports learners’ self-regulation in the learning process.

In the Chinese context, the implementation of the "double reduction" policy, along with the growing maturity of online education technologies, has jointly contributed to the rapid rise of AI-based language learning applications. These tools have demonstrated particularly promising potential in the K-12 stage of children’s English learning. Yang et al. (2021) argue that the popularity of such apps is driven not only by advances in core AI technologies but also by the widespread use of mobile internet, shifting parent-caregiver's attitudes toward education, and the growing demand for personalised and efficient learning experiences.

Several local educational technology platforms, such as Zebra English, VIPKID, Youdao Kids English, and Companion Fish English, have launched AI-powered learning tools designed for children. These applications typically integrate features such as voice evaluation, adaptive exercises, and real-time feedback. Some platforms further enhance learner engagement by incorporating animated content, multimodal input, and gamified designs (Du & Daniel, 2024). For example, Bamma English offers phased speech training and vocabulary instruction through AI-driven personalised learning pathways and picture-book-based materials. VIPKID has also promoted a blended model of "AI + foreign teachers," providing a supplementary solution for regions facing teacher shortages (Miao & Holmes, 2021).

Empirical studies suggest that AI-assisted language learning tools have a positive impact on increasing learning frequency, extending study duration, and enhancing learner autonomy (Shortt et al., 2023). Compared with traditional textbooks and live courses, these tools offer greater flexibility in fragmented time utilisation, more timely correction of spoken errors, and better adaptation to individual learning paths (Qassrawi et al., 2024). In addition, some platforms integrate emotion recognition and voice interaction mechanisms that can dynamically adjust task difficulty based on learners' intonation and emotional responses. These systems provide positive reinforcement, thereby improving learners' motivation and emotional self-regulation (Yang & Zhao, 2024).

Beyond the Chinese context, international research has also confirmed the instructional adaptability and user-friendly nature of AI-based language learning apps. For instance, Woo and Choi (2021) examined studies published between 2017 and 2020, highlighted that app-based AI tools, such as Google Assistant, pronunciation training software, and vocabulary games that effectively enhance learners' listening and speaking proficiency, vocabulary acquisition, and pronunciation accuracy. Tai and Chen (2023) recruited 112 eighth-grade EFL learners, found that using Google Assistant helped increase learners' motivation and reduce anxiety in English language learning after a two-week intervention. Similarly, several pronunciation training apps evaluated by Fouz-González (2020) was found to significantly improve learners' fluency and articulation. Moreover, Rachels and Rockinson-Szapkiw (2018) reported that gamified mobile apps had a positive influence on Spanish vocabulary acquisition among elementary school students. Meanwhile, platforms such as Duolingo have further improved vocabulary learning efficiency through adaptive learning paths and game-based mechanics (Shortt et al., 2023).

In the Chinese context, AI-assisted language learning applications have also attracted increasing attention from researchers. Empirical studies have examined various aspects such as English speaking, writing, vocabulary acquisition, and collaborative interaction. A 20-week experiment conducted by Wang and Han (2021) with college students showed that AI-driven mobile applications significantly improved learners' linguistic complexity, accuracy, and fluency in English monologue production. In the area of collaborative learning, Wu and Miller (2020) found that mobile apps like PeerEval facilitated cooperative communication and high-quality peer feedback among Hong Kong university students, thereby fostering a positive learning community. Additionally, platforms such as IELTS Liulishuo demonstrate strong adaptability in exam-oriented English instruction by providing real-time feedback and personalised reports aligned with the four IELTS scoring criteria: grammar, vocabulary, fluency, and pronunciation (Li, 2020).

However, the pedagogical effectiveness of current AI language learning applications remains a subject of debate. On the one hand, most platforms still rely on pre-programmed options and speech-matching mechanisms, lacking deep semantic understanding and dynamic negotiation capabilities, which restricts their ability to simulate authentic communicative scenarios (Fitria, 2021). On the other hand, assessment mechanisms based primarily on algorithmic recognition and quantitative metrics may overlook learners' cognitive processes and pragmatic comprehension (Bulut et al., 2024).

It is worth noting that although many studies have confirmed the positive impact of AI-based language learning tools among Chinese learners, the target population remains largely concentrated in higher education. Zhou (2021) reviewed 15 empirical studies on oral English MALL in China from 2017 to 2021 and found that the majority of the research focused on university students, with limited attention given to primary and secondary school learners. Similarly, Woo and Choi (2021) systematic review of 53 international empirical studies on AI-based language learning revealed that more than 60% of the participants were college-level learners, further highlighting the narrow demographic scope of this research field.

It is important to recognise that child learners differ significantly from adults in terms of language acquisition mechanisms, cognitive development, and technology acceptance. However, this group remains largely underrepresented in existing studies. This structural gap not only limits our understanding of the pedagogical suitability and effectiveness of AI-based language learning applications but also constrains their refinement and application in K–12 educational settings. Therefore, this thesis focuses on child learners and explores the applicability and potential value of AI language learning apps in children's SLA, aiming to enrich both theoretical perspectives and practical insights in this emerging field.

2.4 Children's Hedonic Experience in AI-Based Language Learning

With the growing integration of AI-based language learning applications in early childhood education, learner experience, particularly hedonic experience has become an increasingly important focus of research. Unlike pragmatic dimensions such as usability or task efficiency, hedonic experience emphasises the overall enjoyment and emotional engagement learners experience when interacting with technology. For young children, who are still in the process of cognitive and emotional development, hedonic experience plays a particularly crucial role in sustaining attention, enhancing learning enjoyment, and fostering long-term motivation.

This chapter aims to provide a comprehensive overview of children's hedonic experience when using AI-assisted language learning applications. It begins by reviewing the theoretical foundations of hedonic experience, drawing on key concepts from user experience design, HCI, and educational psychology. It then introduces the main tools and frameworks currently used to measure hedonic experience in children, analysing their methodological strengths and limitations. Finally, the chapter explores key factors that may influence children's hedonic responses. Through this review, the chapter seeks to offer both theoretical grounding and empirical insight into the role of hedonic experience in the effectiveness of AI-based language learning tools.

2.4.1 Definition and Theoretical Foundations of Hedonic Experience

The term *hedonic* originates from the ancient Greek philosophy of hedonism, which posits that positive emotions and pleasure are central motivators of human behaviour (Moore, 2004). With the rise of the information society and the rapid development of digital technologies and human- computer interaction, UX research has shifted from a primarily functional focus to an emotional orientation. Increasing attention is now given to whether technical products can evoke users' emotional responses and psychological satisfaction, in addition to fulfilling basic task-oriented needs (Alhussayen et al., 2015; Mispa et al., 2019).

In this context, Hassenzahl et al. (2000; 2003) proposed an influential model of UX, which serves as one of the core theoretical frameworks for the present study. This model divides the perceived quality of user experience into two fundamental dimensions: Pragmatic Quality（PQ) and Hedonic Quality (HQ). Within the broader UX model, PQ refers to the task-oriented attributes of a product and highlights its effectiveness as a tool for helping users achieve specific goals. This dimension focuses on whether users are able to complete their intended tasks in an efficient, clear, and controllable manner during interaction. In essence, PQ addresses the question: “Does this application enable me to carry out my learning tasks smoothly and effectively?”

HQ, by contrast, introduces emotional and experiential dimensions into the traditional evaluation framework that had previously emphasised efficiency and usability. According to their view, high-quality products should not only be easy to use and provide clear information, but also elicit positive emotions, reinforce users’ sense of identity, and evoke affective resonance tied to personal life experiences, ultimately fostering deeper user connection and encouraging sustained use. In a follow-up study, Hassenzahl (2018) further categorised hedonistic into three dimensions: stimulation, referring to a product’s ability to offer novelty and cognitive challenge that fosters personal growth; identification, concerning whether users can express their personal style and sense of belonging through the product; and evocation, involving the triggering of emotional associations or memories. These non-instrumental product characteristics are considered key to enhancing pleasure and increasing user engagement. In contrast to PQ, which emphasises utility and task efficiency, HQ highlights the influence of emotional value and subjective experience in shaping user behaviour (Hassenzahl, 2018).

PQ and HQ together form the core dimensions of user experience. The key distinction between them lies in the fundamentally different user goals they address. PQ (also referred to as “instrumental quality”) highlights task efficiency and utility, focusing on “Do-Goals”, that is, supporting users in achieving clear, externally oriented objectives (Hassenzahl, 2018). In contrast, HQ emphasises non-instrumental emotional value and subjective experience, centring on “Be-Goals”, which aim to satisfy users’ internal psychological needs, such as feeling engaged, entertained, or proud.

Although these goals differ, the two dimensions are closely interconnected and jointly shape a complete user experience. Pragmatic aspects form the foundation of the experience. An application that lacks practicality, for example one that frequently crashes or contains functional errors, cannot deliver meaningful hedonic value regardless of how visually appealing it may be, and will instead lead to user frustration. However, hedonic aspects support long-term use. An application that is only practical, powerful in function yet monotonous in experience, may be effective in the short term but is unlikely to maintain continued use and engagement in contexts such as language learning, which require long-term effort and strong motivation. To understand how product quality translates into these long-term behavioural outcomes, it is essential to examine the user’s psychological dynamics before and after interaction. To understand how product quality translates into these long-term behavioural outcomes, it is essential to examine the user’s psychological dynamics before and after interaction.

To understand how product quality leads to these long-term behavioural outcomes, it is necessary to examine the user’s psychological changes before and after the interaction. Building on this concept, scholars have proposed two closely related constructs: hedonic motivation and hedonic experience. The former refers to users’ pre-use expectations whether a product appears enjoyable or entertaining and is often a predictor of behavioural intention (Venkatesh et al., 2012). The latter focuses on the positive affective responses generated during use, such as enjoyment, satisfaction, and engagement, and is frequently treated as a core outcome variable in user experience research (Diefenbach et al., 2014; Hassenzahl et al., 2003). The relationship between the two can be likened to that between expectation and experience: hedonic motivation reflects the starting point, while hedonic experience reflects the emotional outcome.

Directly assessing hedonic motivation often presents challenging when the participants are child users. Particularly among preschool children or early-stage language learners, children have not yet developed clear personal goals or strong language expressive abilities (Borgers, 2000). Their use of technology is often driven by external arrangements from parents or teachers rather than by intrinsic motivation (Neumann, 2015; Plowman et al., 2010). Therefore, compared to measuring motivation directly, a more realistic and feasible approach is to focus on children's emotional responses during hedonic experiences with technology. This variable is not only consistent with children's developmental characteristics but also more accurately reflects the immediate impact of technological products in real usage contexts.

From the perspective of developmental psychology and sociocultural theory, children's cognitive and language development is highly influenced by emotional states (Beitchman & Brownlie, 2005). Piaget's stage theory of cognitive development suggests that children's cognitive activity during the concrete operational stage is significantly influenced by emotional engagement (Piaget & Cook, 1952). Building on this view of development, Vygotsky's sociocultural theory (1978) emphasises that social interaction and emotional resonance are key factors that facilitate language construction and the negotiation of meaning. Extending this emotional perspective, Fredrickson (2001) broaden-and-build theory proposes that positive emotions help broaden attention and cognitive flexibility, thereby enhancing children's engagement in learning and memory retention. Pekrun (2006) also emphasised that positive emotions during learning support the maintenance of motivation and the construction of knowledge.

Therefore, hedonic experience can be viewed as an important mediating variable linking technology use to language learning outcomes (Moorthy et al., 2019). It not only affects children's continued use and engagement but also promotes language comprehension and internalisation through emotional regulation mechanisms (Beitchman & Brownlie, 2005). Especially in AI-supported language learning applications, focusing on children's enjoyable interactive experiences helps to uncover the potential mechanisms through which technological interventions influence learning (Khlaif et al., 2019).

Although some studies have affirmed the positive role of hedonic experience in children's use of educational technology (Read & MacFarlane, 2006), relatively few have focused on evaluating children's hedonic experience, particularly in the context of AI-based applications for L2 learning. Empirical research in this area remains limited. To address this research gap, it is necessary to review existing measurement tools for assessing children's hedonic experience to provide a theoretical and practical foundation for the development and selection of future instruments. The following section will systematically review several categories of evaluation tools commonly used in children's user experience research, with a focus on their structure, advantages, and examples of their application in educational contexts.

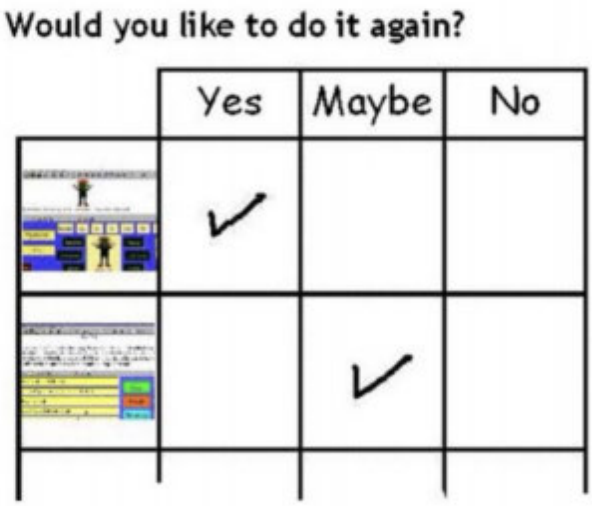
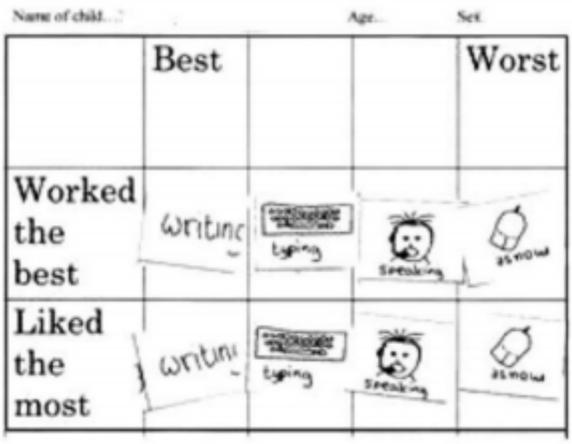
2.4.2 Evaluating Children's Hedonic Experience: Tools and Model Applications

With the growing diversity of children's technology products and their expanding application in educational contexts, researchers have begun to pay more attention to children's subjective experiences and willingness to adopt these technologies, in addition to product usability. In the study of hedonic experience, the methods used to assess children's experiences have become increasingly diverse and integrated. Existing methods primarily include self-reporting, observation, physiological measurements, and data mining (Read & MacFarlane, 2006; Yarosh et al., 2011). Self-reporting has become the most widely used approach in children's user experience research due to its simplicity, low cost, and suitability for large-scale surveys. Common formats include structured questionnaires, image-assisted scoring, and semi-structured interviews (Read & MacFarlane, 2006; Zaman et al., 2013).

Compared to adults, children differ significantly in terms of cognitive development, language expression, and self-reporting ability, which poses substantial challenges for the applicability of traditional questionnaires and interview methods in child-focused research

(Read & MacFarlane, 2006). To address these limitations, researchers have developed a range of evaluation tools specifically designed for children. Among them, the Fun Toolkit (Read & MacFarlane, 2006) has been widely adopted in fields such as education, gaming, and child-computer interaction, and is considered one of the most representative instruments for assessing children's hedonic experience (Looije et al., 2008; Read et al., 2002; Zaman et al., 2013). This tool was specifically designed for children to address the limitations of traditional questionnaires among children with limited cognitive and linguistic abilities. It consists of three core modules: the Smiley-o-meter (a facial expression rating scale), the Again Table (a measure of willingness to engage again), and the Fun Sorter (an activity ranking tool). These components use visual methods such as images, rankings, and tables to significantly reduce the cognitive burden of self-expression for children, thereby improving the comparability, reliability, and validity of the data.

Table 2-1. *Components of the Fun Toolkit (Read & MacFarlane, 2006).*

Smiley-o-meter	
Again table	
Fun Sorter	

Although the Fun Toolkit performs well in capturing children's hedonic experiences, it primarily focuses on children's immediate emotional responses and perceived enjoyment during the usage process. The measurement results largely rely on subjective impressions, such as "Is it fun?" and "Would you like to use it again?” and lack a systematic explanation of the underlying motivational mechanisms, acceptance pathways, and deeper experiential structures involved in children's technology use. While its emotion-oriented and graphical format is suitable for assessing young children, it presents limitations in constructing causal models and predicting behavioural tendencies. Therefore, building on this foundation, this study introduces two well- established theoretical models in educational technology, the UX model (Hassenzahl, 2003) and the TAM (Davis et al., 1989) to enhance the theoretical depth and interpretive validity of the measurement framework.

The UX model focuses on users' emotional responses, perceived value, and overall experience quality during technology use, particularly emphasising the role of HQ in shaping subjective experiences. Its integration facilitates a more systematic understanding of children's emotional engagement, sense of relatedness, and satisfaction in language learning applications, particularly addressing the lack of attention to hedonic experience in previous function-oriented models.

In complement to the UX model, although the TAM model originated in adult information systems research, its core constructs, including PU and PEoU have also showed predictive validity in children (Karisma & Gui, 2023). To make the classic TAM model more applicable to younger children, the materials used within the current thesis simplify the language and adapts the core constructs to suit child users.

Within the current thesis, PEoU is expressed through simplified items such as “Is it easy to use?” and “Is it easy to understand?” Meanwhile, PU is reflected through child-friendly expressions based on task interest and functional feedback, such as “Do you think the game is interesting?” and “Did the pictures help you understand?” These expressions not only preserve the core meaning of the original model but also align more closely with the cognitive development level and linguistic habits of children aged 5 to 7, thereby improving the child-appropriateness and validity of the measurement tool. For more details, please see Section 4.4.3.2 on questionnaire design. Therefore, in this study, the UX model is integrated with the TAM framework and supplemented by the practical instruments of the Fun Toolkit to construct an evaluation model encompassing both functional acceptance and emotional experience. This theoretical integration not only enhances the conceptual soundness and child-appropriateness of the questionnaire design but also provides theoretical and methodological support for a deeper understanding of children's behavioural tendencies and emotional mechanisms in their use of AI-based language learning applications.

The UX model, proposed by Hassenzahl (2003), emphasises users' emotional responses and overall subjective experience during their interaction with technology. This model divides the user experience into two core dimensions: PQ and HQ. PQ relates to the practicality and efficiency of a product in task completion, whereas HQ emphasises the user's enjoyment, interest, aesthetic experience, and sense of personal significance while using it (Hassenzahl, 2003).

In research on children's educational technology, HQ has been widely used to explain children's subjective evaluations of interactive features such as interface design, gamification elements, and animation feedback (Alhussayen et al., 2015; Mispa et al., 2019). Specifically, Mispa et al. (2019) employed the User Experience Questionnaire, based on the UX model, to evaluate the user experience of children aged 5 to 8 with the mobile application FantasyLand. The results showed that the tool effectively captured the emotional responses and engagement motivations of preschool children in gamified contexts. Yarosh et al. (2011) also compared children's perceptions of "functional efficiency" and "emotional pleasure" during interactions with social robots, using the UX framework, further supporting the model's applicability in child-specific contexts.

Unlike the UX model, which focuses on subjective emotional experience, the TAM originated in information systems research and emphasises how users form behavioural intentions. It suggests that PU and PEoU are the two core factors influencing a user's willingness to adopt a technology (Venkatesh & Davis, 2000).

In recent years, the TAM has been increasingly applied in educational technology research, particularly in evaluating mobile learning and children's educational applications. Although the model was originally developed to predict adult users' intention to adopt technology, its core constructs, namely PEoU and PU have shown broad applicability in explaining technology use behaviour (Venkatesh & Davis, 2000)..

As digital tools become increasingly integrated into children's learning, researchers have begun to simplify the TAM structure and adapt its language to enhance its usability and predictive validity with children. For instance, Karisma and Gui (2023) conducted an empirical study with 330 primary school students to examine the applicability of a simplified TAM model in assessing e-learning platform acceptance. The results showed that PU and Attitude toward Using (ATT) significantly predicted intention to use, while PEoU indirectly influenced children's overall attitude toward technology. These findings suggest that, when children possess basic metacognitive abilities, a suitably adapted version of TAM can serve as an effective tool for understanding how they accept and engage with technology.

Building on this foundation, this study integrates the TAM and UX models as the theoretical framework for questionnaire construction. Specifically, the TAM model helps illustrate how users perceive related to functional usefulness and ease of use in children's AI language learning applications, while the UX model emphasises subjective experiences such as pleasure, aesthetic perception, and emotional engagement during technology use.

By introducing the two key aspects of functional acceptance and emotional experience, this thesis seeks to establish an assessment framework aligned with children's cognitive development and to systematically analyse which design elements and interaction features in AI language learning applications influence children's hedonic experience and motivation to continue using the application. This theoretical integration not only addresses the lack of attention to the emotional engagement in traditional single-model approaches to children's learning but also offers theoretical and methodological support for improving future AI-assisted language learning tools.

2.4.3 Potential Influences on Children’s Hedonic Experience

In the preceding sections, it has been established that user experience assessment tools for AI- based language learning apps for children typically encompass two core aspects: PQ and HQ. However, empirical research further suggests that children's user experience with these applications is influenced not only by these assessment factors, but also by a range of design-related factors embedded within the applications themselves (Alhussayen et al., 2015). Previous research has broadly categorised the design factors influencing children's user experience into several key aspects, often related to interface design, content presentation, interaction modalities, and reward features (Cheng & Tsai, 2019; Sun et al., 2022). The remainder of this section will explore how these interrelated elements contribute to children's overall experience.

Interface design, as a fundamental component of children's AI language learning applications, directly influences PEoU and the quality of initial user interaction. Studies have shown that children's interface preferences differ significantly from those of adults. They tend to prefer interface styles featuring a simple layout, bright colours, and intuitive controls (Alhussayen et al., 2015; McKnight & Fitton, 2010). Such designs can effectively enhance children's confidence in navigation and improve their understanding of on-screen information (McKnight & Fitton, 2010; Yan, 2017). Icon recognisability, button size, and layout clarity have all been shown to directly affect interaction efficiency and learning performance (Papadakis & Kalogiannakis, 2017). Druin et al. (2001) pointed out that graphical and animated interfaces are better aligned with children's cognitive development and can enhance both attention and emotional engagement. However, Papadakis and Kalogiannakis (2017) cautioned that interfaces that are either too simplified or too complex may undermine children's motivation to explore and reduce task performance, ultimately leading to negative user experience (Chiasson & Gutwin, 2005).

Beyond the influence of interface structure on initial user interaction, the way information is shown often shaped how users feel during the experience. Multimedia elements (such as images, animations, and audio) serve as key channels for stimulating emotional responses and play a central role in children's user experience (Verhallen & Bus, 2010). Rosas et al. (2003) pointed out that appealing visual design can enhance sustained attention and motivation to participate, while Verhallen and Bus (2010) found that dynamic images support vocabulary acquisition and story immersion. Teng (2023) through a mixed-methods study, further demonstrated that multimodal input integrating images and audio can significantly improve memory retention and enhance emotional engagement during vocabulary learning.

Building on multimedia content, multimodal design further enriches the immersive dimension of user experience. Multimodal interaction not only integrates multiple sensory channels such as vision, hearing, and touch, but also enhances interactivity and realism through synchronous input and feedback (Moreno & Mayer, 2007). Guo and Lan (2023) investigated the impact of multimodal and virtual environments on young EFL learners' reading performance. Using an action research approach, they divided 38 Taiwanese primary school students aged 10-12 into two groups. The control group engaged in picture storybook reading and word games, while the experimental group, in addition to the activities of the control group, incorporated a 3D virtual construction task using the Omni-immersion Vision online VR construction tool. The results showed that students using the VR tool made significant improvements in their English reading ability and showed greater improvement in learning motivation and anxiety levels. This suggests that multimodal input, integrating visual, auditory, and interactive elements, can effectively enhance children's learning motivation, reduce anxiety, and foster stronger emotional engagement and immersion. Li et al. (2023) reached similar conclusions in their study. Through a pilot survey of Learning through Play implemented in several first-tier cities in China, which included questionnaires and interviews, they found that multimodal gamified environments effectively reduced cognitive load disparities faced by children during learning and significantly enhanced their engagement and emotional responsiveness. The research indicated that such environments, by integrating multiple sensory experiences and interactive elements, encouraged greater involvement in learning, improved emotional responses, and thus boosted both learning motivation and outcomes. Furthermore, studies by Cerezo et al. (2019) and Lan (2020)Lan indicate that virtual interaction, XR technology, and multimodal holographic systems can boost children’s confidence in language expression and strengthen learning motivation.

However, Mayer and Moreno (2003) points out that humans have separate systems for processing visual and verbal materials (Dual-channel assumption), with each channel having a limited capacity to process information at one time (limited-capacity assumption). Therefore, using too many sensory modes incorrectly may lead to "modality redundancy" and "sensory overload." Multimodal design should focus on moderation and alignment with learning goals to ensure that each modality supports a balance between learning objectives and emotional regulation (Noroozi et al., 2020).

Building on immersive experiences, gamification mechanisms further help engagement into intrinsic motivation for learning. Typical gamification elements include point systems, task challenges, badge rewards, and leader boards, which can enhance learning engagement and enjoyment through achievement incentives and feedback loops (Smiderle et al., 2020). When game elements align with learners' personal traits, their motivational impact tends to be more enduring (Li & Chu, 2021).

Chapman and Rich (2018) also highlighted the positive impact of immediate feedback and task guidance on cognitive processing and emotional regulation. Plass et al. (2015) further pointed out that the positive emotions prompted during gameplay not only enhance hedonic engagement but also promote deeper processing and better retention of language input.

However, Nicholson (2015) and Smiderle et al. (2020) warned that when game mechanisms rely too heavily on external rewards or create excessive competitive pressure, they are tend to causing "goal shift" (where learners focus more on rewards than on the actual learning content). This shift may disrupt the process of autonomous motivation.

In addition to task-based motivational strategies, HCI plays a crucial role in shaping children's learning experiences. Speech recognition technology improves the fluency of children's expression and response by facilitating more natural input and output, thereby helping establish an emotional connection with the system (Xu & Warschauer, 2020). Bai et al. (2021) found that automatic speech recognition based personalised reading feedback can enhance reading fluency, sense of control, and self-confidence. However, issues such as recognition errors, feedback delays, and semantic mismatches in ASR systems may lead to misunderstandings and user frustration (Mazouz & Maatallaoui, 2025).

For young children, such negative experiences are more likely to lead to reduced motivation or engagement to continue participating. Therefore, ASR systems should emphasise emotional sensitivity and error tolerance, and, when combined with multimodal feedback and human-like language prompts, aim to create a stable, trustworthy, and inclusive emotional interaction environment.

From the perspective of emotional support in user experience, reward mechanisms and virtual character design represent two key pathways which are motivational mechanism and social identification that play an important role in fostering children's learning motivation and sense of social belonging (Ryan & Deci, 2017).

Reward mechanisms stimulate a sense of achievement and task persistence through features such as adaptive progression and phased incentives (Sun & Hsieh, 2018). However, if poorly designed, they may lead to motivation externalisation and overreliance on quantifiable rewards (Ryan & Deci, 2017). Virtual characters, as personified interface agents, serve dual functions of language scaffolding and emotional regulation. Studies have shown that characters with approachable appearances, expressive features, and warm vocal tones are more likely to foster emotional connections with virtual characters, thereby enhancing immersion and promoting a sense of psychological safety (Breazeal et al., 2016; Calvert et al., 2014). However, if character design does not align with children's aesthetic preferences, cultural backgrounds, or gender identities, it may lead to emotional detachment or even user disengagement (Nass & Moon, 2000; Turkle et al., 2006). Furthermore, overly personified characters may cause confusion or discomfort, particularly during early stages of cognitive development (Breazeal et al., 2016). Therefore, the design of rewards and character roles should consider children's developmental psychology and cultural background, using motivation and emotion theories to balance technical incentives with emotional connection.

In conclusion, interface design, multimedia elements, multimodal interaction, gamification mechanisms, speech recognition, reward systems, and virtual characters are key factors that influence user experience in contemporary children's AI language learning applications. These elements not only shape children's motivation, attention, and cognitive engagement during learning but also significantly impact their hedonic experience. In this thesis, these design factors serve as key variables influencing children’s hedonic experience and provide an empirical foundation for the development of the questionnaire, the design of interview outlines, and subsequent empirical analysis.

However, although existing studies have extensively explored the positive effects of these design elements on children's learning engagement, emotional involvement, and motivation maintenance, most discussions remain situated within the context of general educational technology. There remains a lack of systematic and empirical research on how these factors specifically affect children's hedonic experience in AI-based language learning applications.

Addressing this theoretical and empirical gap, the present thesis focuses on the real-world usage of AI language learning applications among Chinese children and explores how these design factors shape children's hedonic experience within specific AI environments. By integrating Hassenzahl's (2003) UX model, this study aims to identify key design features that foster positive emotional responses and sustained learning motivation in children, thereby providing both theoretical guidance and empirical evidence for improving AI language learning tools.

2.5 Parental Roles in Children's Digital Language Learning

Parents have consistently played a key role in children's language development, both in traditional learning contexts and in the digital age. Based on sociocultural theory (Vygotsky's , 1978), their influence is evident not only in early language input and emotional support but also, with the rise of AI applications in education, in their increasingly complex and multifaceted roles. This is particularly evident in the SLA process of children aged 5 to 7 who are learning with the aid of technology, where parental involvement significantly affects both the nature and effectiveness of learning (Harris & Goodall, 2008).

Therefore, understanding the evolution of parental roles and the factors that influence them, as well as exploring parents' decision-making processes and attitudes toward AI-assisted language learning, is essential for comprehensively evaluating the applicability and effectiveness of such tools for children. Section 2.5.1 focuses on two key aspects:

the evolving role of parents in supporting children’s language learning through technology, and their attitudes toward AI language learning applications, including acceptance, trust, and willingness to use them.

2.5.1 Parents’ influence on children’s using digital tool

Language acquisition, as a core aspect of human cognitive development, is influenced by social and cultural factors. Among these factors, the family environment, as the primary context for children's language input and social interaction, plays a crucial role in early language development (Harris & Goodall, 2008). According to sociocultural theory (Vygotsky, 1978), children's language development is gradually constructed through the ZPD during interactions with more experienced individuals (such as parents). In this process, parents not only provide language stimulation but also shape the trajectory of children's language development through emotional regulation, feedback, and other forms of support (Safwat & Sheikhany, 2014; Topping et al., 2013).

Existing studies have consistently shown the positive impact of parental involvement on children's language abilities, including vocabulary accumulation, phonological awareness development, and language comprehension (Wasik & Hindman, 2011). In addition, a positive home language environment has been shown to enhance children's learning motivation, self- confidence, and academic achievement (Goodall, 2018). It is important to note that children aged 5–7 years are in a stage of rapid cognitive language development. Therefore, family language practices play an especially important role during this stage of early cognitive language development.

With the widespread adoption of AI technologies and smart mobile devices, the landscape of children's language learning is experiencing significant changes. The traditional model of human interaction, based on home and school environments, is gradually giving way to a model mediated by technology, consisting of educational applications, speech recognition systems, and virtual characters (Plowman, 2015). The American Academy of Pediatrics (Pediatrics, 2016), while highlighting the educational potential of digital tools, also emphasises that children should be accompanied and guided by parents during their use to ensure proper understanding and critical engagement with the content.

In this context, the role of parents in language education is no longer limited to the traditional language input and emotional support but has further evolved into that of "media gatekeepers" (Livingstone & Blum-Ross, 2020). They need to assess and regulate the selection, use, and frequency of technological tools, thereby indirectly determining the type and quality of language resources that children are exposed to (Chaudron et al., 2018). This role change not only reflects the digitisation of family education practices but also highlights the dual functions of technological guidance and value mediation that parents perform in the process of supporting and regulating children's language learning.

Although existing studies have emphasised the influence of parents' attitudes towards technology on children's learning, most of them focus on the usage behaviours related to general educational technology, and systematic discussion of the underlying mechanisms of parental involvement in AI-based educational applications remains insufficient. Some studies have preliminarily shown that if parents actively select high-quality language learning applications and encourage children to use them in an exploratory way, children may not only perform better in the domains of language and cognition but also develop a stronger sense of learning autonomy (Brito & Dias, 2020; Griffith & Arnold, 2019). In addition, parents’ attitudes towards technology are considered a mediating factor influencing children's learning outcomes (Kong & Li, 2009).

Parents' attitudes towards educational technology are often influenced by their educational beliefs, digital literacy, and cultural values, leading to significant individual differences. Family media ecology theory Nikken and Jansz (2014) suggests that parents' media use is influenced by their education level, value systems, and media experience. Studies have found that parents with higher levels of education and digital literacy are more likely to actively guide their children to explore technological tools with educational value and emphasise their cognitive and developmental benefits (Banić & Orehovački, 2024; Kalaycı & Ergül, 2020). Children growing up in such environments tend to demonstrate higher levels of technological confidence, motivation for exploration, and tendencies toward self-directed learning (Papadakis et al., 2019)(Papadakis et al., 2019). Conversely, parent-caregivers with lower digital literacy may adopt hesitation or restriction strategies that indirectly limit children's developmental opportunities in technology-assisted learning environments (Kalaycı & Ergül, 2020).

In addition, parenting style has also been shown to be significantly related to parents' choices regarding educational technology. According to a survey conducted by (Broekman et al., 2016), involving 316 parents of children aged 3 to 7, authoritative parents tend to choose educational applications with clear structure, well-defined functions, and features that support children's independent exploration. Authoritarian or permissive parents are more likely to choose tools that are entertaining, soothing, or serve as less constructive substitutes for parent-child interaction, which reflects the value-oriented differences that underlie media selection.

In conclusion, parents not only play a fundamental role in traditional language acquisition, but also, as AI technology becomes increasingly embedded in educational practices, their technological attitudes, digital literacy, and parenting concepts more profoundly affect the quality, frequency, and ways of using digital resources that children are exposed to. Especially for 5 to 7 years old children during the critical period of language development, parents' decision- making practices play an important role in shaping their second language learning path. Therefore, this thesis explicitly includes parents in the research perspective, aiming to explore their attitudes towards the use of AI educational applications in children's SLA.

2.5.2 Parent-caregivers' Perspectives on Children's Use of Digital Technologies

Based on a preliminary discussion of parents' roles regarding children's use of technology in the previous section, Section 2.5.2 focusses on parents' attitudes and perceptions of technology- assisted language learning itself. Studies have shown that parents' views on children's early exposure to digital technology, especially their judgment of the potential benefits and risks of technology, significantly affect the frequency of children's exposure, usage patterns, and ultimate learning outcomes of language learning tools (Plowman et al., 2011). A large body of literature reveals that parents' attitudes are complex in nature, often demonstrating a mix of acceptance and skepticism.

Current research on parents' attitudes has predominantly focused on their views regarding the use of digital tools in children's English learning (Chen et al., 2019; Gjelaj et al., 2020; Mikelic Preradovic et al., 2016). A more consistent finding is that parents' attitudes are significantly polarised (Neumann, 2015; Vaala & Takeuchi, 2012). On the one hand, some parents strongly acknowledge the potential of such tools in stimulating learning motivation and improving learning outcomes (Akgun, 2023; Zakaria et al., 2022). On the other hand, some studies have pointed out that parents are generally concerned about the frequency of use, content quality, distraction, and the risk of media dependence (Gjelaj et al., 2020).

Among parents with positive attitudes, many studies have shown that they often regard educational apps as tools for both "learning" and "entertainment", which are believed to improve children's cognitive development and language skills. For example, Zakaria et al. (2022) found that parents particularly value educational apps as tools to extend classroom learning, develop vocabulary, and improve children's spoken expression. Saraswati (2021) cited Duolingo as an example of gamification and interaction that promotes children's initiative in learning, confidence, and vocabulary development. Neumann (2018) further found that children can spontaneously use words learned from apps in daily communication, reflecting the positive role of such tools in language transfer and authentic expression.

Regarding user experience, parents are also generally positive about the design of educational apps, particularly in terms of interface friendliness, visual appeal, sound feedback, and animation rewards, which are believed to help enhance the fun ofthe learning process and extend children's attention span (Papadakis, 2021; Sergi et al., 2017). Among these, apps that include "instant rewards" or target-based incentives are especially favoured by parents and are considered effective in enhancing children's learning persistence and engagement (Read & MacFarlane, 2006).

Parental enthusiasm towards the digital learning apps can be partially explained by the TAM. The model suggests that when parents perceive a technology as having perceived usefulness (PU) and perceived ease of use (PEoU), their willingness to promote it and use it in a collaborative way with their children is significantly enhanced (Tsuei & Hsu, 2019).

While existing research broadly affirms the potential of technological products for children's learning and development, a growing body of studies also indicates that parents have many concerns about the potential risks of technological interventions in real-world use.

While existing research widely acknowledges the potential of technological products for children's learning and development, a growing body of research also indicates that parents have many concerns about the potential risks of technological interventions in real-world use. First, concerns about children's physical development and social skills have been repeatedly raised in multiple studies. Kabali et al. (2015) pointed out that ages 3 to 6 represent a critical period for children's physical activity and the development of social skills, and excessive reliance on touchscreen devices may reduce opportunities for real-world interaction, thereby inhibiting outdoor activity, which is essential for developing peer interaction skills. For example, Ma et al. (2021) have shown that high levels of recreational app use by children are associated with risk factors for obesity. It is worth noting that this concern does not arise from a complete rejection of technology itself, but rather from parents' anxiety about the "substitution effect" of technology use, namely the fear that it may replace parent-child interaction and peer play. As shown by Sergi et al. (2017), some parent-caregivers began to reflect on whether digital device use was encroaching upon family time, and this reflection revealed a subtle shift caused by technological intervention in family relationships.

Second, the discussion of media dependence and addictive tendencies lead to further concerns that parents may have regarding the digital learning tools. Although there is no conclusive evidence that short-term use of high-quality educational apps directly leads to addiction, the lack of certainty has nonetheless triggered ongoing anxiety among parents about the "use boundary", that is, how to establish a boundary-setting strategy between functional and potentially addictive use that can be understood and consistently implemented by the caregivers (Livingstone & Blum-Ross, 2020).

In addition, concerns about content quality and external distractions are key reasons for parents' reservations about digital products. Sergi et al. (2017) pointed out that prolonged immersion in low-quality apps that lack educational value may result in decreased attention span and increased media dependency. Palaiologou et al. (2019) further noted that parents are also concerned about the presence of uncontrolled pop-up ads and inappropriate content for children in apps, which not only undermines parents' trust in technology products, but also challenges their media dominance within the family.

Interestingly, even among parents who are open to technology, the idea that technology must be used with regulation still dominates (Moorhouse & Beaumont, 2020). This shows that parents' attitudes are not strictly polarised but are instead based on a dynamic balance between the perceived benefits and potential risks of technology use. This strategic response reflects the flexibility and complexity of family education practice.

In conclusion, parents' attitudes towards children's use of technological tools for language learning are complex and multifaceted. While affirming the potential of educational apps, they also express ongoing concerns about screen time, content quality, the substitutive effects of media, addiction risks, and the challenges of supervising technology use within the family. For this study, these views not only provide an empirical basis for exploring parents' acceptance and evaluative criteria for AI language learning applications but also highlight the importance of considering parents' active role in technology use and their specific perspectives when designing and evaluating AI educational tools. Understanding parents' perspectives can shed light on whether AI applications can be effectively adopted in children's language learning, as well as on the family's supportive and decision-making roles.

2.6 Summary

This chapter reviewed the key literature informing the current study, focusing on the intersection of early SLA, AI, hedonic user experience, and parental influence in children's digital learning. Section 2.2 examined the developmental and cognitive foundations of children's SLA, highlighting the importance of receptive vocabulary, WM, and ToM. Section 2.3 traced the evolution of AI in language education, from CALL to ICALL, summarised the key design principles of modern AI vocabulary apps, and explored the growth and implementation of AI-based applications in the Chinese context, especially for young learners. Section 2.4 discussed the concept of hedonic experience, its theoretical foundations, and evaluation models, focusing on the need to incorporate emotional engagement into the design of AI learning tools. Finally, Section 2.5 explored the critical role of parent- caregivers in mediating access to and perceptions of digital language learning.

Across these sections, a key research gap was identified: empirical research remains limited regarding the impact of AI-based language learning applications on Chinese children's SLA, the design-related factors that influence their hedonic experience, and parental perspectives on children’s use of such technologies. This review provides a theoretical and contextual foundation for the research framework and methodology presented in the following chapter.

Chapter 3. THEORETICAL FRAMEWORK

The theoretical framework of this thesis draws on multiple interrelated models. These include the Input Hypothesis (Krashen, 1985) and the Interaction Hypothesis (Long, 1996) from SLA theory; MLT (Mayer ,2001); SDT (Deci & Ryan, 2013); the UX Model (Hassenzahl et al., 2003); and the TAM (Davis et al., 1989). The comprehensive integration of these theoretical frameworks provides essential support for the design, analysis, and interpretation of this thesis.

Specifically, based on Input Hypothesis (Krashen, 1985) and Interaction Hypothesis (Long, 1996), this study explores how AI-based language learning applications promote receptive vocabulary development in children aged 5 to 7 years by providing comprehensible input and opportunities for interaction. In order to further understand the impact of design and function in AI applications on children's learning enjoyment (i.e. hedonic experience), this thesis combines MLT (Mayer ,2001) and SDT (Deci & Ryan, 2013) to analyse how the relevant design features in AI applications can align with the cognitive characteristics of children at this developmental stage while also fulfilling their needs for autonomy, competence, and relatedness, thereby enhancing their enjoyment during the learning process.

At the methodological level, the questionnaire developed within the current thesis is based on the UX model and aims to evaluate how the functions and design features of AI language learning applications influence children's hedonic experience. The questionnaire content was structured around the core dimensions of the UX model and TAM to capture children’s subjective perceptions and emotional responses during actual use. In addition, TAM3 is used to inform the design of parent interview questions, aiming to explore factors such as their attitudinal acceptance, behavioural intention, and perceived social norms regarding AI language learning tools. Table 3-1 outlines the connections between the study’s research questions and the theoretical frameworks that guided the study.

Table 3-1. *Research questions, theories, methods of this study.*

Research Question	Theory	Method	Data Type
RQ1	Input Hypothesis (Krashen, 1985) Interaction Hypothesis (Long, 1996) MLT (Mayer ,2001)	Quasi-experiment	Quantitative
RQ2	UX model TAM	Questionnaire	Quantitative
RQ3	TAM3	Semi-structured interview	Qualitative

3.1 SLA Theories

In the field of SLA, the Input Hypothesis (Krashen, 1985) and the Interaction Hypothesis (Long, 1996) are two widely recognised and influential theories. Both emphasise the central role of comprehensible input in language learning and have had a significant impact on technology-assisted language learning. Although the two theories focus on different mechanisms, with the Input Hypothesis stressing the comprehensibility and contextual presentation of language input (Krashen, 1985) and the Interaction Hypothesis highlighting the role of negotiation of meaning during communicative exchanges (Long, 1996), they are highly complementary in explaining the processes of language acquisition.

This thesis adopts these two theories at its foundation, based on two key considerations: first, their strong explanatory power for early childhood language development, particularly in the domain of receptive vocabulary acquisition (Krashen, 1985); and second, their clear theoretical guidance for the design of input delivery and interaction mechanisms in AI-based language learning applications (Y. L. Chen et al., 2022).

The following sections will elaborate on each theory in turn, discussing their relevance and application within the context of this study.

3.1.1 Input Hypothesis

The Input Hypothesis, proposed by Krashen (1985), is one of the core components of Krashen's "Monitor Model". This theory asserts that language acquisition does not rely on explicit instruction or language output, but rather emphasises the importance of understanding comprehensible input (Krashen, 1985). Comprehensible input refers to linguistic material that is slightly above the learner's current language level, namely *i+1*, where *i* represents the learner's current level of language competence, and *1* represents the next level of language input that is just beyond their current understanding but still accessible with the help of context or prior knowledge. Exposure to such input in appropriate contexts can activate learners' subconscious language acquisition mechanisms and promote the natural development of the language system (Krashen, 1982). Especially for children in the early stages of language learning, this view contrasts sharply with behaviourist and other traditional models that emphasise language output (Swain, 1985).

Krashen further proposed that, in order to promote effective acquisition, language input should possess several key characteristics: (1) the input should be comprehensible and moderately challenging (i.e., *i+1*); (2) learning should occur in a state of low affective filter, meaning an environment characterised by low anxiety and high motivation; and (3) language should be presented repeatedly in natural contexts to enhance the potential for meaning-making and long- term retention (Krashen, 1982). These principles provide a theoretical basis for the design of learning environments and hold particular significance for guiding practices in children's language education.

According to the Input Hypothesis (Krashen, 1985), language acquisition is an implicit and subconscious process, emphasising understanding before output as its core mechanism. For children, the early stage of language development is typically characterised by the accumulation of receptive vocabulary, referring to words that children can comprehend but are not yet able to actively produce (Nation, 2001). Based on this theoretical perspective, and to explore the role of AI-based applications in children's SLA, this thesis focuses on the development of English receptive vocabulary in children aged 5 to 7 years.

To operationalise these theoretical concepts within a technological context, this thesis employs Chapelle's (1998) framework for CALL, which translates abstract concepts like 'comprehensible input' into concrete design principles for multimedia learning environments. With the development of AI technology, an increasing number of AI language learning applications have been designed to support children's vocabulary learning. These applications are consistent with what Chapelle (1998) defines as "Input Salience" and "Input Modification" in multimedia CALL. By combining multimodal input formats such as animated stories, voice prompts, and visual highlighting, these applications make linguistic forms more noticeable and "salient" to learners (Chapelle, 1998; L. J. Chen et al., 2020; Piech et al., 2015).

More importantly, many AI systems incorporate adaptive features that dynamically modify input difficulty based on the learner's performance. This technological capability realises the "Learner Fit" criterion proposed by Chapelle (1998), ensuring the input remains at the optimal *i+1* level in actual practice (Fitria, 2021). At the same time, such applications often employed gamification elements to reduce learners' anxiety, enhance learning motivation, and indirectly create a learning environment characterised by a low affective filter (L. J. Chen et al., 2020). These features are highly consistent with the core principles of the Input Hypothesis and offer theoretical grounding for research on the effectiveness of AI applications in children's L2 receptive vocabulary acquisition.

Although the Input Hypothesis emphasises the central role of language input, some scholars have argued that it underestimates the role of language output in linguistic production and syntactic processing, particularly at more advanced stages of language development. They argue that output also plays an important role in the overall process of language acquisition (Swanson & Berninger, 1995). However, in research on early SLA in children, the Input Hypothesis continues to provide an important theoretical foundation for understanding their developmental trajectory and informing instructional design.

In conclusion, the Input Hypothesis not only provides theoretical support for understanding the underlying mechanisms of children's language acquisition but also establishes a conceptual foundation for this study to examine the effect of AI applications in promoting receptive vocabulary development.

3.1.2 The Interaction Hypothesis

The Interaction Hypothesis, proposed by Long (1996) on the basis of Krashen's (1985) Input Hypothesis, emphasises that language acquisition not only depends on comprehensible input, but also relies on the interactional processes that shape and refine such input. More critically, it is the negotiation and dynamic adjustment of input within the interactional process that facilitates acquisition (Long, 1996). Although Krashen (1985) argues that language acquisition is primarily achieved through one-way exposure to *i+1* input, Long contends that in real communication, comprehensibility is not inherent at the outset. Rather, it is constructed through ongoing negotiation between learners and their interlocutors, using interactional strategies such as questions, clarification requests, repetition, and correction. This process is known as the "negotiation of meaning." In this process, learners can not only receive input that is more closely aligned with their current language level, but are also prompted to notice the linguistic forms themselves, thereby enhancing their cognitive processing of language input (Long, 1996).

The Interaction Hypothesis does not negate the fundamental assumptions of the Input Hypothesis but instead builds upon them by proposing that interaction serves as a mechanism to “process and customise” input through social engagement. In contrast to the Input Hypothesis, which emphasises the comprehensibility of language materials, the Interaction Hypothesis highlights the importance of accessibility and adaptability in shaping language input (Krashen, 1985). Studies have shown that interaction can improve the immediacy, relevance, and contextual appropriateness of language input, and contribute to language internalisation by guiding learners to attend to language forms (Loewen & Sato, 2018). Particularly in the context of children's language acquisition, the contextual support and immediate feedback provided by interaction can help alleviate learning anxiety, lower cognitive load, and create a more conducive environment for natural language acquisition (C. H. Chen et al., 2022).

In the AI-supported child language learning environment examined in this thesis, interactional features continue to play a key role, even though learners are not interacting with human communicators. Specifically, many AI systems are equipped with ASR and instant feedback modules that can adjust input forms in real time. Chapelle (1998) describes this process as "Interactional Modification," where the computer acts as a participant in the negotiation of meaning. By simplifying sentence structures, repeating keywords in response to the learner's actions, the AI system moves beyond static delivery to dynamic engagement. According to Chapelle's CALL framework, this computer-mediated negotiation is critical because it provides opportunities for comprehensible output and corrective feedback, which are essential for driving language development (Chapelle, 1998).

In this way, the process of “modified input” and “negotiation of meaning” found in natural communication is effectively simulated (Dennis, 2024). Such automated meaning-negotiation mechanisms are particularly critical for children in the early stages of language development. Studies have shown that instant feedback through HCI is more effective in supporting vocabulary comprehension and retention than passive input (Arifani et al., 2021). In this process, the AI system not only improves the comprehensibility of the input, but also enhances its contextual relevance and accessibility, thereby providing a more supportive environment for children's language acquisition.

However, the application of the Interaction Hypothesis in children's language learning remains controversial. Some studies have pointed out that young children may have difficulty participating in complex interactive structures due to their cognitive and linguistic immaturity, thereby limiting the depth and effectiveness of meaning negotiation (Mackey, 2013). Therefore, when designing an AI language learning system based on the Interaction Hypothesis, factors such as children's cognitive load capacity, the appropriateness of feedback frequency, and the complexity of interactive tasks should be fully considered to achieve a balance between the level of cognitive development and the difficulty of interaction (Mackey, 2013).

In summary, the Interaction Hypothesis not only expands the social-interactional dimension of the Input Hypothesis but also provides a theoretical basis for the design and functional implementation of interaction mechanisms in AI language learning applications. Compared with traditional teaching methods, the more responsive and context-sensitive interactive features in AI systems may be more conducive to children's deeper understanding of receptive vocabulary and the formation of long-term memory within authentic learning contexts.

3.2 Multimedia Learning Theory

MLT was systematically developed by Mayer (2001, 2009, 2021). Mayer (1997, 2021) explains multimedia learning as the process by which learners construct mental representations from the combination of words (such as spoken narration or on-screen text) and images (such as videos, animations, or static illustrations). Since the 1990s, empirical research exploring the specific effects of multimedia instruction has increased significantly (Li et al., 2019), and numerous studies have confirmed the positive impact of multimedia environments on learning outcomes (Akçayır & Akçayır, 2017; Mayer, 2017; Wang et al., 2011). Building on this, Mayer (2009) further developed the Cognitive Theory of Multimedia Learning (CTML), emphasising that the focus of the theory is not the technology itself, but the limitations and possibilities of human cognitive architecture in processing information.

From the perspective of MLT, learning is regarded as an active process of knowledge construction, requiring learners to actively select, organise, and integrate information through both auditory/verbal and visual/pictorial channels. Compared with single-mode representations (e.g., reading text alone), MLT suggests that well-designed multimedia materials can apply the human dual-coding mechanism to promote deeper cognitive processing, thereby significantly enhancing learning outcomes.

Based on these theoretical characteristics, MLT was selected as the core analytical framework for this study. The research object, the Zebra-AI application, essentially constitutes a typical multimodal learning environment. As the application integrates not only text input but also dynamic images (such as contextual animations) and auditory input (such as instructional narration and songs), Mayer's theoretical model provides a framework for analysing how such multi-source information contributes to children's language learning outcomes.

Mayer's CTML (2005, 2009) further clarifies the mechanisms of information processing in multimedia learning environments. The theory is built on three core assumptions of human cognitive architecture: dual channels, limited capacity, and active processing, which form the cognitive psychological basis for analysing the instructional effectiveness of Zebra-AI.

Firstly, the Dual-Channel Assumption integrates Paivio (1986, 1991) dual-coding theory, stating that humans possess two independent systems for processing information: a visual/pictorial channel (including on-screen text) processed through the eyes, and an auditory/verbal channel processed through the ears. This assumption suggests that learners can process information from different modalities in parallel. Zebra-AI apply this mechanism by simultaneously presenting dynamic animations (visual) and instructional narration and songs (auditory), aiming to fully engage children's dual-channel processing capacity for more efficient information encoding compared to a single modality.

Secondly, the Limited-Capacity Assumption draws on Baddeley's (1986) WM model and Miller's (1956) information-processing theory, emphasising that each channel can only process a limited amount of information at a time. When the complexity of instructional content exceeds WM capacity, cognitive overload occurs, hindering learning (Chandler & Sweller, 1991; De Jong, 2010). This assumption is particularly relevant for the target population of this study. Given that children's executive control systems are not yet fully developed and their WM resources are considerably lower than adults' (Gathercole et al., 2004), they are highly vulnerable to cognitive overload in digital learning contexts. Using CTML as a framework allows this study to evaluate whether Zebra-AI reduces unnecessary mental effort (like removing extra information), thereby matching children's limited processing ability.

Thirdly, the Active-Processing Assumption proposes that meaningful learning is not passive reception of information, but an active process in which learners construct coherent mental representations. Mayer (2009) further breaks this process down into three key steps: selecting relevant words and images from sensory input, organising them into logical models in WM, and integrating them with prior knowledge stored in long-term memory. In the Zebra-AI learning context, this assumption highlights the importance of interactive design: the application must provide sufficient scaffolding (e.g., interactive feedback and prompts) to guide children to engage in deep processing of language and images, rather than engaging with the content at a surface level.

Based on the above cognitive architecture, Mayer (2021) proposed a series of empirically supported multimedia design principles. These principles not only serve as practical guidelines for enhancing learning outcomes but also provide a theoretical basis for this study to assess how Zebra-AI can improve cognitive experiences and foster children's hedonic experience.

Firstly, the Multimedia Principle suggests that combining words and images works better than presenting text alone. This study examines how Zebra-AI applies this principle by presenting contextual animations and spoken narration at the same time. For children, such multimodal design goes beyond simple information delivery: vivid animations inherently offer playfulness (Rieber, 1991), and combining visual and auditory information significantly enhances processing fluency (Reber et al., 2004). Theoretically, this cognitive fluency can reduce boredom and provide children with a more enjoyable learning experience.

Secondly, the Coherence Principle emphasises that visual or auditory elements unrelated to the learning goal should be removed to prevent unnecessary cognitive load. Although children's applications often include rich decorative elements, this study examines whether Zebra-AI avoids "seductive details" to protect children's limited attentional resources (Harp & Mayer, 1998). From an emotional perspective, following the Coherence Principle can reduce frustration caused by cognitive interference and remove potential barriers to learning.

Thirdly, the Personalization Principle and the Voice Principle jointly highlight the social aspect of learning. The former suggests that conversational language enhances social presence, while the latter demonstrates that natural, human-like voices promote learning more effectively than synthetic voices. These principles provide a theoretical foundation for analysing anthropomorphic interaction features in Zebra-AI, such as "Zebra Call." This study hypothesises that simulating real interpersonal communication with expressive human voiceovers can reduce foreign language learning anxiety and, through emotional contagion (Hatfield et al., 1994), stimulate children's trust and positive emotional engagement with the learning agent.

In summary, Mayer's MLT and the previously discussed SLA theories form a complementary theoretical framework. The SLA framework defines the "linguistic conditions" necessary for Zebra-AI to facilitate language acquisition, such as comprehensible input and interaction, while MLT establishes the "cognitive conditions" required for children to effectively process this information in digital environments. The sense of competence and autonomy generated by cognitive fluency provides a psychological foundation for fostering children's intrinsic motivation. To further explore the mechanisms of motivation and its relationship with hedonic experience, this study will introduce SDT in the next section.

3.3 Self-Determination Theory

SDT, proposed by Deci and Ryan (2013), is an important theoretical framework for explaining the mechanisms underlying the formation and maintenance of learning motivation. The theory proposes that the quality of learners' motivation depends on the extent to which three basic psychological needs; autonomy, competence, and relatedness are satisfied.

When the learning environment effectively supports these three needs, individuals are more likely to develop intrinsic motivation and demonstrate greater and more sustained engagement in learning (Ryan & Deci, 2000b). This theory has been widely applied in educational research, particularly in the fields of digital learning and child development, gradually demonstrating its theoretical significance and practical applicability.

For children aged 5 to 7, the formation of language learning motivation is influenced not only by external rewards or task goals, but also by whether the learning environment can evoke emotional interest, provide immediate feedback, and generate positive emotional experiences (Ryan & Deci, 2000a). In the early stages of SLA particularly during the period of receptive vocabulary accumulation, children's learning motivation not only determines their willingness to engage in learning tasks, but also directly influences their sustained exposure to language input, attention to linguistic forms, and the development of deeper processing pathways for comprehension and memory (Gass & Selinker, 2008). In recent years, studies in educational psychology have shown that enjoyment, immersion, interest, and emotional satisfaction constitute the key dimensions of children's digital learning experiences, collectively referred to as "hedonic experience" (Hassenzahl et al., 2003; Riconscente, 2013). From a theoretical perspective, hedonic experience can be regarded as an important emotional driving force for intrinsic motivation, and the positive emotional states it fosters can help satisfy the three psychological needs proposed by SDT, thereby enhancing learners' engagement, and learning outcomes.

With the rapid development of educational technology, an increasing number of AI language learning applications incorporate gamified learning combined with a hedonic-focused interface design. For example, Duolingo uses animated characters, task-based rewards, and playful sound effects in its design, with the aim of stimulate children's intrinsic motivation and enhance both learning continuity and emotional engagement. These systems significantly increase enjoyment and engagement in the learning process through various elements, such as virtual rewards, quest completion, narrative context, character interaction, and instant feedback (Hamari et al., 2014). At the level of practical implementation, the three psychological needs proposed by SDT are operationalised in the following ways: autonomy is supported through personalised task selection and pathway control; the sense of competence is enhanced through immediate feedback, task progression, and visualised indicators of achievement; relatedness is fostered through emotional interaction with virtual characters or simulated peer collaboration (Ryan, 2017).

Previous studies have shown that children are more likely to experience positive emotions in hedonic learning environments, thereby enhancing their ability to sustain attention, remain focused, and deepen the processing of language input (Plass et al., 2015). This emotionally driven model of learning engagement is particularly well-suited to the early stages of language development and helps children internalise vocabulary knowledge and retain it over time through the combined support of cognition and emotion.

However, applying SDT to early childhood language learning also presents several notable challenges. On the one hand, children aged 5 to 7 have not yet developed full self-regulatory capacity and often rely on external guidance and adult intervention when making choices about tasks and learning paths, which constrains the development of 'true autonomy' (Reeve, 2006). On the other hand, the extent to which virtual interaction provided by AI systems can evoke a stable sense of relatedness is often constrained by individual differences and technological limitations. Therefore, when applying SDT to the design of AI language learning systems, the psychological characteristics of children at this developmental stage should be fully considered. For children or younger learners, the autonomy component of SDT may fail to function effectively when system features are not developmentally appropriate or cognitively aligned (Jeon, 2022).

In conclusion, SDT provides a solid theoretical basis for understanding the underlying mechanisms of children's motivation in language learning. By integrating the emotional perspective of hedonic experience, this thesis will analyse how functional design in AI-based language applications addresses children's motivational needs to further explore how the satisfaction of these needs influences children's reception, attention, and cognitive processing of language input. This theoretical integration provides a key explanatory framework for exploring the mechanisms of child language acquisition in AI-supported environments.

3.4 User Experience Model

Based on the SDT discussed above, the UX model offers a design-oriented perspective for understanding the mechanisms underlying the generation and maintenance of motivation in children's technology-enabled language learning. SDT emphasises that learning motivation comes from the satisfaction of three basic psychological needs, autonomy, competence, and relatedness (Jeon, 2022) while the UX model further focuses on how technical systems can support these needs through interaction, interface design, and feedback mechanisms, thereby stimulating users' positive emotional responses and encouraging them stay interested in using technology (Hassenzahl, 2018; Hassenzahl et al., 2003).

From a pragmatic perspective, the UX model can be regarded as the technical implementation pathway of SDT at the level of system design and user behaviour. By optimising the user interface and interaction processes, it indirectly facilitates the satisfaction of learners' psychological needs, thereby driving the formation and maintenance of intrinsic motivation (Hassenzahl et al., 2003). These two frameworks are theoretically highly complementary. The UX model not only emphasises the role of emotional experience during system use but also provides an empirical framework for measuring the effectiveness of motivational support and is therefore integrated as an essential component of the theoretical framework in this study.

The UX model, proposed by Hassenzahl (2003) is one of the foundational theories in the field of educational HCI and technology-enhanced learning design.

The model divides the user experience into two primary dimensions: PQ and HQ. Specifically, PQ focuses on whether the system is easy to operate and efficient in task completion, such as whether the interface is clear, the process is streamlined, and the feedback is timely. HQ emphasises the emotional satisfaction elicited by the system, including fun, immersion, emotional connection, and self-expression. Hassenzahl (2003) defined "HQ" as the self-related emotional satisfaction and psychological value obtained by users during system use, emphasising the emotional significance that goes beyond mere functional accomplishment. Building on this foundation, he further proposed that an effective system should not only be usable, but also easy to use, desirable to use, and enjoyable to use (Hassenzahl, 2010).

For children aged 5 to 7, hedonic experiences play an especially critical role. At this stage, children have not yet formed a stable structure of learning motivation, and their learning behaviour is often significantly influenced by the appeal of external environments and immediate emotional feedback (Riconscente, 2013). If AI language learning applications are designed to effectively induce positive emotional experiences, such as pleasure, interest, a sense of accomplishment, and a sense of belonging has been shown to increase children's willingness to engage in language tasks and maintain prolonged participation in language tasks and prolong their engagement, thereby enhancing sustained attention and deeper processing of language input. In addition, previous studies have shown that UX-oriented design can significantly improve children's acceptance of and willingness to engage with language learning apps (Read & MacFarlane, 2006), which provides empirical support for the applicability of UX theory in children's language learning contexts.

In addition, the UX model is widely applied to evaluate system functionality and user acceptance. In terms of PQ, the system should ensure that the task flow is concise, the verbal or visual instructions are clear, and the mode of operation is consistent with the child's cognitive level and behavioural habits. In terms of HQ, immersive and emotionally rich learning experiences are typically created through contextual storytelling, virtual companion characters, dynamic animation, sound feedback, and reward mechanisms. Such a design not only enhances children's sense of engagement in learning but also helps satisfy their psychological needs throughout the learning process, particularly autonomy, competence, and relatedness as emphasised by SDT. Consequently, children are more likely to remain intrinsically motivated.

In summary, the user experience model not only provides an analytical foundation for understanding children's emotional engagement and motivation in AI-based language learning but also serves as a structured analytical tool for system design and evaluation. Based on this, the UX model is introduced in this thesis to analyse how AI application interfaces, interaction mechanisms, and feedback designs influence children's processing of language input. At the same time, the UX model also serves as a dimensional foundation for the hedonic experience questionnaire developed in this thesis. The relevant questionnaire items are constructed around the two core dimensions of PQ and HQ, with the aim of quantitatively evaluating the effectiveness of AI language learning applications in supporting children's autonomous motivation.

3.5 Technology Acceptance Model

TAM, proposed by Davis (1989), is a widely used framework for explaining how users accept and use information systems. Based on the Theory of Reasoned Action (Fishbein & Ajzen, 1975), TAM suggests that a person's intention to use a system is mainly influenced by two beliefs: PU and PEoU. PU refers to the extent to which a user believes that a system can improve performance, while PEoU refers to the extent to which the system is perceived as easy to use and requires little effort (Davis, 1989).

Although TAM was originally developed to explain adults' acceptance of workplace technologies, it has since been extended and refined. For example, TAM3 (Venkatesh & Bala, 2008) provides a more detailed account of the factors shaping PU and PEoU. More recently, research in Child–Computer Interaction has shown that TAM can also be applied to children and young learners (Bourgonjon et al., 2010; Shih et al., 2011).

However, when TAM is used with children, their cognitive development needs to be taken into account (Read, 2008). Unlike adults, who tend to focus on usefulness and outcomes, children's acceptance of technology is often driven by intrinsic motivation, particularly enjoyment (Van der Heijden, 2004). This aligns with the hedonic motivation construct highlighted in TAM3.

Therefore, when TAM is applied in child-centred contexts, researchers have commonly adapted the framework to account for children's cognitive capacity. In previous studies, PU has often been operationalised as children's perceived learning support, while PEoU has focused on the absence of physical or cognitive difficulty during interaction (Bourgonjon et al., 2010; Read, 2008). In addition, research consistently highlights the central role of hedonic motivation, as enjoyment is frequently the primary factor sustaining children's engagement with educational technologies (Heidig et al., 2015). These adaptations allow TAM to remain theoretically grounded while being accessible and meaningful for young learners (Read, 2008).

The TAM3, developed by Venkatesh and Bala (2008), builds upon its predecessors, TAM (Davis, 1986) and TAM2 (Venkatesh & Davis, 2000) , by providing a more comprehensive framework for understanding the psychological mechanisms underlying individuals' acceptance of new technologies. Central to TAM3 are two foundational constructs inherited from earlier models: PU, which refers to the extent to which an individual believes that using a particular technology will enhance their performance, and PEoU, which reflects the degree to which the technology is perceived as easy to operate and free of effort (Davis, 1986). Beyond these core elements, TAM3 expands the model by incorporating a wide range of influencing variables that shape these perceptions. Factors affecting PU include subjective norm, job relevance, output quality, and result demonstrability, while determinants of PEoU encompass computer self-efficacy, perceived external control, computer anxiety, and cognitive load (Venkatesh & Bala, 2008). In addition, TAM3 introduces emotional and experiential components through the inclusion of hedonic motivation, such as enjoyment during technology use and considers the moderating role of users' prior experience in shaping their perceptions and acceptance behaviours (Venkatesh & Bala, 2008). This enriched framework enables a more holistic analysis of how cognitive, social, and emotional variables interact to influence technology adoption.

In the present thesis, TAM3 is primarily used to analyse parents' attitudes toward their children's use of AI-based language learning applications. Through semi-structured interviews, this thesis explores how parents perceive the usefulness of such tools in supporting children's language development, whether they consider these applications easy to use for young learners, and whether they are willing to support continued use. TAM3 provides the theoretical foundation for structuring the interview questions, such as: "*Do you believe this app helps your child learn English effectively?*" "*Is it easy for your child to operate?*" "*Are there any concerns about technological difficulty or negative side effects?*"

These questions reflect the perception pathways outlined in the TAM3 framework. A more detailed explanation of how TAM3 informs the interview design will be provided in the following chapter.

One of TAM3's key strengths lies in its ability to uncover not only behavioural intentions, but also the underlying cognitive, emotional, and social factors that shape technology acceptance (Faqih & Jaradat, 2015). In the context of this study, PU corresponds to whether parents believe AI tools can enhance their child's language proficiency; PEoU reflects whether they think the child can navigate the tool independently or whether usability presents a challenge. Furthermore, the inclusion of subjective norms in TAM3 is particularly relevant, as it enables the study to examine how parental attitudes are shaped by influences from teachers, other parents, or broader societal discourse. More importantly, the hedonic motivation construct in TAM3 reflects the hedonic experience aspect within UX models, providing a theoretical connection between cognitive acceptance and affective evaluation. This connection allows the study to analyse how parents consider their child's enjoyment, immersion, and emotional engagement when assessing the value of AI learning tools.

In summary, TAM3 serves as a foundational framework in this study for guiding the design of interview instruments and interpreting parental responses. By applying TAM3, the study systematically examines the cognitive beliefs, emotional responses, and social influences that work together to impact how parents support or resist their children's use of AI-based language learning technologies. Consequently, informs a deeper understanding of the feasibility and sustainability of such tools in early childhood education.

3.6 Theoretical Framework and Hypotheses

To summarise, this study establishes a comprehensive theoretical framework integrating instructional design, cognitive mechanisms, and affective experience. It aims to systematically explain the potential mechanisms through AI-assisted vocabulary learning and the sources of individual differences.

First, regarding the instructional design dimension, this study draws on Krashen's (1985) Input Hypothesis and Mayer's (2005) MLT to provide learners with comprehensible input via multimodal scaffolding. This design aims to optimise learners' cognitive processing, thereby supporting the effective encoding of new vocabulary items. Specifically, the synergistic presentation of visual and auditory stimuli is hypothesised to enhance the salience of linguistic forms, enabling learners to more easily notice and process key linguistic information. At a mechanistic level, this multimodal input is considered more effective to facilitating the initial acquisition of receptive vocabulary than traditional single-mode methods. These theoretical assumptions provide an important foundation for evaluating the instructional effectiveness of the AI tool in vocabulary teaching.

Second, regarding the cognitive mechanism dimension, the learning effectiveness of the tool is assumed to be moderated by individual learner differences. This study identifies WM and ToM as key predictors, as they likely perform complementary cognitive functions during vocabulary acquisition, with WM supporting the formation of robust form–meaning connections and ToM facilitating perspective-taking in interactive learning contexts (e.g., understanding usage or contextual cues).

The effective operation of this associative aspect of WM draws directly upon the mechanisms described in Baddeley's (2000) model. In particular, the episodic buffer is viewed as a temporary, multimodal information integration system capable of linking phonological and semantic information, thus providing essential cognitive support for the formation of sound–meaning bindings. In the context of vocabulary learning, this integration process is considered critical for constructing representations of new vocabulary. Meanwhile, although the Interaction Hypothesis emphasises the facilitative role of feedback in language acquisition, its actual effectiveness depends on the learner's specific social-cognitive capabilities. In this study, ToM is regarded as a crucial prerequisite for effective HCI: it enables children to interpret the AI agent's behaviours as interactive cues with communicative significance and, on this basis, infer the 'communicative intent' presented. This transforms surface-level, procedural HCI into a more meaningful communicative experience for the learner. Consequently, differences in learners' cognitive resources regarding working memory and ToM may, to some extent, predict individual variations in receptive vocabulary learning outcomes.

Finally, at the level of affective experience, this study emphasises that cognitive mechanisms do not operate in isolation but are embedded within specific affective and motivational contexts.

Drawing on the TAM and UX design principles, the application aims to elicit positive emotional responses from learners by enhancing the system's HQ. Furthermore, incorporating SDT, when the learning context satisfies learners' needs for competence and relatedness, their intrinsic motivation is expected to be strengthened. This, in turn, may encourage learners to invest a higher degree of cognitive effort and engage in greater usage behaviour (i.e., a higher usage dose). Theoretically, this process may foster a virtuous cycle of affective experience, learning engagement, and language acquisition.

Based on the theoretical deductions above, this study proposes the following research hypotheses:

Hypothesis 1a (H1a): Children in the experimental group (using the Zebra AI) will demonstrate significantly higher gains in receptive vocabulary (PPVT-5 scores) compared to the control group.

Hypothesis 1b (H1b): WM and ToM will serve as significant positive predictors of vocabulary learning outcomes.

Hypothesis 2a (H2a): Children will report a generally positive hedonic experience.

Hypothesis 2b (H2b): Children's reported hedonic experience will be significantly and positively correlated with two indicators: actual usage dose and receptive vocabulary learning gains.

The hypotheses outlined above require an experimental research design. To assess the efficacy of AI-driven tools, this study compares the learning outcomes of the Zebra AI with a non-educational alternative. In early childhood education, digital engagement is becoming widespread. Therefore, it is crucial to distinguish the educational impact from simple digital exposure. The research design includes both an experimental intervention and a control condition, resulting in two distinct groups:

Experimental Group: Intervention using Zebra AI

Control Group: Engagement with a non-educational entertainment app

This setup leads to the following two RQs. RQ 1 is intended to address H1a and H1b, while RQ 2 can be linked to H2a and H2b.

RQ 1: To what extent does the AI-based app improve children's receptive vocabulary, and how are these outcomes predicted by their WM and ToM?

RQ 2: Which functions and design interface influence children's hedonic experience with the AI-based language learning app?

Furthermore, the perceptions of parents and caregivers regarding the utility of such apps for language acquisition have not been fully explored. In this study, it is essential to contextualise the quantitative results through the lens of those who facilitate the children's learning. Therefore, a mixed-methods research design (see Chapter 4) was adopted to combine qualitative insights with experimental findings.

Based on this qualitative perspective, the final RQ is proposed:

RQ 3: What are parent-caregivers' views regarding the effectiveness of using AI-based learning apps for learning a second language?

Chapter 4. METHODOLOGY

This thesis aims to investigate the impact of AI-based language learning applications on SLA in Chinese children aged 5 to 7, with a specific focus on receptive vocabulary development, WM, and ToM. In addition, it explores the design and functional factors influencing children's user experience, as well as parent-caregivers' perspectives on the use of AI in early language education. To address these diverse research questions, the study adopts an embedded mixed methods design (Creswell & Creswell, 2017).

Mixed Methods Research is a methodological approach that combines quantitative and qualitative methods, aiming to make use of the strengths of each to provide a more comprehensive understanding of complex research problems (Creswell & Clark, 2017). Depending on the sequencing, emphasis, and integration of data collection and analysis, mixed methods research typically falls into one of six basic designs: convergent parallel design, explanatory sequential design, embedded design, transformative design, and multiphase design. This thesis adopts an embedded mixed methods design, in which qualitative data collection and analysis are embedded within a primarily quantitative framework. The qualitative data aims to enhance and situate the findings of the experimental intervention, offering insights that may not be fully captured through quantitative measures alone. This design is particularly well-suited for intervention-based research, as it helps maintain statistical validity while enhancing the understanding of participants' subjective experiences and the social context of the study (Creswell & Creswell, 2017).

In this thesis, quantitative data were collected through standardised pre- and post-tests to evaluate the causal effects of AI intervention on children's cognitive and language development. In the meantime, qualitative data were collected via questionnaires and semi-structured interviews, guided by TAM and UX models, to capture children's hedonic experience and parent-caregivers' technology acceptance. Through an embedded mixed methods design, this research ensures the validity of quantitative findings while also capturing children's subjective learning experiences and the perspectives of their families, thereby providing a comprehensive evaluation of the educational value of AI-based language learning applications in young children's L2 learning.

4.1 Research Questions

The purpose of this thesis is to evaluate the effectiveness of AI-based language learning applications in supporting L2 learning among Chinese children aged 5 to 7. The thesis focuses on children's receptive vocabulary development, cognitive outcomes (including WM and TOM), as well as their user experience and parent-caregivers' acceptance of the technology. To address these aims, the following research questions were formulated:

RQ1: To what extent does the AI-based app improve children's receptive vocabulary, and how are these outcomes predicted by their WM and ToM?

RQ2: Which functions and design interface influence children's hedonic experience with the AI-based language learning app?

RQ3: What are parents' views regarding the effectiveness of using AI-based learning apps for learning a second language?

To ensure a robust alignment between the RQs and the methodology, this study follows a structured design logic as summarised in Table 4-1.

Table 4-1. *Links between Research Questions, Measures, and Analytical Methods.*

RQ	Measure(s)	Analysis
RQ1	PPVT-5 (Section 4.3.3.1.2) WM Task (Section 4.3.3.1.3) ToM Tasks (Section 4.3.3.1.4)	Mixed-measures ANOVA to test intervention effects (Section 5.1.1.3; Section 5.1.1.4.1; Section 5.1.1.4.2) Multiple linear regression to explore cognitive predictors of gains (Section 5.1.1.4.3)
RQ2	Children's Hedonic experience questionnaire (Section 4.3.3.2.2)	Descriptive statistics about questionnaire data (Section 5.1.1.5) Descriptive statistics about adherence and the usage dosage (Section 5.1.1.6.1) Pearson correlation analyses (Section 5.1.1.6.2)

RQ3	Semi-structured interview (Section 4.3.3.3)	Thematic Analysis (Section 5.2.1)

4.2 The Pilot Study

The main aim of the pilot study was to test the feasibility of the selected design, materials, and procedures in the target sample population. It was also important to evaluate the data collection methods and make any necessary adjustments before the main data collection phase. Based on the feedback from participants and the results of the pilot study, slight modifications were made to the materials and to the way the tasks were presented to participants.

4.2.1 Design

The main study employed a repeated-measures intervention design, where participants were randomly allocated to either a control group (no intervention) or an experimental group (AI- based intervention). Their performance was assessed at baseline and after a 12-week intervention period. The pilot study tested the randomisation procedure and assessed the suitability of selected materials for measuring the target outcomes. It also evaluated the feasibility of the intervention, including participants' engagement with the AI-based language learning application.

4.2.2 Participants

A total of 80 participants (33 males, 47 females, Mage (month) =73.66, SD =9.302) took part in the pilot. 40 participants were recruited from the urban area (15 males, 25 females, Mage (month) =73.6, SD =9.134), while 40 participants were recruited form the rural area (18 males, 22 females, Mage (month) =73.72 , SD =9.584).

During the pilot phase, participants were allocated using a basic randomisation method: the first 40 were assigned to the experimental group and the next 40 to the control group. In the main study, this was improved by using a ping-pong ball draw method (see Section 4.4.1), in which children drew coloured balls to determine their group assignment. This method enhanced the fairness and randomness of group allocation and reduced potential bias.

4.2.3 Materials

The primary outcome measure selected for the study was the early L2 learning assessed using the PPVT. The secondary outcome measures were WM and ToM.

The primary outcome measure was children's early second receptive vocabulary development, assessed using the Chinese revised version of the PPVT. This version contains 120 items, each presented in PowerPoint format with four black-and-white images. However, in the main study, the latest fifth edition of the PPVT was adopted, featuring updated items and coloured illustrations to enhance children's attention and better reflect real-world visual stimuli.

The WM measure remained unchanged between the pilot and the main study, with the WM task used as the assessment tool.

To assess ToM, the pilot study used a basic false belief task developed by (Wellman & Liu, 2004) in order to test one of the young children's ToM skills tasks. In this task, children listened to a simple story involving a character whose object was moved without their knowledge. The children were then asked to predict where the character would look for the object. The task was repeated twice, each comprising one false belief question and two control questions. Given the limited variability in responses during the pilot, the main study adopted a more comprehensive approach, incorporating the full five-step developmental scale by Wellman and Liu (2004), as well as a sarcasm comprehension task developed by Filippova and Astington (2008) to capture a broader range of ToM skills.

4.2.4 Procedure

Before data collection, informed consent was obtained from the headteacher, teachers, and parents or caregivers of the children. The researcher visited the participating schools and verbally reminded participants of their rights before the data collection began. Each child also gave verbal assent before participating in the tasks.

All data were collected in classrooms. Before the experiment, the researcher explained each step of the task in detail to ensure that the children understood how to participate. Each child was tested individually in a one-on-one session. After the test, the children were not informed of their results and were accompanied back to class by their teacher. Each test lasted no more than 15 minutes.

There were no significant differences in procedures between the pilot and main studies; both were conducted in school classrooms. However, in the main study, a warm-up conversation was added before testing to help create a relaxed atmosphere and reduce anxiety.

4.2.5 Results

The data obtained in the study were analysed with descriptive statistics based on the normality test results. Shapiro-Wilk Test was used to assess the normality test since the study group included 80 subjects.

Table 4-2. *Descriptive statistics and Normality tests of Peabody Picture-Vocabulary Test, As sociative Working memory test, and Theory of mind test.*

	N	**Mean**	**SD**	**Median**	**Range**	**Shapiro-Wilk**
PPVT	80	22.49	8.38	23	4-43	.900
WM task	80	6.16	2.49	6	0-11	.814
ToM Tasks	80	4.89	1.09	5	2-6	<.001

Table 4-2 indicates that the PPVT and WM scores were normally distributed ($p > 0.05$), justifying the use of parametric statistical methods for these variables. In contrast, ToM scores significantly deviated from normality ($p < 0.01$), suggesting that non-parametric approaches maybe more appropriate for analysing this measure.

Based on the analysis of standard deviation and score range, it can be inferred that the results of the PPVT and WM task showed relatively high variability. This reflects a wide distribution of scores among participants, thereby providing sufficient variance for meaningful statistical comparisons and inferences. In contrast, ToM scores showed a more limited distribution, indicating that most participants achieved similar results on this task. This reduced variability maybe attributed to the use of a single ToM tasks as the assessment criterion, which likely limited the ability to capture a broader range of ToM competencies.

To address this limitation, the main study adopted a more comprehensive assessment approach, incorporating the five-step developmental scale proposed by Wellman and Liu (2004) and the sarcasm comprehension task developed by Filippova and Astington (2008), in order to more accurately evaluate multiple dimensions of children's ToM performance.

4.2.6 Discussion

In summary, the results of the pilot study confirmed the feasibility of the selected materials, methods, and procedures for measuring children's early L2 learning in both rural and urban contexts. Specifically, based on the data analysis, the recruitment strategy and sample selection methods were found to be effective. Based on children's performance during the experimental process, the adopted PPVT was found to be a suitable tool for measuring children's English receptive vocabulary proficiency. Furthermore, the WM task and ToM tasks assessments were shown to effectively evaluate children's cognitive abilities. Although most of the materials and procedures were appropriate, several adjustments were made during implementation to improve overall effectiveness. Details of these modifications are presented in Table 4-3.

Table 4-3. *Details of Improvements in Pilot Study.*

	Pilot Study	**Main Study**	**Enhancement Description**
Grouping process	1-40: Experimental group 41-80: Control group	Each child participated in drawing differently coloured balls to determine their group	Aimed to ensure randomness and fairness of group assignment while mitigating bias influences
Before the experiment begins		Added a warm-up chat section	Engaged children in casual conversation to create a comfortable, relaxed atmosphere and alleviate potential anxiety
PPVT	The Chinese revised edition (black and white images)	Latest 5th edition (colour images)	Enhance children's attention and better reflect real-world visual information
ToM test	The false beliefs task (Wellman & Liu, 2004).	Five-step developmental scale (Wellman & Liu, 2004) Sarcasm (Filippova & Astington, 2008)	Enhance Assessment Accuracy: by introducing more assessment tools and tasks, the researcher can more accurately assess children's cognitive performance, leading to more comprehensive and in- depth research results

4.3 The Main Study

4.3.1 Participants

To ensure transparency and reliability in participant selection, a structured recruitment procedure was applied. Participants were recruited from Nanyang City, Henan Province, China. The following section sets out the recruitment and sampling procedures in detail.

Two primary schools in Nanyang were selected to represent both rural and urban contexts. Although the study did not aim to compare rural and urban populations, the inclusion of participants from both contexts was intended to ensure diversity in the sample and to enhance the generalisability of the findings. Children from different educational and sociocultural backgrounds were included to provide a more comprehensive understanding of the potential applicability of the intervention.

The rural school, located approximately 30 minutes from the city centre, has six classes (one per grade). The recruitment targeted students from Grades 1 and 2. The urban school, located in the city centre, serves a larger population with six classes per grade and approximately 60 students per class. Similarly, recruitment in the urban school focused on two randomly selected classes from Grades 1 and 2 to ensure a manageable and comparable pool of potential participants.

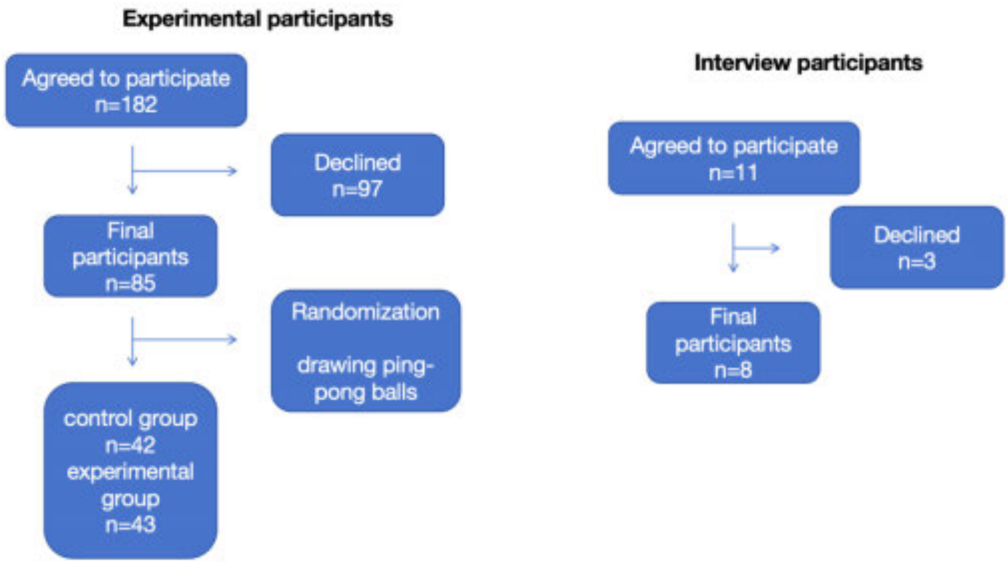
To ensure that children and their parent-caregivers were well informed about the study, the researcher prepared a Child Information Sheet and a Parent Information Sheet (see Appendix 1). In the rural school, these sheets were distributed to all Grade 1 and 2 teachers, who then passed them on to parents and caregivers. In the urban school, the information sheets were randomly distributed to two classes in each of Grades 1 and 2. The dissemination process was supervised by teachers to ensure that all families received accurate and comprehensive information about the research.

Participation in the study was entirely voluntary. After reviewing the study information, parents or caregivers returned signed consent forms to the class teachers, who then forwarded the forms to the researcher. All participants were informed that they had the right to withdraw from the study at any time without providing a reason.

4.3.2 Sample Selection and Grouping

In the main study, 85 participants were selected from a total of 182 returned consent forms using a systematic recruitment procedure to ensure fairness and representativeness (see Figure 4-1).

Figure 4-1. *Recruitment process.*



All consent forms were first reviewed to confirm that participants met the study's inclusion criteria, particularly being between 5 and 7 years of age. From the eligible pool, 85 participants were randomly selected using a computer-generated random number system. Among them, 40 were from rural schools and 45 from urban schools.

An a priori power analysis was conducted using G*Power 3.1 (Faul et al., 2007) to determine the necessary sample size for the primary regression analysis. A medium effect size ($f^2 = 0.15$) was anticipated based on Cohen's (1988) guidelines. With an α error probability of .05, a statistical power of .80, and four predictors included in the model, the calculation indicated that a total sample size of N = 85 was required. Consequently, 85 participants were recruited to ensure the study was adequately powered.

To further validate the robustness of this sample size N = 85, a post-hoc sensitivity analysis was performed to ascertain the smallest effect size the study could reliably detect. With the significance level set at $\alpha = .05$ and power at .80, the analysis confirmed that the change-score regression model (with four predictors) was sensitive to a minimum effect size of $f^2 = 0.149$. This value aligns closely with the standard definition of a medium effect size ($f^2 = 0.15$; Cohen, 1988). Collectively, these analyses demonstrate that the study is robustly powered to detect meaningful statistical relationships.

To assign these 85 participants into the experimental and control groups, a ping-pong ball draw method was employed. A total of 85 balls were prepared in advance: 43 of one colour (representing the experimental group) and 42 of another colour (representing the control group). All balls were placed in an opaque container and thoroughly mixed. Each child was invited to draw a ball without knowing its colour. The colour of the ball determined the participant's group assignment. The drawing sequence was not associated with any demographic variables such as gender or school type, and children drew in the order of their registration.

This approach constitutes a form of constrained random assignment, designed to preserve the fairness of randomisation while ensuring balanced group sizes. Although not a fully unrestricted random process, it minimises researcher interference and maintains equal assignment probability for each participant. Additionally, the method is child-friendly and ethically transparent, as it allows for active, voluntary participation in group allocation.

As a result, 43 children (25 males, 18 females) were assigned to the experimental group, and 42 children (20 males, 22 females) to the control group. Demographic details for both groups are presented in Table 4-4.

Table 4-4. *Detailed data descriptions: Experimental Phase.*

Group	**Male (n)**	**Female (n)**	**Age Range**	M_{Age} **(SD)**
Experimental	25	18	61-94	78.09 (7.32)
Control	20	22	60-95	80.21 (8.18)

Note. Age data measured in months. SD in brackets.

All allocated participants received the intended intervention and completed both pre- and post-intervention assessments. Consequently, data completeness was 100% across all outcome measures (PPVT-5, working memory task, and ToM tasks). This zero-attrition rate is attributable to the study design, as the intervention sessions were integrated into the children's regular school timetable. This structural arrangement ensured consistent attendance and avoided the dropout often associated with voluntary after-school programmes. Accordingly, the final analysis was conducted on a complete-case basis with all 85 randomised participants.

To support the qualitative data of the study, parent-caregiver participants were recruited from both rural and urban primary schools. Eleven parents and caregivers' volunteers submitted their contact details through a questionnaire. Due to time constraints and analytical considerations, eight individuals were selected using a computer-generated random number system to reduce selection bias. Efforts were made to ensure diverse representation from both school contexts.

The selected parents and caregivers were contacted to confirm their participation and to schedule interview times. The remaining three were thanked for their interest and placed on a waiting list in case any selected participants withdrew. In the end, all eight interviews were successfully completed. The final group included one grandfather, two rural mothers, two urban fathers, and three urban mothers.

This recruitment process aimed to enhance the diversity and contextual depth of the qualitative data while ensuring ethical research standards through transparent communication and informed consent. The inclusion of participants from different school settings enabled the study to reflect a broader range of parent-caregivers' perspectives on children's L2 learning experiences.

4.3.3 Data Collection Instruments

This research employed six research instruments to conduct an in-depth investigation of the research questions. Four instruments were used to assess the learners, including the PPVT, WM task, ToM tasks, and a child-friendly questionnaire designed to assess their hedonic experience with the AI-based language learning app. The PPVT, WM task, and ToM tasks were conducted before and after the intervention to allow for comparison between baseline and post-test performance. This methodological approach enabled the evaluation of the effectiveness of the learning and intervention strategies applied to the participants over a set period. In addition, parents and caregivers completed two instruments: a demographic questionnaire and semi-structured interviews. In the following section, each instrument is described in detail to ensure transparency and replicability.

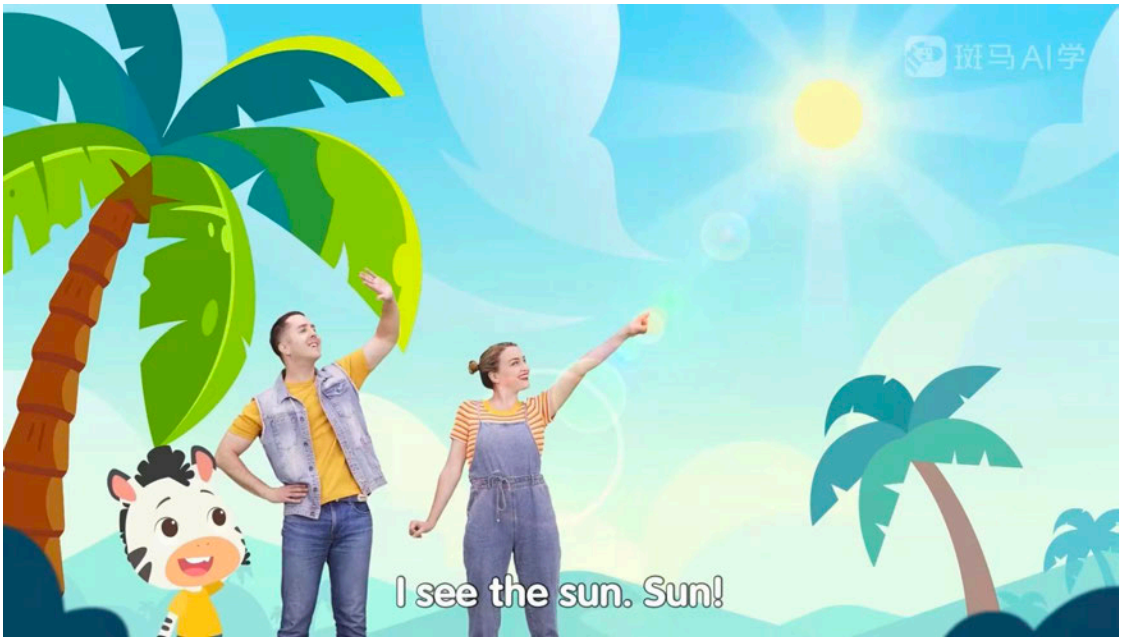
4.3.3.1 Instruments for the Quasi-Experimental Study Data Collection

4.3.3.1.1 Description of Intervention and Control Applications

In this study, Zebra AI (https://banmaapp.com) was selected as the digital tool for a vocabulary intervention experiment in the context of Chinese children's English learning. Developed by the educational technology unicorn Yuanfudao, Zebra AI is currently one of the largest online learning platforms for children in China, with over 20 million users. Zebra AI, launched with investment from Tencent in 2017, has become a leading online English education brand in China.

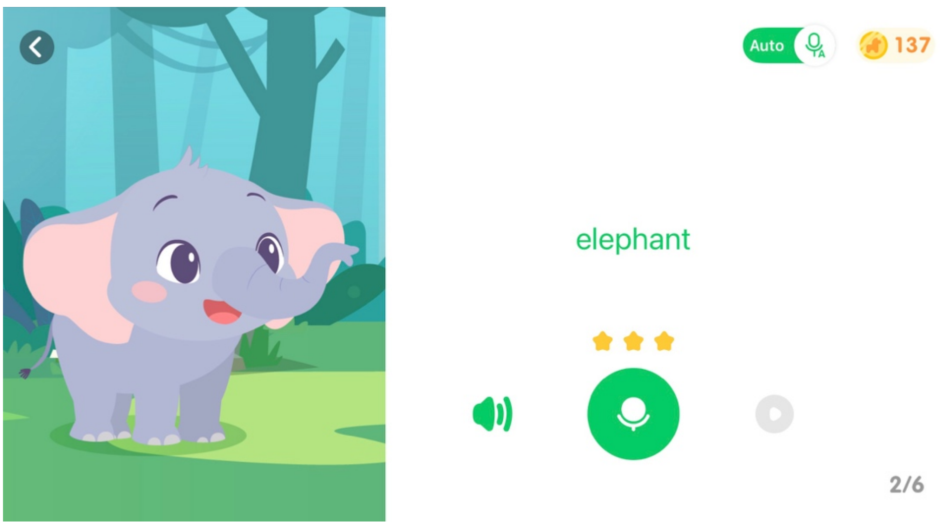
Zebra AI is specifically designed for children aged 3 to 8. Its curriculum covers three core areas: language literacy (English and Chinese), STEM, and art. This study focuses specifically on its English language learning module. The platform integrates multiple AI technologies and design features customised for children's cognitive development, aiming to enhance language learning outcomes and user experience. Specifically, when learning new words, Zebra AI facilitates multimodal interactions by integrating images, animations, and touch-based responses. Its human-like virtual characters (e.g., "Zebra Teacher"; see Figure 4-2) support engaging and immersive interaction through voice, facial expressions, and actions, creating an emotionally supportive environment that facilitates language acquisition in a social context.

Figure 4-2. *Screenshot of Zebra Teacher.*



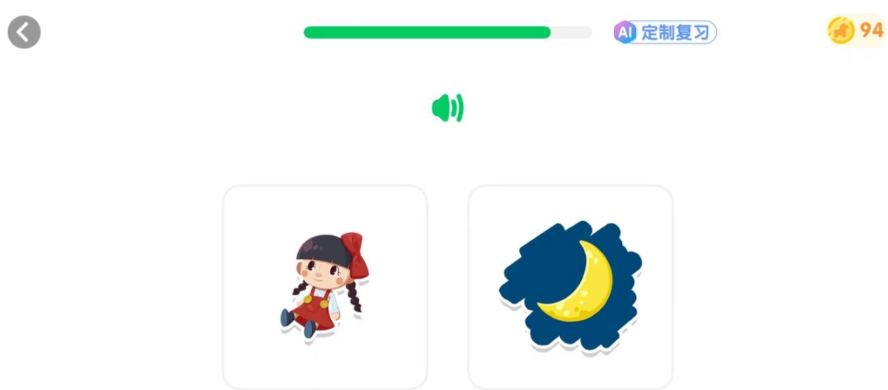
At the core of the system is the ASR function, calibrated specifically for young children's speech. It evaluates the learner's oral production by comparing the acoustic features of their utterances with native-speaker models, providing immediate and developmentally appropriate feedback. Accurate pronunciation results in positive reinforcement, such as the animated 'Three Stars' reward (see Figure 4-3), whereas significant phonological deviations trigger corrective scaffolding, such as 'Try Again' message. Progression through the learning pathway is similarly governed by mastery: the system controls pacing by requiring repetitions if the learner fails to meet the predefined acoustic threshold.

Figure 4-3. *Screenshot of Three Starts.*



Furthermore, within the Quiz module, Zebra AI incorporates an adaptive remediation mechanism known as the 'AI-Customised Review' (see Figure 4-4). Unlike static feedback systems, this feature leverages algorithms to analyse learner performance in real-time, identifying specific knowledge gaps based on error patterns during the assessment. Consequently, the system dynamically generates a personalised review sequence that targets the user's individual weaknesses. This approach facilitates precision learning by shifting the focus from generalised repetition to targeted reinforcement, thereby optimising the efficiency of the revision process.

Figure 4-4. *Screenshot of AI-Customised Review.*



Regarding to the conversational interface, it is important to note that although the broader Zebra AI includes generative NLP chatbots at higher proficiency levels (e.g., Level S4; see Figure 4-5), the children in this study were beginner learners aged 5–7 who had not yet reached this stage. Consequently, these generative features were not used.

Figure 4-5. *Photo of Zebra AI Chatbot.*

In addition, in the weekly Friday video lessons, Zebra AI employs a hybrid multimodal interface, which integrates a real-human instructor within a dynamic virtual narrative environment. This is complemented by the simultaneous presentation of animated imagery and target vocabulary (Figure 4-6). Such a design adheres to the Coherence Principle and the Signaling Principle of multimedia learning. Through the embodied guidance provided by the human instructor, the system effectively reduces the cognitive load on young learners when processing complex visual information, whilst significantly enhancing the sense of immersion.

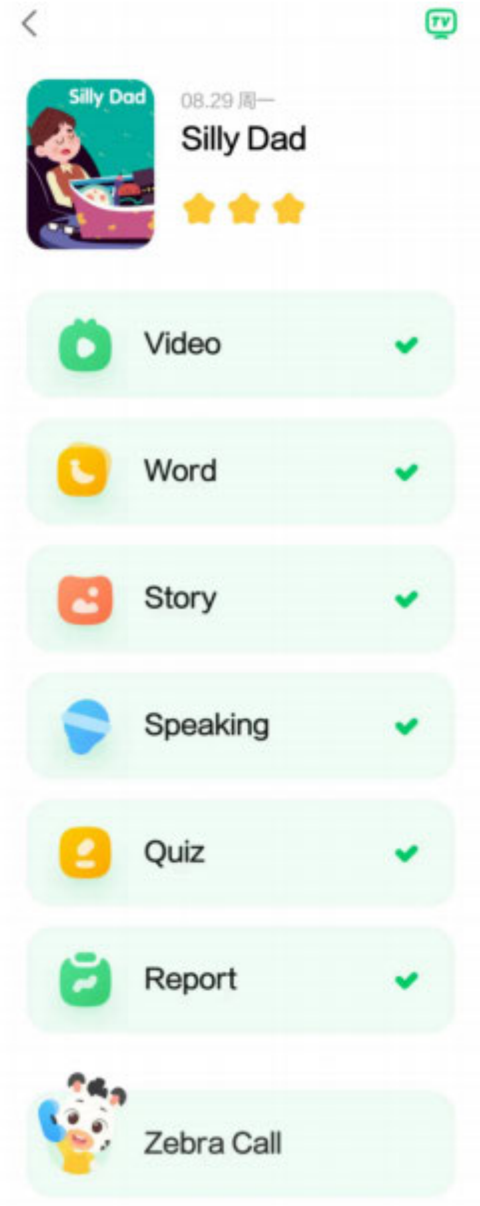
Figure 4-6. *Screenshot of multimodal interface.*

Finally, the system includes a background learning-analytics engine that collects performance data, such as phonetic accuracy, task completion times, and vocal output frequency (speaking counts). These data are automatically compiled into daily summary reports accessible to parents, providing a visual overview of each child's lexical mastery and overall learning trajectory (see Figure 4-7).

Figure 4-7*. Screenshot of Learning-analytics.*

The English course was divided into six progressive learning stages from S1 to S6, each lasting 48 weeks. Given that the participating children were aged 5 to 7 and most of them with no previous knowledge of English, the research selected a 12-week course comprising three units from the S2 stage as the instructional material. Each weekly module consisted of five daily classes from Monday to Friday. New content was introduced from Monday to Wednesday, followed by review sessions on Thursday to help consolidate the learners' understanding. On Friday, learners were invited to participate in a hybrid multimodal video lesson. Each class lasted approximately 15 minutes and could be accessed at any time via mobile devices. Each class consisted of seven parts: a) video, b) word, c) story, d) speaking, e) quiz, and f) report g) zebra call (see Figure 4-8).

Figure 4-8. *Structure of a single Zebra AI English lesson.*



a) **Video**: This part introduces the topic of the lesson through an animated teaching video, usually hosted by a virtual teacher character (such as “Zebra Teacher”). The content includes core vocabulary, basic sentence patterns, and contextual situations, aiming to stimulate children’s interest and attention.

b) **Word**: Key vocabulary is presented through a combination of text and images, supplemented by standard pronunciation and contextual examples, to reinforce understanding through dual- channel (visual and auditory) input.

c) **Story**: This part strengthens language input through short and engaging stories while guiding children to follow and read aloud sentence by sentence. The story content is consistent with the lesson’s core vocabulary and grammar structures, as well as incorporates daily life contexts and basic social interactions to support contextual understanding.

d) **Speaking**: This part guides children to practice pronunciation of target language items introduced in the video. The system adopts ASR technology to provide real-time feedback on pronunciation accuracy, thereby enhancing children’s speaking skills.

e) **Quiz**: This part provides immediate assessment through mini-games, multiple-choice, or matching tasks to evaluate children’s mastery of key vocabulary and listening and speaking skills. The system gives instant feedback after each response and rewards ‘little stars’ as positive reinforcement to promote motivation and sustained engagement.

f) **Report**: At the end of each class, the system generates a personalised learning report, including task completion, accuracy rate, pronunciation score, and system recommendations, enabling parent-caregivers or teachers to track progress and make timely instructional adjustments.

g) **Zebra Call**: This feature mimics the visual and auditory cues of an incoming video call (e.g., the ringing tone and answer button). Upon 'accepting' the call, the learner enters a pseudo-synchronous interaction with a pre-recorded native English speaker. While the content is pre-scripted, the first-person camera angle and direct address are designed to create an authentic communicative scenario.

It is important to note that the Zebra AI used in this research was a commercial language learning product that required an annual subscription. A subscription was normally required to gain full access to its content. During the intervention phase of the study, the three learning units assigned to children in the experimental group were selected from the free trial content offered by the application at that time. These units covered core functions such as vocabulary learning, pronunciation practice, and speech recognition feedback. Although the application was a paid system, this study did not involve any paid features or financial transactions. All participants were able to complete the intervention tasks without financial burden, thereby ensuring consistency in experimental conditions and comparability of the data.

In contrast, children in the control group played *Candy Crush Saga*, a widely used non-educational mobile game. This application was selected as a time- and engagement-matched control activity to avoid any exposure to language-related content. Children in the control group used the application on the same schedule as the experimental group, for 10 to 15 minutes per session, three to five times a week. This arrangement ensured that both groups had comparable digital usage time, while isolating the educational effect to the specific content of the intervention tool.

However, it should be noted that this control-group design primarily accounted for screen time and exposure to non-educational content, but it was unable to effectively isolate the specific technological mechanisms within the Zebra AI (e.g., AI-driven immediate feedback) from the application's core educational content. Consequently, the present design is better suited to examining the overall content effect of the Zebra AI learning system, rather than the independent contribution of its individual technological features.

4.3.3.1.2 Peabody Picture Vocabulary Test

Vocabulary plays a crucial role in language learning (Alqahtani, 2015). As a foundation for reading and comprehension, vocabulary acquisition is essential for expanding language knowledge and developing communicative skills. It is also strongly associated with learners' overall language proficiency (August & Shanahan, 2006; Baker, 2014). As discussed in the previous sections, receptive vocabulary plays a particularly important role in the early stages of SLA, especially among young children (Poarch & Van Hell, 2012). Compared to productive vocabulary, receptive vocabulary typically emerges earlier in language development. It reflects children's ability to understand, recognise, and semantically process language input before they are able to produce language themselves (Florit et al., 2009). Therefore, it serves not only as a key indicator of language comprehension but is also commonly regarded as an important predictor of children's future language development potential (Bialystok et al., 2010; Dongsun Yim et al., 2016).

Based on the above, receptive vocabulary tests have been widely used in studies involving young bilingual children and L2 learners, serving as an important tool for evaluating their language development. These assessments are employed not only to measure children's vocabulary comprehension in a specific language, but also to indirectly assess their overall language proficiency and the development of their language processing abilities (Bialystok et al., 2010; D. Yim et al., 2016).

Among the various instruments designed to assess children's receptive vocabulary, the PPVT is one of the most widely recognised and employed tools in both research and practice. Initially developed by Dunn (1959), the PPVT has been revised multiple times and is currently in its fifth edition (Dunn, 2019). It has been extensively applied in studies on monolingual language development, bilingualism, and SLA (Cohen et al., 2020). The test evaluates children's comprehension of spoken vocabulary by asking them to select the corresponding image from a set of options, making it developmentally appropriate and accessible for young learners.

The PPVT format offers several advantages for assessing young children. It does not require advanced literacy skills or verbal responses; children simply point to the picture that best represents the word they hear. This simplified format reduces the linguistic load and minimises test-related anxiety or fear of failure (Yim et al., 2016). Moreover, the use of pictures in test items enhances visual engagement and is especially effective for maintaining children's attention and motivation throughout the assessment (Dahl & Vulchanova, 2014). In the main phase of this study, the fifth edition of the PPVT was used, which features updated vocabulary items and full-colour illustrations. These improvements make the test more appealing and better aligned with children's real-world experiences, thereby enhancing the validity of the assessment.

Given that the target participants in this study were Chinese children aged 5 to 7, who are at an age group in which expressive language skills are still developing, the PPVT was selected as a suitable tool due toits standardisation, ease of administration, and developmental appropriateness. In addition, the test's widespread use in international research provides a strong methodological foundation for ensuring comparability and reliability across different studies. Therefore, the PPVT served as the primary measure for assessing receptive vocabulary in both the pre-test and post-test stages of this intervention.

However, it is important to acknowledge a potential limitation of the PPVT: its test items are rooted in Western cultural contexts (e.g., the United States and the United Kingdom) and may contain culturally specific content. For children from different cultural backgrounds, particularly Chinese children. This could lead to misunderstandings or unfamiliarity with certain items. For example, the word *salad* is common in Western food culture but may be unfamiliar to many Chinese children, potentially affecting their ability to respond accurately (Pasquarella et al., 2011).

To reduce the influence of cultural differences on test validity, the researcher provided brief explanations in Mandarin when children showed confusion about specific items. These clarifications were limited to objective descriptions of the visual or situational context, and great care was taken to avoid giving any clues that might influence the child's choice. This approach aimed to achieve a balance between maintaining the standardisation of the test and ensuring that participants could meaningfully engage with the assessment content.

The experimental instrument used in this study was the PPVT-5 (Dunn, 2019), a standardised tool designed to assess individuals' receptive vocabulary in English, suitable for participants aged 2 years and 6 months to over 90 years. The PPVT-5 demonstrates strong psychometric properties and has been widely applied in both monolingual and bilingual language research. The test has shown high levels of reliability and validity across diverse linguistic and cultural contexts. According to the official manual, the PPVT-5 has an internal consistency reliability (Cronbach's α) of 0.97, indicating a high degree of coherence among test items. Its test–retest reliability coefficients typically exceed 0.90, reflecting strong score stability over time (Dunn, 2019). In addition, prior research has reported significant positive correlations between the PPVT-5 and other standardised assessments of language ability, such as the CELF (Clinical Evaluation of Language Fundamentals) and the Verbal Comprehension Index of the WISC (Wechsler Intelligence Scale for Children), supporting both its construct and criterion-related validity (Dunn, 2019).

Furthermore, the applicability of the PPVT-5 to Chinese-English bilingual children has been confirmed in a recent validation study. Ji et al. (2022) employed a cross-classified mixed effects model to analyse data from a sample of Mandarin- and Cantonese-speaking children. The results indicated extremely high reliability (0.99) for the test in this population. The study also provided multiple sources of validity evidence, including demographic fairness (e.g., no significant gender or language-group bias), lexical appropriateness, and alignment between item structure and developmental language acquisition patterns. These findings further support the scientific robustness and cultural applicability of the PPVT-5 as a measure of receptive vocabulary in children with a Chinese linguistic background.

The full test consists of 240-word items (i.e., auditory stimuli), each provided with a set of four coloured illustrations. The child is asked to select the picture that best corresponds to the word spoken by the researcher. The test items span multiple lexical categories including nouns (e.g., animals, objects), verbs (e.g., actions), and adjectives (e.g., emotions or descriptive terms). These items are presented in order of increasing difficulty and organised into 13 item sets consistent with developmental age bands, allowing for a systematic assessment across a wide ability range.

The administration of the test followed standardised procedures based on the participants' ages, as recommended by the PPVT-5 manual: children aged 5 began at item 26, those aged 6 at item 53, and those aged 7 at item 66. This approach ensured that the starting difficulty level was appropriate and reduced the likelihood of presenting overly easy or difficult items. If a child answered three consecutive items correctly at the starting point, testing continued forward until reaching the ceiling criterion, defined as six or more consecutive incorrect responses. Items below the established basal level (i.e., the point at which three correct responses occurred consecutively) were automatically scored as correct. Consequently, not all 240 items were administered to each child; only a tailored subset was presented, corresponding to the individual's comprehension level.

According to the PPVT-5 administration manual, raw scores are calculated using a binary scoring system. All items below the basal level are assumed to be correct and awarded full marks, while the number of correctly answered items between the basal and ceiling sets is added to this total. Each administered item above the basal level is scored as 1 (correct) or 0 (incorrect), and the raw score is derived accordingly.

In terms of data analysis strategy, this study prioritised raw scores over standardised scores adjusted for age norms. The primary rationale for this decision lies in the sensitivity of the metric: while standardised scores are predominantly designed to assess an individual’s relative position within a peer cohort, raw scores provide a more direct quantification of the absolute magnitude of growth in children’s vocabulary. Particularly within the context of the short-term intervention and within-subject design employed in this study, raw scores effectively circumvent the risk of age-based corrections masking actual, subtle linguistic gains. This methodological approach is consistent with the findings of (Sullivan et al., 2014). To rigorously validate this methodological choice and rule out potential confounding effects of developmental maturation, a robustness check controlling for age and baseline proficiency was conducted (for detailed results, see Section 5.1.1.2).

An illustration of a typical PPVT-5 test item is shown in Figure 4-9. Each page includes four images, and the child selects the one that matches the spoken word.

Figure 4-9. *PPVT-5 word example stimuli - A single image contains four objects.*

4.3.3.1.3 Working Memory Task

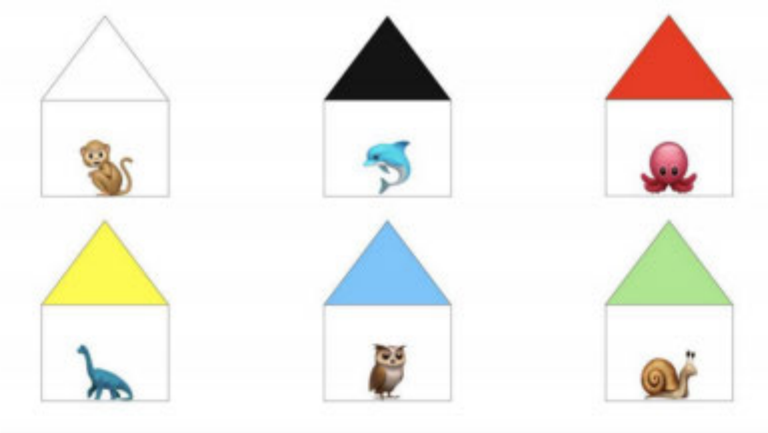
WM supports children in forming and retaining connections between distinct informational units, and is therefore considered a core element of cognitive ability underlying language aptitude (Ellis, 1996). Within language acquisition, the associative aspect of WM plays a critical role as it facilitates the binding of arbitrary phonological forms to their visual referents. In the context of L2 learning, the relationship between associative capabilities and vocabulary acquisition is fundamental. Prior research indicates that learners with superior associative learning abilities tend to possess larger vocabularies and enhanced language proficiency (Baddeley, 1998). This is particularly relevant during the early stages of SLA, where the ability to synchronously process and bind unfamiliar phonological forms to meanings supports successful word learning (Gathercole, 2006; Nakata, 2011).

To examine whether the AI-based language learning application influences children's cognitive development, this study assessed WM. Given the young age of the participants (5–7 years) and their varied English proficiency, a non-verbal WM binding task adapted from Willoughby et al. (2012) was used to specifically target this narrower associative capacity. Unlike simple digit span tasks, which mainly assess sequential storage, this task requires children to bind visual items (animals) to specific contexts (coloured houses). This associative form of WM binding is conceptually distinct and highly relevant to vocabulary learning, as word acquisition depends on forming arbitrary links between phonological forms and their referents (Cowan et al., 2005; Gathercole et al., 1997). In addition, the task reduces reliance on literacy and advanced language skills, making it suitable for measuring core associative memory processes in young L2 learners (Alloway et al., 2004).

This task has been widely used and validated in previous research with preschool and early primary school children (Torrington Eaton & Ratner, 2016; Willoughby et al., 2012). Studies have reported high internal consistency and strong correlations with other executive function and academic outcome measures, providing evidence of both its reliability and construct validity. Moreover, its engaging and visually appealing format helps to maintain children's attention and motivation during testing.

In this study, the task was administered using PowerPoint slides adapted from the original version developed by Willoughby et al. (2012). Prior to the test, the researcher introduced the children to six animal figures and six roof colours and asked them to name each item to ensure familiarity with the materials, thereby ensuring that the subsequent recall depended on the association rather than the recognition of individual items. During the testing phase, children were shown images of six houses, each with a different coloured roof and a different animal ‘living’ inside (see Figure 4-10). They were given one minute to establish and memorise each animal–colour pair. The detailed testing procedure is described in the following section.

__Figure 4-10.__ Example of the WM task – the pairings of all animals and their different coloured houses.



4.3.3.1.4 Theory of Mind task

ToM is a core construct in children's social cognitive development and has been widely recognised as a foundational element for successful social interaction (Peterson et al., 2005). ToM is defined as the capacity to attribute mental states, such as beliefs, desires, intentions, and emotions, to oneself and others, and to use this understanding to interpret and predict behaviour (Milligan et al., 2007). Extensive research has established a close and complex relationship between ToM and language development, though this relationship is neither linear nor unidirectional (de Villiers, 2007; De Villiers & de Villiers, 2014). On the one hand, language provides the representational tools necessary for expressing and reasoning about mental states; on the other hand, effective language use presupposes a certain level of ToM ability (Siegal, 1999). Morgan and Kegl (2006) further argue that language development is not merely correlated with but may scaffold the cognitive architecture for mental state reasoning. While meta-analyses such as Milligan et al. (2007) report moderate to strong correlations between language and ToM, the mechanisms underlying this association remain contested, especially across different cultural contexts.

Building on these empirical insights, AI-based language learning applications which often simulate interactive dialogue and context-rich communication, create new possibilities to examine whether technologically mediated language input can influence children's social cognitive development. However, empirical support for the effectiveness of AI-based learning tools in enhancing ToM remains limited. This research aims to address this gap by assessing children's ToM abilities before and after exposure to AI-supported language learning experiences.

To examine this, the research adopted the five-step developmental scale originally proposed by Wellman and Liu (2004), using the Chinese version adapted by Wellman et al. (2006). This scale reflects a theoretically grounded developmental progression of ToM, beginning with the understanding of desires and moving through belief, knowledge access, hidden emotions, and first-order false beliefs. The tool demonstrates strong psychometric reliability. Wellman et al. (2011) included a longitudinal sample of 31 preschool children from Beijing, assessed between ages three and five. To ensure linguistic and cultural appropriateness, the researchers localised task materials (e.g., replacing the Band-Aid box used in U.S. tasks with a familiar potato chip tube) and adapted task wording to match Mandarin language structure. Results showed that Chinese children generally followed the same developmental sequence, with minor variations (e.g., earlier acquisition of knowledge access), confirming the scale’s cross-cultural validity. These findings provide a strong empirical foundation for using the Chinese version of the ToM scale in the present research.

To enhance the sensitivity of the measure for older children within the target age range of 5 to 7 years, the study incorporated an additional sixth task: a brief ironic vignette developed by Filippova and Astington (2008), in which a picnic is disrupted by unexpected rain. This story-based task introduces cognitive and emotional conflict, allowing for the assessment of more advanced ToM skills such as irony comprehension and affective perspective-taking.

In this study, ToM is operationally defined as the child’s ability to identify and reason about mental states through task performance. To ensure scoring precision, coding criteria were differentiated based on task complexity. Specifically, the first five standard tasks employed a binary scoring system (0 = incorrect, 1 = correct), yielding a maximum sub-score of 5, whilst the sixth task (irony comprehension) used a three-point scale (0 = incorrect/irrelevant, 1 = partially correct, 2 = fully correct) to capture nuances in understanding, yielding a maximum sub-score of 2. Consequently, the total score for the ToM measure ranged from 0 to 7. All responses were recorded verbatim and coded by the researcher according to established rubrics (see Appendix 2). Regarding the statistical treatment, the aggregate total score was treated as an approximately interval measure in the subsequent analyses. This decision was supported by an inspection of the pre-test score distribution (N = 85), which demonstrated normality across both the experimental (Skewness = 0.24, Kurtosis = -0.72) and control groups (Skewness = -0.15, Kurtosis = -0.18), thereby meeting the assumptions for parametric analyses.

To further validate the sixth task, a small pilot test (N = 12) was conducted prior to the main study, confirming its comprehensibility and age-appropriateness. While the irony task has not yet been widely validated in other cultural contexts, it was treated as an exploratory, non-core component in the present study, with appropriate analytical controls applied.

Taken together, this ToM assessment framework offers strong theoretical grounding, developmental sensitivity, and methodological rigor. It serves as a reliable tool for detecting nuanced changes in children's mental state reasoning potentially influenced by AI-mediated language learning.

The details of the five-step measurement and the sixth step of sarcasm are as follows: (1) Different Desires (different individuals may have different desires for the same things), (2) Different Beliefs (recognising that people may hold different, and possibly true, beliefs about the same things), (3) Knowledge Acquisition (understanding that certain information remains unknowable unless observed first hand), (4) Contents false-belief (people may have formed misbeliefs as a result of misinformation) (5) Hidden Emotions (people may hide their true feelings and expressing emotions externally inconsistent with internal feelings) (6) Sarcasm (people may experience emotions that are inconsistent with the emotions expressed in their words).

Each task followed a standardised structure and included:

1.At least one ToM tasks question to evaluate the child's level of social-cognitive reasoning.

2.A comprehension control question to check understanding of the literal meaning of the story.

3.In some tasks, an additional emotional control question to assess the child's ability to infer characters' emotional states.

For example, Task 6—a brief ironic vignette developed by Filippova and Astington (2008), and involved an illustrated short story (see Figure 4-11: Picnic in the Rain).

Figure 4-11. *Picnic in the Rain.*

The story describes how a little bear, and a little fox are caught in a sudden downpour during a picnic. The researcher read the story aloud in a neutral tone:

"The little bear and the little fox are going on a picnic. It's the little bear's idea. He says it will be a lovely sunny day. But when they take out the food, dark storm clouds appear. It starts raining, and all the food gets wet. The little fox says, 'Today is really a good day for a picnic.'"

After hearing the story, children were asked the following questions:

1.**Literal comprehension question**: "Do you think what the little fox said is true?'

2.**ToM tasks question**: "Why did the little fox say, 'Today is really a good day for a picnic'?"

3.**Emotion control question**: "Was the little fox happy about the rain?"

Children's responses were recorded verbatim by the researcher on a scoring sheet and were later coded according to pre-defined criteria. Each answer was scored on a three-point scale: 0 = incorrect or irrelevant, 1 = partially correct, and 2 = fully correct. All responses were scored by the researcher based on clearly established rubrics, which included detailed descriptors and representative answer examples for each task type. All scoring procedures were conducted strictly following standardised criteria to ensure consistency and comparability across participants. Detailed descriptions of each task and the related scoring guidelines are provided in Appendix 2.

4.3.3.2 Instruments for the Questionnaire Data Collection

Questionnaires are widely used in research due to their convenience and cost-effectiveness, particularly because they enable quicker and more straightforward data collection and analysis compared to verbal methods (Marton-Williams, 1986). Furthermore, questionnaires can be distributed to large numbers of participants, either in paper format or online, at relatively low cost.

There are two main types of questionnaires: closed-ended and open-ended (Taherdoost, 2019). Closed-ended questions ask respondents to choose from pre-determined options provided by the researcher, such as single-choice, multiple-choice, or Likert-type scales indicating intensity (Reja et al., 2003). This format makes the questions easier and faster to complete, reduces the writing burden on participants, and facilitates efficient data aggregation and analysis (Willams, 2007). However, a limitation of closed-ended questions is that they do not allow participants to express their views freely or explore issues in depth and detail (Reja et al., 2003).

Open-ended questions address this limitation by allowing respondents to articulate their thoughts in their own words, offering opportunities for more nuanced and potentially insightful responses (Sun et al., 2018). However, such responses tend to be longer and more varied, which may complicate the analysis process for researchers (Taherdoost, 2019).

In this research, two questionnaires were developed incorporating both closed-ended and open- ended items. The Demographic Questionnaire was used to collect basic background information about the participants, while the Hedonic Experience Questionnaire was designed to assess children's perceptions of different features of the AI-based language learning app. These questionnaires are described in greater detail in the following sections.

4.3.3.2.1 Demographic questionnaire

The demographic questionnaire was designed to collect background information about the participating children and their primary parent-caregivers, serving as an important component of the research's sample characterisation. Among the collected variables, children's age, a key developmental factor and was included in the main statistical analysis. Other variables such as gender, home language environment, family income, and parent-caregivers 'occupation were not directly analysed but were used to describe the composition of the participant sample. Although these variables were not part of the formal modelling process, they provide valuable context for understanding the socio-cultural diversity of the sample and offer a foundation for future research.

In the context of child development and SLA, individual and family background characteristics can influence language input, access to learning resources, and levels of parent-caregivers' involvement. Therefore, even if these factors are not included in the inferential analysis, they should still be systematically documented to ensure the comprehensiveness and rigour of the study.

To ensure data confidentiality while enabling linkage between questionnaire data and experimental outcomes, each participant was asked to enter a unique identifier (e.g., a combination of initials and birthdate digits). This approach allowed for secure data matching while maintaining participant anonymity.

The questionnaire included 11 items (see Appendix 3) covering aspects such as the child’s gender, age, home language use, family income level, parental occupation, any special developmental circumstances, and parental concerns regarding the child’s growth. The questionnaire was distributed in paper form and completed by the children’s parents or primary caregivers.

4.3.3.2.2 Children's hedonic experience questionnaire

To investigate children's user experience and acceptance of the Zebra AI language learning application, this study developed a child-friendly questionnaire based on the TAM and the UX model. The questionnaire was designed to assess four core dimensions: PEoU, PU, hedonic experience, and behavioural intention. Given the cognitive characteristics of young children (ages 5–7), the questions were constructed using simplified language and evaluated through visual analogue formats to enhance comprehensibility and engagement.

This study adopts Hassenzahl’s (2003) UX model as its core theoretical framework. This model conceptualises user experience as comprising two key dimensions: PQ and HQ.

PQ refers to the product's capacity to support users in achieving 'Do-Goals' (i.e., task-oriented objectives), which relates to its fundamental functionality and usability. Within the language learning context of this study, PQ is operationalised through core constructs from the TAM. These include: PEoU, reflecting the simplicity of operation; and PU alongside Comprehensibility, which serve to measure the application's effectiveness in supporting learning tasks (e.g., Immediate Corrective Feedback and content understanding).

Conversely, HQ concerns the product's capacity to fulfil user 'Be-Goals' (i.e., self-oriented needs), emphasising non-instrumental emotional rewards, such as enjoyment, aesthetic appeal, and motivational experiences.

Furthermore, this study employs Behavioural Intention as an outcome variable to assess the influence of both pragmatic and hedonic qualities on the user's propensity for sustained use. Grounded in this framework, a total of seven questionnaire items were designed to measure PQ, HQ, and Behavioural Intention, respectively. The specific items, their corresponding application features, and their theoretical mapping are detailed in the Table 4-5 below.

Table 4-5. *Theoretical Dimensions, Measured Constructs, Questionnaire Items, and Corresponding App Features.*

Theoretical Dimension	**Measured Construct**	**Questionnaire Item(s) (Q#)**	**Corresponding App Feature**
PQ	PEoU	Q1a: "Do you think the Zebra AI is easy or difficult to use?"	Overall app operation
	Content clarity	Q5b: "Do you think the images and videos are easy or difficult to understand?"	Visual learning materials
	PU	Q6: "Do you think the ASR-based feedback is useful or useless?"	ASR-based feedback
HQ	Stimulation	Q1b/c: "Do you think the Zebra AI is fun or boring/exciting or dull?"	Overall app
	Stimulation	Q4: "Do you think the games on Zebra AI are interesting or boring?"	Gamification features
	Visual Appeal	Q5a: "Do you think the images and videos are look great or look terrible"	Visual learning materials
Outcome variables	Intention to continue use	Q2: "Would you like to use it for a longer time or a shorter time?"	N/A
	Behavioural intention	Q3: "Would you like to use it again?"	N/A

In terms of measurement format, the questionnaire adopted the Smiley-o-meter, a visual analogue scale derived from the Fun Toolkit (Van Dijk et al., 2012). This child-friendly tool is an adaptation of the traditional Likert scale and has been widely used in child–computer interaction research due to its strong developmental appropriateness and ability to visually capture emotional responses. For young children, especially during early cognitive development, graphical rating tools, such as facial expressions are generally easier to understand and interact with than purely textual or numerical scales (Barendregt et al., 2006).

In this study, the Smiley-o-meter employed a five-point scale, ranging from 1 (a disappointed face) to 5 (a very happy face), allowing children to express their subjective experiences with different app features in a simple and engaging way.

To ensure cross-cultural and content validity, a rigorous, multi-stage development process was followed. Initially, the original English items underwent a forward-translation process by the researcher, followed by a review by a second bilingual colleague to ensure semantic equivalence.

To further establish face validity and content clarity, the draft questionnaire was reviewed by two supervisors. Based on their feedback, revisions were made to simplify wording and enhance age-appropriateness. Subsequently, a pilot study was conducted with five children (aged 5–7) prior to the main data collection. This pilot confirmed that the pictorial smiley-face indicators were clearly understood as representing ordinal degrees of enjoyment, and that the children could comprehend and respond to the items appropriately.

Following data collection. The reliability of the questionnaire was evaluated using Cronbach's alpha. The scale demonstrated adequate internal consistency for this age group ($\alpha = .66$). Although item analysis indicated that excluding question 6 ("Do you think the ASR-based feedback is useful or useless?") would increase the α to .73. However, this item was retained to ensure content validity. Given the AI-based nature of the intervention, the PU regarding the ASR-based feedback feature is a critical component of the children's overall user experience (Davis, 1989). Furthermore, in the context of exploratory research involving young children (aged 5–7), a reliability coefficient between .60 and .70 is widely considered acceptable (Hair et al., 2019).

Finally, regarding construct validity, although a full-scale factor analysis was not conducted due to the sample size constraints, the combination of a strong theoretical foundation, expert consultation, and successful pilot testing provides robust initial evidence for the questionnaire's validity.

The questionnaire was administered in paper format within a classroom setting. Children completed the questionnaire using pencils. To ensure full comprehension, the researcher read each item aloud and provided explanations as necessary. Children were encouraged to ask questions if they encountered any difficulties during the process. The entire activity typically took 5 to 10 minutes per participant. The English version of the questionnaire is presented in Table 4-6. It is worth noting that the version shown to children when completing the questionnaire was the Chinese translation.

Table 4-6. *Example of questions in the Hedonic experience questionnaire.*

1. Do you think the Zebra AI is		
easy to use		difficult to use
fun		boring
exciting		dull
2. Would you like to use the app		

for a longer time		for a shorter time
3. Would you like to use it again?		
Yes / A lot		No / Not at all
4. Do you think the games on Zebra AI are		
interesting		boring
5. Do you think the images and videos are		
look great		look terrible
easy to understand		difficult to understand
6. Do you think the ASR-based feedback is		
useful		useless

7. What do you like the most about the Zebra AI?
[Open-response option]

4.3.3.3 Instruments for the Interview Data Collection

Interviews provide a deeper understanding of participants' experiences and perceptions, thereby enabling researchers to obtain in-depth and detailed qualitative data (Rubin & Rubin, 2011). Therefore, interviews are considered an influential method for collecting qualitative data and are widely used in qualitative research (Gillham, 2001; Walliman, 2021). Powney and Watts (1984) defined an interview as a discussion between the researcher and a small number of participants on a topic of interest. By recording and analysing this interaction, researchers are able to explore participants' views and opinions on specific issues. There are three main types of interviews: structured, unstructured, and semi-structured (Kallio et al., 2016).

Semi-structured interviews were adopted in this study. The format allows researchers to generate follow-up questions based on participants' responses, thereby facilitating in-depth discussions. As a result, participants have greater opportunities to express their perspectives in a more detailed and meaningful way (Cachia & Millward, 2011). Given the significant influence of parent-caregivers on children's L2 development, this research considered not only the child's perspective but also the parent-caregivers' perspective regarding their child's use of AI-based applications during the intervention. The aim was to gain insight into how parent- caregivers perceived their child's learning experience, as well as the perceived advantages and disadvantages of using such apps for L2 learning.

Semi-structured interviews were used to obtain rich and detailed responses from parent-caregivers. However, this method also presented certain limitations. For example, some interviewees spent time discussing issues unrelated to the core themes of the study, and in a few cases, the researcher encountered difficulties in maintaining focus during the conversation. These challenges may have contributed to the collection of less relevant or inconsistent data (Schmidt, 2004). To avoid this, the interview questions were designed in accordance with the themes of the study, which also served as the framework for conducting the semi-structured interviews.

Following the quasi-experiment, eight parents or caregivers from the experimental group were invited to participate in semi-structured interviews to examine in depth their perceptions of their children’s use of the AI-based application. The discussion focused on three key aspects: the AI- based application’s impact on L2 learning, the children's overall learning experience, and the perceived advantages and disadvantages of the application.

To explore parent-caregivers' views on the role of AI-based language learning applications in children's L2 learning, this study designed a set of semi-structured interview questions. The questions aim to gain a comprehensive understanding of parent-caregivers' attitudes, experiences, and suggestions regarding the use of AI-assisted language learning tools.

This study adopts the TAM3 as the primary theoretical framework to guide the design and analysis of the interview component. Building upon the original TAM (Davis et al., 1989), TAM3 offers a more comprehensive approach by incorporating not only the core cognitive constructs, namely PU and PEoU, but also emotional, social, and individual difference factors, such as hedonic motivation, subjective norms, and prior experience. These factors are particularly pertinent to understanding technology acceptance behaviour.

The semi-structured interview outline in this study includes 10 open-ended questions organised into four thematic sections. The first module, which focused on overall attitudes and PU, explored parent-caregivers’ general views on AI-based language learning applications and whether they believed these tools were effective in supporting children’s L2 learning. The second section, centred on observations of learning outcomes and behavioural changes, focused on parent-caregivers’ perceived changes in their child’s language learning and how previous experience with similar technologies might have influenced current use. The third section explored parents’ evaluations of user experience and app design, focusing on features, usability, and children’s emotional engagement. It drew on TAM3 constructs such as PEoU, output quality, and hedonic motivation. The final section focused on usage barriers and suggestions for improvement, aimed to identify practical challenges in the app’s use and gather parent-caregivers’ recommendations for future AI-assisted language learning applications. Each question was carefully aligned with key constructs from TAM3, including PU, PEoU, hedonic motivation, subjective norm, prior experience, and cognitive load, to gain a comprehensive understanding of parent-caregivers' acceptance and attitudes toward AI-based tools in children's L2 learning, as shown in Table 4-7.

Table 4-7. *Table of Interview Questions and Their Corresponding TAM3 Constructs.*

Interview Question	**Thematic Module**	**TAM3 Construct(s)**
1. What do you think about using AI apps as a tool in children's L2 learning?	Overall Attitudes; PU	PU; Subjective Norm
2. Do you think this app is helpful for your child's L2 learning? Could you provide an example?	Overall Attitudes; PU	PU; Result Demonstrability
3. Do you think using the app has made a difference in your child's learning? If yes, what kind of change have you observed?	Learning Outcomes; Behaviour Change	Output Quality; PU
4. Has your child used any similar learning apps before? If so, do you think that experience influenced how they used this one?	Learning Outcomes; Behaviour Change	Experience-Related; Moderators; PU
5. In your opinion, what is the best feature of the app? Why do you like it?	User Experience; Design Evaluation	Hedonic Motivation; Output Quality
6. Did your child face any difficulties when using the app? Did they need your help?	User Experience; Design Evaluation	PeoU; Computer Anxiety; Perceived External Control
7. What do you think are the main pros and cons of this app?	User Experience; Design Evaluation	PU; PeoU; Hedonic Motivation
8. Roughly how long does your child use the app each day? Do they ever feel bored, overwhelmed, or tired of it?	Usage Barriers; Suggestions	Cognitive Load; Hedonic Motivation

9. Would you support your child in continuing to use this app in the future? Why or why not?	Usage Barriers; Suggestions	Behavioural Intention; Subjective Norm
10. What features would you like to see added or improved in future AI language learning apps?	Usage Barriers; Suggestions	Design Expectations; PEoU

While TAM3 serves as the main framework, the design of the interview questions is also informed by relevant insights from Multimodal Learning Theory (Mayer, 2002) and Parental Mediation Theory (Livingstone & Helsper, 2008), which help broaden the scope of inquiry. For example, questions comparing AI-based applications with traditional learning materials draw on the importance of sensory input emphasised in multimodal learning, while questions related to home usage challenges and parent-caregivers' support are informed by parental mediation theory, which highlights parent-caregivers' roles in managing children's digital media engagement.

4.3.4. Procedure

The data collection process in this study consisted of four main components, encompassing demographic profiling, pre- and post-intervention experimental tasks, an app-based language learning experience, and interviews with parents and caregivers.

The data collection was carried out in sequential phases. The baseline phase took place between September and October 2023 and included the demographic questionnaire and the first round of experimental tasks. The intervention phase began in October 2023 and lasted for 12 weeks. During this period, children in the experimental group used the Zebra AI language learning app at home on a daily basis for autonomous practice, while children in the control group engaged with the game Candy Crush as an equivalent non-linguistic digital activity. The post-intervention phase was conducted between January and February 2024, with the same assessments administered as in the baseline phase. The resulting data were used for comparative analysis. Parent interviews and questionnaires were completed within three weeks following the post-test to gather feedback on the intervention process and children's performance.

The detailed procedures are outlined below.

4.3.4.1 Demographic questionnaire

Parents and caregivers completed a demographic background questionnaire including information about the child's age, gender, home language exposure, and digital media usage habits. These data supported descriptive analyses and covariate control in subsequent modelling.

A total of 85 children and their parents or caregivers participated in the filling of this demographic questionnaire. The questionnaire was prepared in paper format by the researcher and provided to the teachers at the partner schools. To ensure the uniformity of the questionnaire distribution process and the accuracy of the interpretation, the researcher provided necessary guidance and instructions to the teachers before distributing them.

The teacher then distributed the questionnaires to the classes participating in the research and briefly explained to the students in class the purpose, importance, and operation method of filling them out. On the premise of obtaining the informed consent of the child, the student takes the questionnaire home and completes it under the assistance and supervision of the parents. All questionnaires are in an anonymous form, identified only by numbers, and do not involve any personal identity information to ensure the privacy and data security of participants.

After filling it out, students should return the questionnaires to the school on time. The teachers will collect them uniformly and return them to the researcher. Each questionnaire consists of 10 questions. It is expected that the filling time will be within 10 minutes. The questions are designed based on the principles of concise language and easy understanding to ensure that both children and parents or caregivers can complete them smoothly.

4.3.4.2 Pre- and Post-Intervention Experimental Tasks

All participating children individually completed three standardised tasks in a quiet classroom environment at their school. These tasks were designed to assess receptive single-word vocabulary breadth, WM, and ToM. The three tasks included: the PPVT-5, the WM Task, ToM Tasks.

Each task was administered at two time points: pre-intervention (T1) and post-intervention (T2) to measure change over time.

4.3.4.2.1 Peabody Picture Vocabulary Test

One of the main instruments used was a localised version of the PPVT-5, consisting of 240 words, each paired with four related images. The test was delivered via a standardised PowerPoint slideshow. Prior to the assessment, the researcher explained the procedures using age-appropriate language and emphasised that participation was voluntary and part of a fun activity. The goal was to build rapport, reduce anxiety, and foster a relaxed atmosphere that would support active engagement.

All assessments were conducted in a quiet room within the school to minimise environmental distractions. Each child was tested individually under comparable conditions to ensure consistency. Before each session, the researcher standardised the testing setup to maintain procedural uniformity.

During the test, each word was presented verbally by the researcher while four corresponding pictures were simultaneously shown on a slide. Children were asked to identify the picture that best matched the word, either by pointing to it or stating its corresponding number. The task took approximately 10–15 minutes per child. To support comprehension and participation, children were guided with playful, easy-to-follow instructions, such as:

"Let's look at some pictures together. I will say a word, and you can point to the picture you think matches it—or tell me the number. If you're not sure, just give your best guess."

Children's responses were recorded by the researcher using a binary scoring system, with correct answers coded as 1 and incorrect ones as 0. In addition to scores, the researcher also documented children's behavioural responses throughout the session. Observable reactions such as hesitation, silence, or unusual responses to specific items were noted to provide supplementary qualitative data. These observations offered valuable insights into individual differences in engagement and emotional response.

Informed consent was obtained from both the parents or caregivers and the children prior to participation. All data were anonymised and securely stored to ensure privacy and ethical compliance.

4.3.4.2.2 Working Memory Task

To ensure experimental validity and data reliability, a structured WM task was administered using the open-source software OpenSesame (https://osdoc.cogsci.nl) running on a standardised laptop. The software allowed for precise control of stimulus presentation and response timing, which supported consistent data collection across participants.

The experiment consisted of three sequential phases: a warm-up phase, a familiarisation phase, and a recall phase.

Before the task began, each child participated in a brief warm-up session with the researcher. This involved simple question-and-answer interactions designed to build rapport and reduce anxiety. The researcher also checked the child's familiarity with the colours and animals used in the task to minimise the risk of cognitive bias. This initial interaction helped the children feel relaxed, focused, and better prepared to engage with the subsequent testing activities.

In the familiarisation phase, participants were given one minute to memorise paired stimuli, which consisted of cartoon animals and houses with differently coloured roofs. The duration was controlled by the experimental software. All stimuli and instructions were age-appropriate and pre-tested with children of similar age to ensure clarity and engagement. This standardised exposure reduced individual differences in memory retention time, thereby enhancing result reliability.

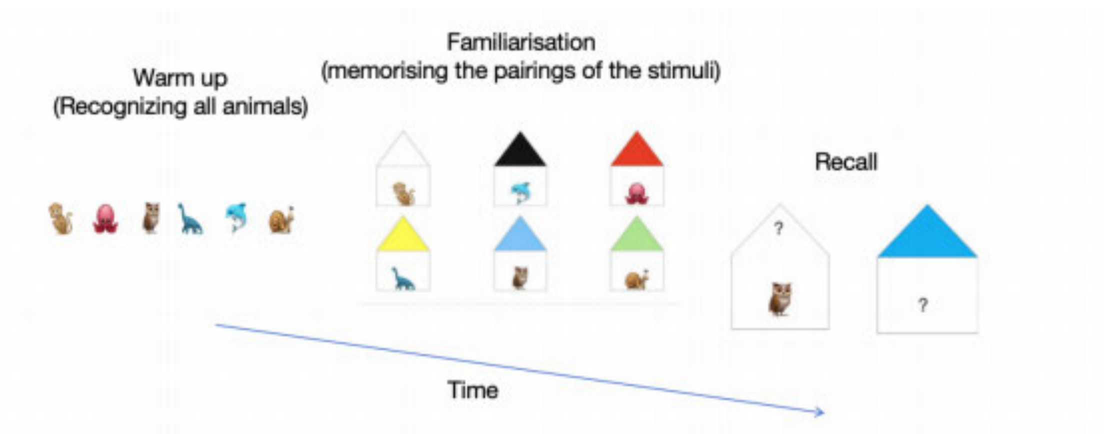
The recall phase was divided into two parts. In the first part, children were shown images of empty houses with coloured roofs and asked to recall which animal had lived in each house. In the second part, they were shown animal images and asked to recall the colour of each animal's house. The order of stimuli in both parts was pseudorandomised, meaning they were randomised while controlling for sequence and repetition effects. Each child completed six trials in total.

Children responded verbally, and the researcher recorded their answers in real time by pressing predefined keys on the keyboard. These responses were automatically logged by the software. The entire task lasted approximately 10 minutes per child, with each child tested individually in a quiet room to reduce external distractions.

In addition to accuracy scores, the researcher also observed and noted specific behaviours exhibited during the task. These included changes in facial expression, hesitation, or unusual verbal reactions. Such observations were documented as supplementary qualitative data, providing valuable insights into individual engagement levels and emotional responses.

A visual illustration of the three-phase procedure is provided in Figure 4-12.

Figure 4-12. *The three steps of the WM Task.*



4.3.4.2.3 Theory of Mind Tasks

To assess children's ToM abilities, this study designed and implemented six structured story- based tasks adapted from previous research (Filippova & Astington, 2008; Wellman & Liu, 2004). Each task was presented using a combination of text and illustrations and was designed to assess different levels of cognitive reasoning, including first-order and second-order beliefs, understanding of irony, and emotional attribution. The specific content of each story-based task is detailed in Section 4.3.3.1.4. The characters featured in the stories were toy animals and dolls familiar to children. The visual design of the figures reflected Chinese cultural features, such as black hair and East Asian facial characteristics, to enhance cultural relevance and familiarity.

All tasks were administered individually in a quiet classroom setting to ensure that children completed the tasks in a relaxed and distraction-free environment. Before the formal testing began, the researcher introduced the children to the toys, props, and story settings used in the tasks, in order to reduce novelty effects and promote better comprehension and focus.

The six tasks were presented in a fixed order, gradually progressing from relatively simple belief reasoning to tasks involving more complex inferential skills, such as irony. This scaffolded task structure was intended to help children gradually adapt to the testing process and better understand the task requirements.

4.3.4.3 App-Based Learning Intervention: Parent Log

Between the pre- and post-tests, children in the experimental group continuously used the Zebra AI language learning application over a 12-week intervention period. Each week, they were required to complete one learning section, which included new lesson content, review tasks, and interactive activities. Children were allowed to flexibly schedule their learning across 3 to 5 days per week, engaging with the app for approximately 10 to 15 minutes per session.

Given the ethical and technical constraints preventing the researcher from directly accessing the back-end usage data of the Zebra AI application, a structured home-monitoring instrument, the Zebra AI App Usage 12-Week Parent Log (see Appendix 4), was issued to parents in the experimental group. This tool was employed to systematically document the children's home learning activities throughout the 12-week intervention period. The log served as the core instrument for monitoring the intervention’s adherence and dosage, and its design incorporated three complementary data modules.

Firstly, the Daily Task Completion Record (Quantitative) required parents to mark, on a daily basis, whether the child completed the specified Zebra AI learning task (Yes/No). One completed task was defined as the minimum unit of usage, with the target dosage consequently set at five units per week, reflecting the application's five mandated usage days. It is important to note that although the log's original design included seven daily tracking slots to permit flexibility for make-up or extra learning on non-designated days (such as weekends), only the completion of the five specified task days was included in the subsequent scoring and analysis. This ensured the adherence metrics were consistent with the application's target dosage (five units per week).

Secondly, the Weekly Adherence Summary (Quantitative) was completed by parents at the end of each week, serving to confirm the full completion of the five required tasks and allowing for the calculation of the weekly full adherence rate.

Finally, the log featured a Qualitative Comments Section (Qualitative), which enabled parents to document the child’s learning disposition, engagement level, encountered difficulties, or any special circumstances that might have influenced usage and adherence. This provided essential contextual supplementation to the quantitative data.

All record sheets were collected following the intervention. The resultant data were used to calculate key adherence indicators, such as the total task completion rate and the average weekly days completed, forming a crucial basis for the descriptive analysis presented in the results section. By simultaneously providing structured quantitative data and explanatory qualitative insights, the Parent Log functioned as a vital source of process data. This enabled researchers to verify actual child participation, monitor the integrity of the intervention's execution, and ultimately enhance the interpretability of learning outcomes and the credibility of the study's conclusions.

In addition to usage monitoring, children's subjective experience was also assessed. At the end of the intervention, children in the experimental group completed a Hedonic Experience Questionnaire, which was adapted based on the UX framework. The questionnaire covered key aspects such as perceived enjoyment, ease of use, and willingness to reuse.

After the pre-test, children in the experimental and control groups installed two different applications on the mobile devices of their parents or caregivers and used them for the following 12 weeks. The experimental group used Zebra AI (https://banmaapp.com), an AI-based language learning application, while the control group used Candy Crush Saga (https://www.king.com/zh_CN/game/candycrush), a popular non-educational game.

The researcher provided all parents and caregivers with detailed installation instructions and download links to ensure the successful setup and configuration of the apps. To maintain the stability and responsiveness of the intervention process, the researcher established real-time communication channels with teachers, allowing them to promptly offer technical guidance and resolve any issues reported by parents during the installation or usage phases.

During the initial setup, teachers were invited to supervise the installation to ensure that all devices successfully ran the assigned applications under consistent conditions. Throughout the intervention, any technical difficulties reported by parents and caregivers were addressed in real time via phone, WeChat, or in-person support, thereby minimising disruptions to the intervention.

During the intervention, children in the experimental group were required to complete one standard learning section per week. Each section was systematically distributed by the Zebra AI platform, consisted of three types of tasks: new lesson input, review activities, and interactive feedback.

The weekly schedule was as follows: from Monday to Wednesday, children completed one new lesson per day, focusing on vocabulary acquisition, sentence comprehension, and pronunciation practice, each lasting approximately 10–15 minutes. On Thursday, they engaged in a review session designed to consolidate newly learned content and reinforce listening and speaking skills. On Friday, they were invited to participate in an AI-guided “Interactive TV Live Class” or a gamified “Weekly Challenge” activity, aimed at enhancing language input and contextual understanding in a playful format.

All children in the experimental group were required to use the application for 10 to 15 minutes per session, at least 3 days and no more than 5 days per week. The control group followed the same usage schedule with the designated game application to ensure consistency in intervention duration and comparability across groups. To accommodate individual learning rhythms, flexible scheduling was permitted, allowing children to choose the most suitable time slots for their learning tasks.

Throughout the intervention, parents and caregivers were encouraged to provide appropriate support during app usage and to observe their children’s engagement, including usage frequency, session duration, and notable behavioural patterns.

As the researcher was unable to directly access the backend usage data of the Zebra AI app, a structured weekly log sheet (see Appendix 4) was provided to the parents in the experimental group. They were instructed to mark daily task completion and provide brief comments at the end of each week to document their child's learning progress or any special circumstances. These record sheets were collected at the end of the intervention and served as a key reference for assessing children's task completion and participation throughout the study.

To systematically evaluate the children's subjective experience with the AI-based learning application, children in the experimental group completed the Hedonic Experience Questionnaire during the post-test phase, after the 12-week intervention. This questionnaire was based on the UX model proposed by Hassenzahl (2003) and incorporated the Smiley-o- meter from the Fun Toolkit framework. It was further adapted with age-appropriate illustrations to align with the cognitive and expressive abilities of children aged 5 to 7.

The questionnaire items were designed based on the content and interaction style of the intervention platform (Zebra AI), combining text with supportive images. All items were scored using a visual smiley-face scale. The final version was printed in colour on paper, making it easy for young children to understand and complete.

The questionnaires were administered in familiar classroom settings, with both researcher and classroom teachers present to provide supervision and guidance. Before completion, researcher explained each item in simple language and showed how to respond using the smiley-face scale. To accommodate individual differences in comprehension, additional assistance was offered as needed, and children were encouraged to ask questions throughout the process. The questionnaire activities were designed in a playful, non-test format to encourage a more relaxed and enjoyable learning environment.

Each child was given no more than 10 minutes to complete the questionnaire. Upon completion, all responses were collected immediately, anonymised with ID codes, and securely stored. The data were later entered into Excel for descriptive statistics and comparative analysis. These self- reported responses served as supplementary evidence of children's emotional engagement and user experience during the intervention, helping to explain observed differences in learning motivation and outcomes.

4.3.4.4 Parent-caregivers' Interviews

Following the intervention period, semi-structured interviews were conducted with parents and caregivers to explore their attitudes toward AI-based learning applications, their observations of the child's engagement and behavioural changes, and suggestions for future application development. These qualitative data complemented the quantitative results and offered valuable insights from the family context.

The interviewees were from urban and rural primary schools in Nanyang City, Henan Province, China. A total of eight parents and caregivers participated, including two mothers and one grandfather from a rural school, and two fathers and three mothers from an urban school. Among them, two parents were interviewed face-to-face, while the remaining six completed the interviews via phone or online voice messaging platforms such as WeChat.

Each interview was based on approximately ten core questions (see Appendix 5), with each session lasting around 30 to 45 minutes. The interview questions were developed with reference to existing literature on educational technology acceptance and user experience, and were adjusted to suit the research context, ensuring relevance and clarity.

Before the interview, researcher provided written information sheets and informed consent forms outlining the study's objectives, procedures, and data usage policies. Participation was entirely voluntary, and parents and caregivers were informed of their right to withdraw at any time without consequences. They were also assured that all interviews would be audio-recorded with their consent, and that the data would be kept strictly confidential and used solely for academic research purposes. These measures helped build trust and encouraged parents and caregivers to engage openly with the process.

During the interviews, researcher used open-ended and follow-up questions to obtain parent-caregivers' detailed opinions and emotional reflections. Their perspectives were respected without judgment or interference, ensuring the objectivity and reliability of the findings. At the end of each interview, researcher summarised the main points discussed to confirm understanding and strengthen participant engagement.

All interviews were transcribed verbatim, and the resulting texts were analysed thematically following Braun and Clarke (2006) framework. The interview data served as a valuable qualitative supplement to the study, offering insights into children's behavioural and emotional responses to AI-based language learning and providing important contextual support for interpreting the quantitative findings.

4.4 Reliability considerations

To ensure the reliability and consistency of data collected during the pre-test and post-test phases, this study implemented rigorous control measures across multiple aspects.

First, in terms of environmental control, both the pre-test and post-test were conducted in the same classroom setting, ensuring that all participating children completed the tasks under identical physical and environmental conditions. This approach minimised the influence of external variables on participants' performance.

Second, to maintain instrument and procedural consistency, the same testing tools, materials, and presentation formats were used in both phases. The sequence of test items remained unchanged, ensuring that all children followed the same order and procedures when completing the tasks.

The language used in instructions and prompts was standardised in advance, and no modifications were made to the difficulty levels or structure of the test content. To further avoid procedural errors, all test materials were pre-tested and reviewed to ensure clarity of content, precision of instructions, and developmental appropriateness for the target age group.

In terms of procedural steps, all tests were administered only by the researcher to ensure a high level of consistency and standardisation in task delivery. Before the formal testing, the researcher prepared a comprehensive set of testing guidelines, specifying the execution of each step, the sequence of tasks, the wording of instructions, and time management. All participants in both the pre-test and post-test phases were guided and tested by the same researcher, thus minimising potential variations caused by different administrators.

Additionally, for test segments requiring verbal guidance, the researcher employed pre-scripted language to maintain consistency in verbal expression across participants. This not only improved the clarity of instructions but also reduced the risk of variability in children's responses due to differences in delivery. The entire testing process was strictly consistent with the pre- defined procedures, and a test log was used to systematically document testing details and children’s responses to ensure that each participant completed the tasks under the same conditions.

To assess the stability of the measurements over time, this study also prioritised the control of test-retest reliability. The content and structure of the pre-test and post-test remained highly consistent, with no changes to the measured constructs (e.g., receptive vocabulary, WM, and ToM). The number of items, types of questions, scoring criteria, and administration procedures were identical across the two testing phases, ensuring that the measurements targeted the same underlying abilities and were comparable over time.

Moreover, the researcher maintained detailed observational records during the entire testing process, monitoring children's behaviour and responses in real-time to ensure accurate comprehension of instructions and independent task completion. For children who showed irregular behaviour or interruptions during the test, notes were made in the original records and taken into account during data analysis.

In summary, by controlling environmental conditions, standardising test tools and procedures, maintaining consistency in administration, and adopting multiple strategies to enhance reliability, this study established a strong methodological foundation for the comparability of pre- and post-test data, thereby ensuring the validity and trustworthiness of the experimental findings.

4.5 Ethical considerations

Before starting any research, researchers must first obtain approval from the ethics committee of their institution (Guillemin et al., 2012). Therefore, to follow the rules of the official organisation and to better guarantee the rights of the participants, the study was approved by Research Ethics Committee of the College of Social Sciences at the University of Glasgow (See Appendix 7). Also, as my research was conducted in two primary schools in Nanyang, China, I applied to the Education Department of Nanyang and the local community and received their approval to conduct the experiments locally and to collect data from children and parents (See Appendix 6).

The research aims to provide an evaluation of an AI-based language learning app for potential use in support of second langue learning in primary school-aged children. Although the evaluation of the app itself possesses minimal risks to all the participants and settings, the study should be considered some important ethical issues as some of the participants involved children between the ages of 5-7 years old. One of the most important issues is the safety of the participating children. To reduce this risk, the children were not taken away from their normal school activities as all experiments were being conducted during the day in the school classrooms.

And there was no scenario where the researcher was lone working/one-to-one working with a child. In addition to the researcher, a teacher who was familiar with the children was present to help deal with any emergencies that might happen. Also, public health recommendations were followed throughout the process. Data collected from children's interactions with the app was collected by the researcher during home activities at the time mutually agreed with the parents and children.

To reduce the stress that maybe caused to participating children, I provided children and parent-caregivers with a full verbal debriefing before participating in the experiment, including details of the methodology, the wider implications of the research questions, and the ability to answer any questions they may have. They were also given a written version of the information sheet explaining the study. When they signed the consent form the experiment was officially started. All experiments on children consist of fun animated pictures and stories and children will be asked to participate in games rather than being told to take tests. This is to make them feel relaxed. The format of the game will also avoid making them think that the results will relate to their educational outcomes. Furthermore, I also discussed with the teacher and the children the best time to take part in the quasi-experimental task to ensure that the children felt comfortable and in a good state of mind. They were all free to withdraw from the study at any time, either before or during the experiment. They were also told that this would not affect the assessment of their learning in any way, nor would it have any other negative impact on their relationships with others in the school.

To ensure that all participants had equal access to learning opportunities after the experiment, children in the control group were also provided with training and guidance on using the Zebra AI application after the post-test. This ensured that they received the same digital resources and support as the children in the experimental group. This measure was implemented to compensate for any information disparity caused by group assignment and to uphold the overall fairness and ethical integrity of the study.

As well as the personal safety of the children being considered, the safety of the researcher was also considered as the experiments were conducted in China. I created a schedule of fieldwork activities, and this was shared with the supervisors and a close relative living in data collection to ensure that the supervisors were aware of the researcher's whereabouts. Similarly, my contact details were known to the supervisor in case of an emergency. If the public health advice changes and there is any potential risk to the researcher or participants data collection will be terminated immediately. I familiarised myself with the area before fieldwork and had access to an up-to- date map of the area and avoided traveling outside daylight hours. If traveling at night, I should take a taxi to make sure safety.

During the data collection process, all data is collected and coded numerically rather than by the name of the participant to protect the privacy of the participants. Meanwhile, experimental data will be stored in computer files that can only be accessed by password and will be securely destroyed after analysis. When the experiment is complete, the data will be organised and structured using STR and SPSS software. All the files containing data will be organised on the computer provided by the university. Another copy will be saved on OneDrive (University server). The data will be destroyed once the data has served its purpose. Once the analysis and any publications resulting from the data have been completed, the data will be securely destroyed. Assurances regarding confidentiality will be strictly adhered to unless evidence of wrongdoing or potential harm is found, and participants will be informed that in such cases the University may be obliged to contact the relevant statutory body/authority. All data used in the thesis will be reviewed by the supervisors to ensure that any identifying information is removed so that the risk of identifying the participant is minimised.

4.6 Conclusion

This chapter presented a comprehensive overview of the research methodology employed in this research. It began by outlining the research questions and situating the study within an appropriate research paradigm. The pilot study was then introduced as an essential preliminary step that informed the refinement of the main study design.

The main study was described in detail, including information on participants, sample selection, and grouping procedures. A multi-method data collection strategy was adopted, comprising a quasi-experimental study, child questionnaires, and semi-structured interviews with parents and caregivers. Each data collection instrument was carefully designed and justified in alignment with the research objectives.

The chapter also described the research steps comprehensively, covering all key stages of data collection: the demographic survey, pre- and post-test tasks, the implementation of the AI-based learning app intervention, and the parent-caregivers' interviews. Measures taken to ensure the reliability of the data collection process were explained, as well as ethical considerations, such as informed consent, participant safety, confidentiality, and post-study fairness.

In conclusion, this chapter established a strong methodological foundation for the analysis of the collected data. The following chapter will present the results of the study, drawing on both quantitative and qualitative findings to address the research questions in depth.

Chapter 5. RESULTS

This chapter presents the results of both quantitative and qualitative data analyses and integrates these findings to explore the effectiveness of AI-based learning applications (Zebra AI) in Chinese children’s L2 learning. It also examines parent-caregivers’ attitudes toward their children's use of this technology for learning English. As described in Chapter 4, the quantitative data collection tools included the receptive vocabulary test (PPVT-5), cognitive ability tests (including WM task and ToM tasks), and a questionnaire assessing children’s hedonic learning experience. The qualitative data were derived from post-experiment interviews with parents. This study was structured around the following RQs.

RQ1: To what extent does the AI-based app improve children’s receptive vocabulary, and how are these outcomes predicted by their WM and ToM?

RQ2: Which functions and design interface influence children’s hedonic experience with the AI-based language learning app?

RQ3: What are parent-caregivers’ views regarding the effectiveness of using AI-based learning apps for learning a second language?

Specifically, this study addressed RQ1 and RQ2 primarily through quantitative analysis, which evaluated the impact of AI-based learning applications on Chinese children’s SLA and its association with their hedonic learning experience. Additionally, qualitative analysis is conducted mainly to address RQ3 by exploring parent-caregivers’ perceptions toward their children's use of AI-based learning apps. Some qualitative insights are also compared with quantitative findings to complement and/or interpret the results of RQ1 and RQ2, thereby providing a more comprehensive understanding of the study’s findings.

5.1 Quantitative Data Findings

The quantitative study involved children from two primary schools in the Nanyang region, with the intervention was lasted for a duration of 12 weeks (see Chapter 4 for details). The main focus was to examine the growth of children's receptive vocabulary, changes in cognitive abilities, specifically WM and ToM, and the influence of AI-based learning application features on their hedonic learning experience.

To assess receptive vocabulary and cognitive skills, evaluations were conducted at two points: a pre-test before the quasi-experiment and a post-test following the 12-week intervention. In the experimental group, students used the AI-based application Zebra AI to learn English, while the control group engaged with a standard entertainment application, Candy Crush. The pre-test established baseline vocabulary and cognitive abilities between the two groups. The post-test evaluated changes in these areas within the experimental group, specifically as a result of using the AI-based learning application and identified any statistically significant differences compared to the control group.

To gain deeper insights into which features of the AI-based learning application influence children's enjoyment and engagement during the learning process, a separate questionnaire was administered to the experimental group following the post-test.

The data from the quasi-experiment and the questionnaire were applied to address RQ1 and RQ2.

5.1.1 Quasi-experiment Data Analysis

Prior to statistical analysis, data completeness was examined. Data completeness was 100% across all outcome measures (PPVT-5, WM task, and ToM tasks), as all 85 allocated participants completed the full intervention and assessments. Consequently, no missing data imputation methods were required, and the final analysis was conducted on a complete-case basis with the full sample ($N = 85$).

Before conducting the analysis, data were tested for normality and homogeneity of variance to ensure the validity of analyses. When these assumptions were not met, non-parametric alternatives, such as the Mann–Whitney U test or Wilcoxon signed-rank test, were applied. These alternative approaches ensured that the statistical methods used were appropriate for the dataset, thereby enhancing the robustness and reliability of the analysis.

To address the research questions, a mixed measures ANOVA was employed. This analytical method allows for the simultaneous comparison of within-subject and between-subject factors while identifying interactions between these factors (Field, 2013). This method is particularly suitable for experimental designs involving repeated measurements across independent groups, as it systematically analyses time-point changes and group differences, improving statistical efficiency and minimising error variance (Singmann & Kellen, 2019).

Specifically, pre- and post-test data were analysed using mixed measures ANOVA, with time points (pre-test and post-test) treated as within-subject variables and group (experimental vs. control) treated as a between-subject variable. The analysis aimed to evaluate the intervention effects of the AI-based learning application (Zebra AI) on receptive vocabulary levels and cognitive abilities in different groups of children. Additionally, interaction effects between time and group were examined to determine whether the intervention's impact varied across groups over time.

Following the primary analysis, a sensitivity analysis was performed using one-way ANCOVA to verify the robustness of the intervention effect on receptive vocabulary. Specifically, post-test receptive single-word vocabulary breadth were compared between groups while controlling for baseline pre-test scores and age as covariates. This approach was adopted to account for potential pre-existing individual differences and to isolate the specific impact of the intervention on the primary outcome.

To investigate the potential underlying mechanisms of the intervention, specifically whether cognitive gains drove vocabulary acquisition, a multiple linear regression analysis was performed using change scores (gain scores). This approach was selected to capture dynamic developmental changes and control for pre-existing baseline individual differences, thereby allowing for a direct assessment of how improvements in cognitive domains relate to vocabulary growth. The dependent variable was defined as the gain in receptive vocabulary（Δreceptive single-word vocabulary breadth）(post-test minus pre-test scores). The predictor variables included group assignment (experimental vs. control), age, and the calculated gains in WM (ΔWM) and ToM (ΔToM).

To control for Type I error rates associated with multiple comparisons, a hierarchical approach was adopted by pre-specifying the outcome measures. Receptive single-word vocabulary breadth was designated as the primary outcome, reflecting the main objective of the intervention. Cognitive measures (WM task and ToM) were treated as secondary outcomes, serving primarily to explore underlying cognitive mechanisms. Consequently, statistical inference for the primary outcome was based on a standard α level of .05, while secondary analyses were interpreted as exploratory.

All quantitative data were analysed using SPSS version 29 to ensure precision in data handling and the reproducibility of results. Throughout the analysis, strict adherence to statistical standards was maintained, with detailed reporting of significance levels (*p*-values) and effect sizes. For mixed-design ANOVA, effect sizes were calculated using partial eta squared (ηp2) to assess the impact of both time and group effects. For non-parametric tests, effect sizes were calculated as *r* (using the formula $r = z / \sqrt{N}$) to quantify the practical significance of observed differences.

5.1.1.1 Test of normal distribution

To ensure the validity of the statistical methods used in this study, the distributional properties of the pre-test and post-test data for all outcome measures (receptive single-word vocabulary breadth, WM, and ToM) were examined separately for the experimental and control groups.

Although the Shapiro–Wilk test was initially considered, it is recognised as being highly sensitive to minor departures from normality, particularly in moderate sample sizes and in datasets with discrete scoring structures (Field, 2024). This is directly relevant to the ToM and WM measures in the present study, both of which employ integer-based scores with restricted ranges. For this reason, and in line with recommended practice for General

Linear Models, normality was assessed primarily through the inspection of skewness and kurtosis values. According to George and Mallery (2024), values between −2 and +2 are generally regarded as acceptable indicators of approximate univariate normality.

As summarised in Table 5-1, the descriptive statistics indicated normality across all measures.

Table 5-1. *Tests of normality.*

		Group	Mean (M)	*SD*	Skewness	Kurtosis
Pre-test	Receptive single-word vocabulary breadth	Experimental	21.79	6.23	.84	1.21
		Control	22.55	6.45	.15	-.90
	WM	Experimental	6.91	2.01	.37	-.60
		Control	7.33	1.63	-.64	-.02
	ToM	Experimental	3.95	1.27	.24	-.72
		Control	4.21	1.00	-.15	-.18
Post-test	Receptive single-word vocabulary breadth	Experimental	36.12	11.79	.47	-.49
		Control	28.40	9.69	.42	-.43
	WM	Experimental	8.09	1.88	-.10	-.76
		Control	7.36	1.61	.24	.49
	ToM	Experimental	4.26	1.05	.24	-1.15
		Control	4.55	.97	-.14	-.88

The distribution of receptive single-word vocabulary breadth was approximately symmetrical. In the experimental group, pre-test skewness was 0.84 and kurtosis was 1.21; in the control group, skewness was 0.15 and kurtosis –0.90. Post-test distributions similarly met normality criteria, with skewness and kurtosis values remaining within the ±0.50 range in both groups.

WM scores also demonstrated good normality despite the discrete scoring scale. In the experimental group, pre-test skewness was 0.37 and kurtosis –0.60, while the control group showed skewness of –0.64 and kurtosis of –0.02. Post-test skewness and kurtosis values also remained within the ±1.0 range in both groups.

The ToM total scores displayed acceptable distributional characteristics consistent with parametric assumptions. Pre-test skewness was minimal for both the experimental (0.24) and control (–0.15) groups, with kurtosis values (–0.72 and –0.18) respectively indicating a slightly platykurtic yet normal distribution. Post-test skewness and kurtosis values also remained within the ±2.0 range in both groups.

Taken together, these indices confirmed that the receptive single-word vocabulary breadth, WM and ToM variables were suitably normally distributed within each group, supporting the use of parametric analyses (mixed-design ANOVA) for evaluating intervention effects."

5.1.1.2 Pre-test

Pre-testing is one of the key components in a quasi-experimental design, as it helps ensure that the research's results are not influenced by bias. If significant differences exist between the groups at the outset, these differences could potentially account for any post-intervention disparities (Christensen et al., 2011). In this study, the primary purpose of the pre-test was to confirm that the vocabulary levels and cognitive abilities of children in both the experimental and control groups were comparable. Moreover, pre-testing aids researchers in mitigating threats to validity, such as bias (Harris et al., 2006). Consequently, the pre-test was employed in this study to ensure that any observed group differences could be appropriately attributed to the intervention using the AI-based learning app.

The pre-test focused on assessing children's baseline receptive vocabulary and cognitive abilities. Before testing, neither the control group nor the experimental group had been exposed to the AI-based application intervention. The test data indicated no statistically significant differences between the two groups at baseline, whether in overall vocabulary levels (as determined by a t-test) or in cognitive ability dimensions. This finding reinforces the confidence that any differences observed in the post-test can be attributed to the intervention, specifically the impact of he AI-based learning application on children's vocabulary development and cognitive skills.

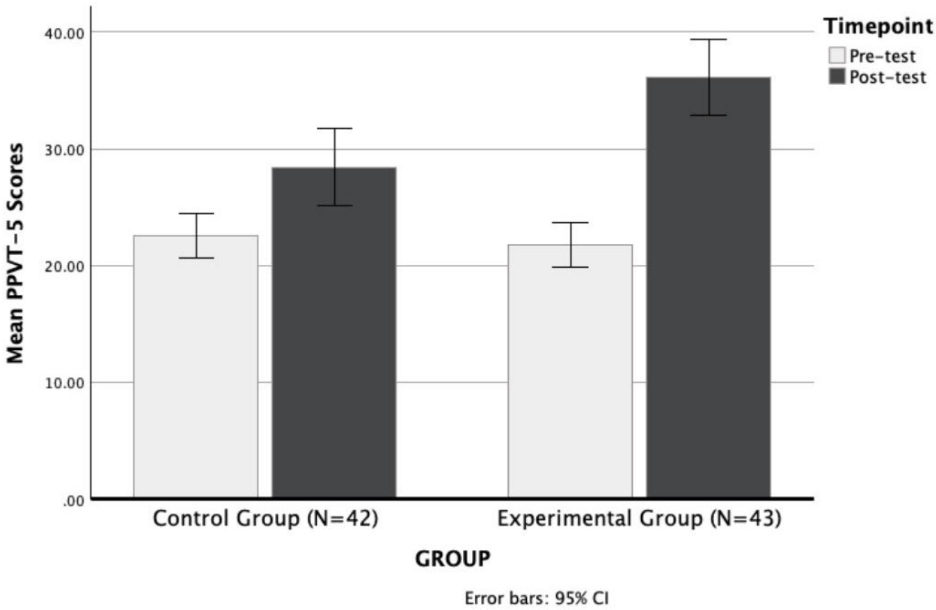
5.1.1.3 Effect of AI-based Learning Apps on Vocabulary Proficiency

A mixed measures ANOVA was conducted to examine the intervention effects. The results revealed a significant main effect of Timepoint (i.e., pre-test vs. post-test) was significant [F (1,83) = 297.75, p <.001, η^2_p = .782]; the average score generated at pre-test was M= 22.16, SD = 6.31 and M= 32.31, SD = 11.42 at post-test.

The effect of Group was non-significant [F (1,83) = 3.61, p =.061].

Furthermore, there was a significant interaction between the effects of Timepoint and Group [F (1,83) = 52.42, p <.001, η^2_p = .387], suggesting differing variability in PPVT-5 across different groups over time. This is illustrated in Figure 5-1.

Figure 5-1. *PPVT-5 - Timepoint x Group.*



At the pre-test stage, there was no significant difference between the experimental and control groups (mean difference = .76 scores, p = .584), indicating that both groups were comparable at baseline. During the intervention period, children in the control group showed an average improvement of 5.86 points on the receptive single-word vocabulary breadth (p <.001), whereas children in the experimental group demonstrated a significantly greater improvement, with an average gain of 14.33 points (p <.001). By the post-test, the experimental group outperformed the control group, with a significantly higher mean score averaging 7.71 points more (p =.001). This suggests that the AI-based learning app had a positive impact on children's receptive vocabulary development, thereby supporting Hypothesis1a.

Given that the primary analysis used PPVT-5 raw scores to capture absolute vocabulary growth, it was necessary to rule out the potential confounding influence of developmental maturation and baseline differences. To verify the robustness of the findings, a one-way ANCOVA was conducted to compare post-test PPVT scores between the experimental and control groups, with pre-test scores and age (in months) entered as covariates. The results indicated that the pre-test covariate was significantly related to the post-test scores [F (1, 81) = 508.88, p < .001, η^2_p = .86], justifying its inclusion. The effect of age was not significant (p = .181). Importantly, after adjusting for pre-test scores and age, there was a statistically significant difference in post-test PPVT scores between the groups [F (1, 81) = 113.75, p < .001], with a large effect size (η^2_p = .58).

Analysis of the estimated marginal means revealed that the experimental group achieved a significantly higher adjusted post-test score (M_{adj} = 36.78, *SE* = 0.59) compared to the control group (M_{adj} = 27.72, *SE* = 0.60). The mean difference between the groups was 9.06 (95% CI [7.37, 10.75]). These findings provide robust evidence that the AI-based language learning application (Zebra AI) significantly enhanced children's receptive vocabulary.

5.1.1.4 Effect of AI-based Learning Apps on Cognitive Development

5.1.1.4.1 Working Memory Task

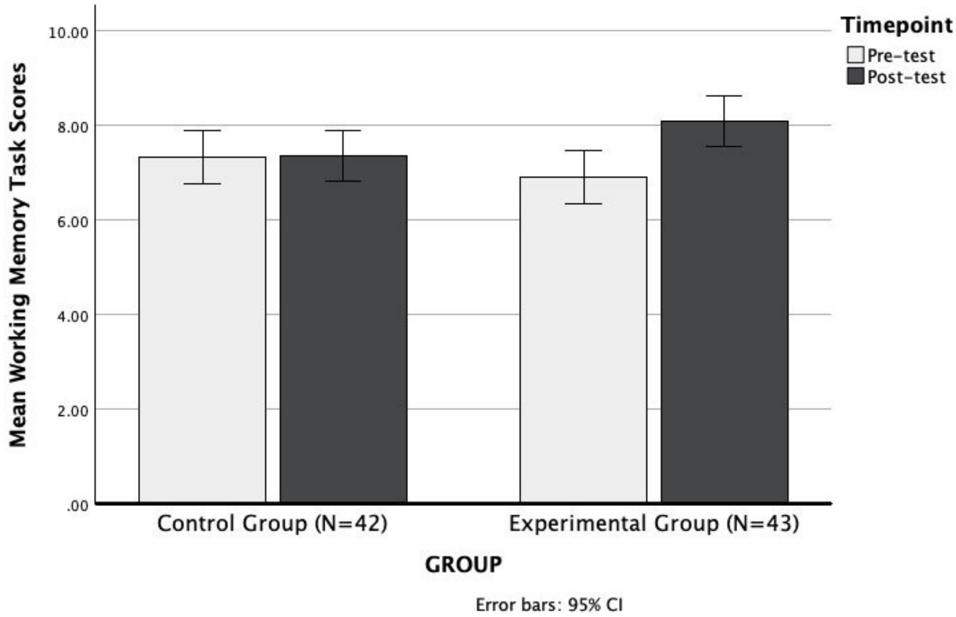
The effect of Timepoint was significant [F (1,83) = 14.54, p <.001, η^2_p = .149]; the average scores generated at pre-test was 7.12 and 7.73 at post-test.

The effect of Group was non-significant [F <1].

There was a significant interaction between the effects of Timepoint and Group [F (1, 83) = 13.42, p <.001, η^2_p = .139].

These data are illustrated in Figure 5-2.

Figure 5-2. *WM - Timepoint x Group.*



Follow-up comparisons further revealed that the experimental group showed a significant improvement in WM task scores during the intervention period, with an average increase of 1.19 points (p <.001), while the control group exhibited no significant change (p =.916).

In summary, the results of the WM task indicate that children's performance improved overall from the pre-test to the post-test, with the experimental group showing a particularly notable improvement following the intervention. These findings suggest that the AI-based learning application (Zebra AI) may have a positive impact on children's cognitive development, particularly in enhancing WM.

5.1.1.4.2 Theory of Mind Tasks

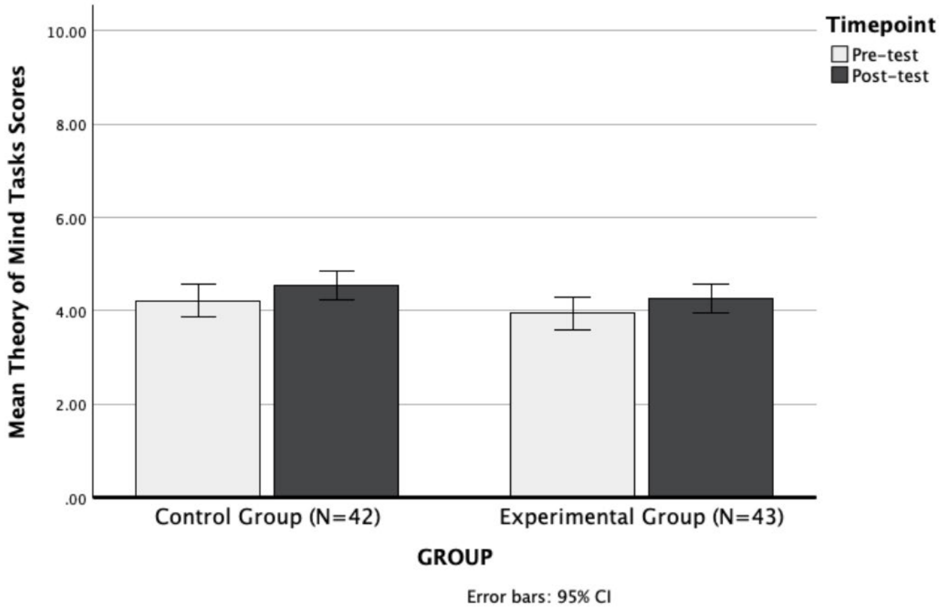
The effect of Timepoint was significant [F (1, 83) = 16.048, p <.001, η^2_p = .162]; the average scores generated at pre-test was 4.08 and 4.40 at post-test.

The effect of Group was non-significant (p =.214).

There was no evidence of an interaction between the effects of Timepoint and Group [F <1].

These data are illustrated in Figure 5.3.

Figure 5-3. *Theory of Mind - Timepoint x Group.*



The results showed that children's overall performance in the ToM tasks improved over time. However, the intervention did not have a significant impact on their ToM abilities (p = .214).

5.1.1.4.3 Receptive Single-word Vocabulary Breadth Regression Analysis

To investigate the underlying mechanisms of the intervention, specifically whether cognitive gains drove vocabulary acquisition, a change-score regression model was constructed. This model aimed to predict Δreceptive single-word vocabulary breadth from intervention group assignment, age, and ΔWM and ΔToM. It should be noted that while the variable 'Dose' (total completion rate) was initially considered for inclusion, it was excluded from the final model due to severe multicollinearity with the group variable (*VIF* > 700).

Prior to interpreting the model results, statistical assumptions regarding linearity, homoscedasticity, and independence of errors were verified. Crucially, collinearity diagnostics for the final model indicated that *VIF* values for all predictors were well within acceptable limits (ranging from 1.02 to 1.22), thereby confirming the independence of each predictor.

The overall regression model was statistically significant [F (4, 80) = 18.06, MSE = 25.890, p <.001], explaining approximately 44.8% of the total variance in receptive vocabulary growth [R =.688, R^2 =.474, R^2adj =.448]. The detailed contributions of individual predictors are summarised in Table 5-2.

Table 5-2. *Receptive Single-word Vocabulary Breadth – Regression.*

	B	*Std. Error*	*β*	*t*	*p*	*95% CI*
Constant	-18.611	6.187		-3.008	.004	(-30.923, -6.300)
Age (month)	.211	.072	.240	2.920	.005	(.067, .355)
ΔWM	.777	.392	.178	1.983	.051	(-. 003, 1.557)
ΔToM	-1.381	.785	-.147	-1.760	.082	(-2.943, .181)
Group	7.970	1.203	.586	6.627	< .001	(5.577, 10.364)

Note. B (and its Std. Error) is an unstandardised coefficient; β is a standardised coefficient. The 95% Confidence Interval is the population estimate of the β coefficient (i.e., extrapolating the estimate from the sample to the 'population' of individuals who reflect the sample characteristics). Significant contributors are highlighted in bold.

The results of the change-score regression analysis revealed that the group variable was the most robust predictor of receptive vocabulary growth (β = .586, p < .001), indicating that the experimental group maintained a significant advantage even after controlling for other variables. Regarding cognitive mechanisms, the growth in WM demonstrated a marginally significant positive predictive effect (β = .178, p = .051). Although this *p*-

value falls slightly above the .05 threshold, the positive direction of the coefficient (B = .777) suggests a potential association between WM gains and receptive vocabulary growth. In contrast, gains in ToM did not significantly predict receptive vocabulary growth (β = -.147, p = .082). Hypothesis H1b, which predicted that both WM and ToM would be significant positive predictors of vocabulary learning outcomes, was only partially supported. Furthermore, age was confirmed as a significant positive predictor (β = .240, p = .005).

5.1.1.5 Influence of Functions and Interface Design on Children's Hedonic Experience

The second research question aimed to explore children's hedonistic experiences with apps and which app features and designs influence their using experience. In this data analysis, to quantify the respondents' feedback, I converted the smiley emoticons into numbers and used a five-point scale for measurement. The specific conversion rules are as follows: 2 = 😀, 1 = 🙂, 0 = 😐, -1 = 🙁, -2 = ☹️. This approach enabled a more effective analysis and interpretation of the participants' answers. As illustrated in Figure 5-4, descriptive statistics were generated to examine the data related to the questionnaire. This included calculating the percentages and frequencies of responses to the nine questions across the experimental group.

Figure 5-4. *Participants hedonic experience with the app.*

From the perspective of overall emotional experience (Q2, Q3), 88% of participants (N = 38) reported feeling very happy while using Zebra AI, and 86% (N = 37) found its user experience to be very exciting. Only 5% (N = 2) of participants expressed a neutral attitude, and no negative feedback was recorded. These data indicate that most children demonstrated positive emotional experiences when using Zebra AI.

Firstly, from the perspective of usability (Q1), 86% of participants (N = 37) believed Zebra AI was very easy to use, with only 5% (N = 2) providing neutral feedback and no one considering it difficult to use. This result suggests that the app's user interface and operational design align well with children's usage habits and have gained their approval.

Regarding visual experience (Q7), 93% of participants (N = 40) provided positive feedback on the image and video quality of the application, indicating that most children enjoyed its visual content.

For content comprehensibility (Q8), 81% of participants found the learning material easy to understand. However, the proportion of neutral and negative feedback was slightly higher (12%) compared to that for visual elements, suggesting that there is room for improvement in aligning content presentation with children’s language comprehension levels.

Among all features, the fun aspect of the games (Q6) received the highest positive rating at 90%, with no negative feedback reported.

Regarding usage intention (Q4, Q5), 84% of participants (*N* = 36) expressed a willingness to use the application again, and 79% (*N* = 34) indicated a desire for long- term use. This data reflects that most children show a high willingness to use the application and maintain a positive attitude toward its continued use.

Although the overall feedback was positive, the ASR-based pronunciation feedback (Q9) obtained a relatively weaker emotional response from children. While 72% of participants provided positive feedback, neutral and negative responses accounted for 24% which was highest proportion among all functions. This suggests that this feature may not fully meet children's emotional expectations in terms of ease of operation and prompt feedback, thereby impacting the overall pleasantness of the experience.

Based on the questionnaire data, Zebra AI demonstrates strength in enhancing children's enjoyable experiences. First, the application's ease of use, visual appeal, and game features play a crucial role in the children's experience, especially the game and video call simulation features, which are particularly popular among children. These features not only enhance emotional engagement but also boost their interest and participation in learning. However, there is still room for optimisation and improvement in certain functions, such as ASR-based pronunciation feedback. Overall, the Zebra AI application, with its user-friendly interface design, rich interactive elements, and high- quality visual effects, provides an enjoyable and highly engaging learning experience, making it an ideal language learning tool for children.

At the end of this survey, I asked about children's favourite features in the application. The results, as shown in Table 5-3, indicate that the most popular feature among children is games, accounting for 28% (*N* = 12). Following closely behind is simulating a video call with a foreigner, which accounts for 23% (*N* = 10). Animations and specific cartoon characters represent 16% (*N* = 7) and 12% (*N* = 5), respectively. Although these two features are less popular compared to games and video calls, they still demonstrate a certain level of appeal. English songs also received 12% (*N* = 5) of user preference. It is noteworthy that 9% (*N* = 4) of users did not specify their favourite feature. This may reflect these users' general satisfaction with various aspects of the application or a lack of specific focus on any feature.

Table 5-3. *Ranking of children's favourite app features.*

Q10. What do you like most about the app?	%	*n*
It was a game	28%	12
Simulate a video call with a foreigner	23%	10
Animations	16%	7
Cartoon character	12%	5
English songs	12%	5
Unknown / Not specified	9%	4

5.1.1.6 Intervention Implementation and Participant Experience

Following the report on the primary effects of the intervention on receptive single-word vocabulary breadth, WM, and ToM (see Section 5.1.1.4), this section turns to an evaluation of the intervention implementation and children's hedonic experience. This analysis aims to provide the behavioural context and explanatory mechanisms for the observed learning outcomes. To fully contextualise the findings, this section first quantifies the actual dosage received by the experimental group via parent logs (Section 5.1.1.6.1) and subsequently explores how children's hedonic experience served as a potential driver of their behavioural adherence and resulting learning gains (Section 5.1.1.6.2).

5.1.1.6.1 Adherence and Dosage Metrics

To ensure transparency regarding the implementation of the intervention and to quantify the actual dosage received by the experimental group ($N = 43$), a descriptive analysis was conducted on data derived from the *Zebra AI App Usage 12-Week Parent Log* (as detailed in the Methods section). This analysis aimed to quantify participant adherence and the dosage of usage throughout the 12-week intervention period.

The quantitative analysis revealed a high level of overall adherence within the experimental group, with all metrics approaching or meeting the pre-defined targets (target dosage: 5 units per week). The summary statistics for the key adherence indicators are presented in Table 5-4.

Table 5-4. *Descriptive Statistics of Intervention Usage (N = 43).*

	Mean (*M*)	*SD*	Min	Max	Target
Weekly Frequency (Units/week)	4.91	0.13	4.50	5.00	5.00
Adherence Rate (%)	93.80	7.74	75.00	100.00	100%
Total Completion Rate (%)	98.10	2.69	90.00	100.00	100%

The results demonstrate that the intervention was executed with high efficiency and consistency within the experimental group. The Total Task Completion Rate of 98.10% confirms that the intervention was supported by sufficient dosage. Regarding Mean Weekly Days, participants completed an average of 4.91 days per week ($SD = 0.13$), a figure that aligns closely with the ideal target dosage of 5 days per week established for this study. Furthermore, 93.80% of the intervention weeks achieved full adherence ($SD =$

7.74). This indicates that, on average, participants met the strict target dosage (5 days/week) for approximately 11.3 out of the 12 weeks. Notably, even the lowest individual adherence rate recorded was 75.00%, highlighting the exceptional stability of usage frequency across the entire experimental group.

As the parent log did not directly record duration in minutes, a single completed task was defined as the minimum unit of usage. Overall, these indicators, particularly the high completion rates and average weekly usage, confirm that the intervention achieved and sustained the intended dosage and frequency. These robust adherence data provide a solid foundation and credibility for attributing the subsequent learning outcomes to the Zebra AI intervention.

5.1.1.6.2 Relationship between Children's Hedonic Experience and Engagement

Having established the descriptive profiles of intervention adherence, the study proceeded to examine the underlying mechanisms driving these patterns. Specifically, this section investigates the interplay between children's hedonic experience, their objective behavioural engagement, and their learning outcomes. The primary aim of this analysis was to cross-validate children's self-reported feedback against objective metrics, whilst determining whether positive emotional experiences effectively translated into sustained usage and receptive vocabulary gains. To this end, Pearson product-moment correlation analyses were conducted for the experimental group (*N* = 43). To ensure the robustness of the estimates, bootstrapped 95% confidence intervals (CIs) were calculated.

Table 5-5. *Intercorrelations between Hedonic Experience, Adherence, and Receptive Vocabulary Gains (N = 43).*

	Total Task Completion Rate	ΔReceptive single-word vocabulary breadth	Hedonic Experience Index

Total Task Completion Rate	1	$r = .599^{**}$ $p < .001$	$r = .753^{**}$ $p < .001$
ΔReceptive single-word vocabulary breadth	--	1	$r = .338^{*}$ $p = .027$
Hedonic Experience Index	--	--	1

Note: *Statistical significance is indicated by asterisks: * $p < .05$, ** $p < .01$. All reported p-values are based on two-tailed tests.*

First, the analysis assessed the link between children's self-reported hedonic experience and their actual usage of the application. The results revealed a significant positive correlation between the composite Hedonic Experience Index and the objective Total Task Completion Rate ($r = .753$, $p < .001$, 95% CI [.59, .86]). The coefficient of determination ($r^2 = .567$) indicates that children’s hedonic experience explained approximately 56.7% of the variance in their behavioural adherence. This substantial overlap suggests that the participants' intrinsic enjoyment was a primary driver of their sustained engagement and high task completion rates throughout the 12-week intervention.

Furthermore, the relationship between subjective experience and cognitive learning outcomes was examined. A moderate but statistically significant positive correlation was observed between the Hedonic Experience Index and standardised Δreceptive single-word vocabulary breadth ($r = .338$, $p = .027$, 95% CI [.04, .58]). While the strength of this association is lower than that of behavioural adherence ($r^2 = .114$), it nonetheless indicates that hedonic experiences during learning are significantly associated with superior vocabulary acquisition. This finding provides empirical support for the link between affective quality and cognitive performance within the context of AI-supported learning.

5.1.2 Summary of Quantitative Findings

This section summarises the quantitative findings related to RQ1 and RQ2. Based on the changes observed between the pre-test and post-test scores, the results reveal the complexity and diversity of learning patterns addressed by RQ1. In the experimental group (children who used the AI-based language learning application), a significant improvement in receptive single-word vocabulary breadth was found following the intervention. Consequently, Hypothesis H1a is fully supported.

Regarding cognitive development, the results revealed a nuanced pattern. The intervention appeared to specifically bolster WM as evidenced by significant gains to the experimental group. In contrast, improvements in ToM were observed across both groups, suggesting these changes were attributable to natural maturation rather than the specific effects of the AI application.

Hypothesis H1b was only partially supported by the change-score regression analysis. While the experimental group maintained a robust advantage and age proved to be a significant positive predictor, the cognitive mechanisms showed mixed results: WM growth displayed a marginally significant positive trend, but ToM gains did not significantly predict receptive vocabulary growth.

Regarding RQ2, the survey results indicated that children showed a high level of satisfaction with Zebra AI, thereby confirming Hypothesis H2. Features such as ease of use, high-quality videos and animations, and the app’s interactive and entertaining design all contributed to enhancing children’s hedonic experience to varying degrees.

Additionally, the intervention demonstrated exceptional adherence and consistency, confirming that the intended dosage was successfully delivered to the experimental group. Analyses revealed that children's hedonic experience significantly predicted both their behavioural engagement and receptive vocabulary acquisition. These findings underscore the pivotal role of hedonic experience in driving sustained usage and fostering cognitive development within AI-supported learning environments.

5.2 Qualitative Data Findings

This section analyses the qualitative data, which were collected from interviews with parents from both urban and rural primary schools, comprising ten questions. The interview data were coded through thematic analysis, identifying key themes to address RQ3.

RQ 3: What are the parent-caregivers' views regarding the effectiveness of using AI-based learning apps for learning a second language?

5.2.1 Qualitative data analysis

This section discusses the methods for qualitative data analysis and describes the steps taken to ensure the reliability and validity of the process. In this research, Thematic Analysis (TA) was chosen as the primary approach for analysing interview data. TA is a widely recognised method in qualitative research, which aims to identify, analyse, and report themes within data (Clarke, 2016). The process involves identifying themes or patterns, segmenting and categorising the data, then summarising and reconstructing it to reveal underlying patterns and core concepts. This approach enables a deeper understanding of implicit meanings within the data (Given, 2008).

TA's versatility and widely applicability make it particularly effective for examining complex qualitative phenomena, such as participants' lived experiences, perspectives, and behavioural patterns (Clarke, 2016). Given its structured yet flexible framework, TA is a practical and accessible choice for researchers new to qualitative and interview data analysis (Knott et al., 2022).

In TA, researchers often develop themes using both deductive and inductive approaches. The deductive approach is guided by existing theories or literature, relying on a pre-defined framework to direct coding and theme identification. In contrast, the inductive approach allows themes to emerge directly from the data, providing greater openness and flexibility to capture participants' authentic experiences and perspectives (Azungah, 2018). Although deductive and inductive methods have distinct theoretical foundations and unique strengths, Proudfoot (2023) highlights that, in practice, researchers cannot fully exclude prior knowledge or subjective influences. Consequently, achieving a 'purely inductive' or 'purely deductive' approach presents notable methodological challenges.

In conducting the thematic analysis for this study, I integrated both deductive and inductive approaches to enhance the comprehensiveness and depth of the analysis. Initially, I employed a deductive approach, constructing a preliminary analytical framework informed by existing theories and research questions. This framework, developed through interview questions, facilitated the generation of initial codes in the early stages of data processing, ensuring alignment with the study's objectives.

As the analysis progressed, however, I increasingly adopted an inductive approach. Through iterative data analysis, new themes and subtle distinctions became apparent, extending beyond the original analytical framework. Consequently, the inductive approach became central to the latter stages of analysis, allowing for the exploration of organically emerging themes and a deeper capture of participants' perspectives and experiences.

To ensure the validity and rigor of the analysis, the thematic analysis process in this study generally followed the six-step guide proposed by Braun and Clarke (2012): "1) Familiarising yourself with the data; 2) Generating initial codes; 3) Searching for themes; 4) Reviewing themes; 5) Defining and naming themes; 6) Producing the report" (p87).

According to Braun and Clarke (2006), the first stage of Thematic Analysis (TA) involves familiarising oneself with the interview data. As all participants were Chinese, the interviews were conducted in Mandarin. To preserve the original meaning and subtle contextual nuances, I initially analysed the data in Mandarin, translating only key terms and sentences to aid in interpreting the findings. To further ensure translation reliability, I engaged a bilingual colleague proficient in both Chinese and English to review the translations, thus minimising potential discrepancies or loss of meaning. Additionally, I included explanatory notes for expressions with specific contextual meanings in

Mandarin that lack direct English equivalents, aiming to help readers understand how these cultural and linguistic differences may affect interpretation.

At the initial stage of transcription, I transcribed the interviews verbatim, applying a 'denaturalised transcription' approach (Halcomb & Davidson, 2006). Denaturalised transcription, or 'full verbatim', captures all utterances, errors, repetitions, and grammatical flaws (Bucholtz, 2000). This approach prioritises the substantive content of the interviews, such as the meanings and insights generated during conversations, and places less emphasis on details like accent or involuntary sounds.

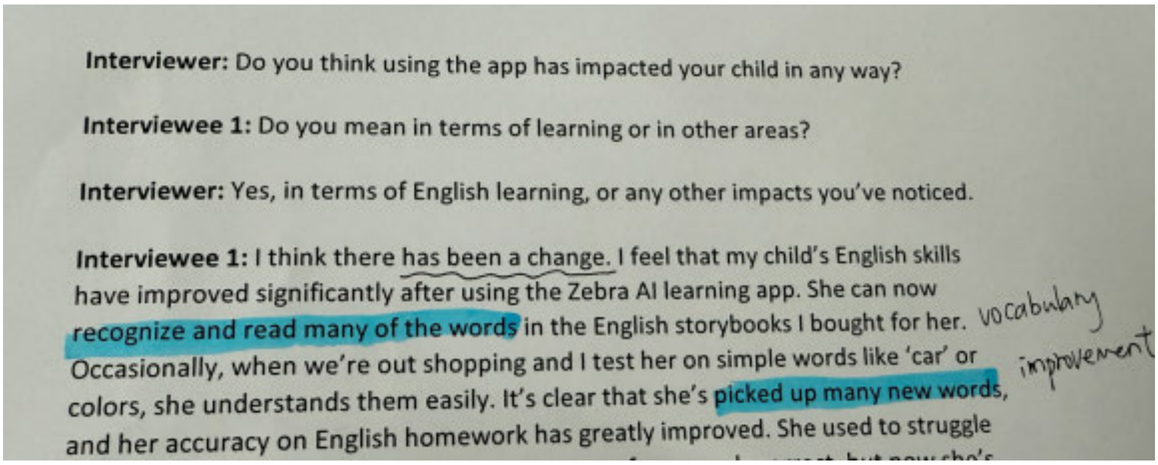
To ensure participant privacy and data confidentiality, I anonymised the interview data. Each participant received an anonymous identifier (e.g., 'Interviewee 1' to 'Interviewee 8'), replacing real names and thereby protecting personal information. This anonymisation reduces potential biases and minimises impacts that may arise from identity disclosure (Knott et al., 2022).

Following transcription and anonymisation, I transitioned to the coding phase for in-depth data analysis. Coding serves as a bridge between qualitative data collection and analysis, allowing researchers to systematically categorise and label raw data, thus setting the stage for further interpretation (Saldaña, 2021). Initial coding was guided by the study's research questions, with a specific focus on parent-caregivers' perspectives on their children's use of AI-based learning apps for L2 learning.

Given the small sample size of this study, comprising 8 participants, manual coding was selected over computer-assisted coding tools at the initial analysis stage. Saldaña (2011, 2021) suggests that in small-scale studies, manual coding can deepen the researcher's understanding of the data and foster reflective engagement, wherein the researcher actively interprets and connects with the data's nuances. Consequently, manual coding was employed as the initial analytical strategy to facilitate a more detailed and intuitive analysis. This approach is consistent with the recommendations (Miles et al., 2014), who advocate manual coding in small-scale qualitative research for its capacity to yield richer insights. While computer-assisted tools can improve coding efficiency, their use in this study's context may introduce additional procedural complexities that could detract from the analytical process.

Initial coding was guided by the content of the 10 interview questions, summarising and categorising key themes systematically. Responses were reviewed and analysed to identify and extract essential information. To enhance data management, I printed all interview transcripts and conducted manual coding on paper. Following Saldaña's (2021) recommendations, I used colour coding to highlight keywords and key segments, facilitating a visual organisation of content. This method enhanced clarity and intuitiveness while supporting the subsequent categorisation of themes and identification of patterns. For instance, recurring mentions of 'vocabulary improvement' after AI-based learning application use were marked in blue, with detailed notes beside each segment to facilitate later tracing of sources and contexts, as illustrated in Figure 5-5.

Figure 5-5. *Labelling and Identification of Transcription Segments*



Interviewer: Do you think using the app has impacted your child in any way?

Interviewee 1: Do you mean in terms of learning or in other areas?

Interviewer: Yes, in terms of English learning, or any other impacts you've noticed.

Interviewee 1: I think there has been a change. I feel that my child's English skills have improved significantly after using the Zebra AI learning app. She can now recognize and read many of the words in the English storybooks I bought for her. Occasionally, when we're out shopping and I test her on simple words like 'car' or colors, she understands them easily. It's clear that she's picked up many new words, and her accuracy on English homework has greatly improved. She used to struggle

vocabulary improvement

Following a comprehensive review of the colour-coded data, all descriptions pertaining to vocabulary improvement were consolidated to enhance thematic precision. I combined repetitive expressions and refined similar information to capture both explicit mentions of vocabulary growth and detailed feedback, such as parent-caregivers' observations that 'many new words and phrases were learned', as well as improvements in recent English test scores. This layered approach enabled a detailed and detailed coding framework. The remaining interview transcripts were analysed using a comparable coding process, systematically identifying relevant text units and organising them into meaningful themes. Through iterative categorisation and refinement, a total of 52 initial codes emerged upon reaching data saturation, indicating that no new information or themes could be identified. These codes encompass a range of insights into parent-caregivers' feedback on AI-based learning apps, particularly regarding children's performance, progress, and challenges in English learning. Table 5-6 presents an overview of all initial codes along with their associated content, providing a structured foundation for further analysis.

Table 5-6. *The 52 Initial Codes from the Interview Transcripts*

1. Vocabulary improvement
2. Improved classroom performance (children's performance in class received praise from teachers)
3. Pronunciation praised by teachers
4. Specific pronunciation issues improvement
5. Limitations of AI apps in vocabulary acquisition (achieving complete mastery of a language through the app alone can be challenging)
6. Improvement in pronunciation
7. Pronunciation improvement in repetition practice
8. Increased willingness to raise hands and speak in class
9. Improved English test scores
10. Learning new words and phrases
11. Specific pronunciation issues improvement
12. Increased interest in learning
13. Enhanced motivation for learning
14. Attention captured by multimedia features (animations, games, and interactive features increase children's learning interest)
15. Gamified learning experience (learning English feels like playing a game, making children more willing to participate)
16. Appeal of animations and cartoon characters
17. Age-appropriate game design (games in the app are suitable for different age groups)
18. Real-time feedback system (the app provides immediate feedback on pronunciation, helping children correct their pronunciation errors more quickly)
19. Increased confidence in class

20. Improved participation in English class
21. Overcoming fear of speaking English
22. Changed study habits
23. Increased patience in learning
24. Enhanced self-directed learning (children can independently focus on learning for longer periods)
25. Reduced reluctance to studying
26. Development of independent learning habits (children no longer need reminders and can study on their own)
27. Convenience of learning anytime, anywhere
28. Rich learning resources (AI tools provide a lot of learning materials)
29. Savings on learning expenses
30. Reduced teaching pressure on parents (the app can help parents check their child's pronunciation)
31. Young children lack self-discipline (young children need constant supervision from parents)
32. Low Learning Efficiency Due to Distractions
33. Difficult content (especially for children with weaker foundations)
34. Loss of interest when facing challenge vocabulary
35. Operational difficulties at first use (children need parental guidance when first using the app)
36. Negative effects of animated characters (characters in opposition or intense scenes in animations may negatively influence children)
37. Parental concerns about eye health
38. High-performance device requirement
39. Impact of unstable network connection (especially in rural area)
40. Parental concerns about costs (parents reconsider using the app if it is too expensive)
41. Lack of familiarity with technology among rural parents (some parents are unfamiliar with new technology, facing difficulties in using the app)
42. Limited supervision time for parents (parents are busy sometimes and cannot supervise children for extended periods)

43. Importance of school and teacher recommendations (parents' willingness to use AI learning apps is influenced by school recommendations and teacher experience)
44. Expectation for personalised learning features (hope the app can automatically adjust difficulty based on the child's level)
45. Need for dialect support
46. Offline mode requirement
47. Need for a parental user guide (parents hope the app provides a simple user guide to better support children's learning)
48. Peer-support
49. Enhanced feedback feature (parents hope the app provides insights on children's learning progress to better support their development)
50. Potential of AI to replace real teachers
51. Parent-App communication and progress updates
52. Integration of AI Apps with Classroom Learning

Following the initial identification of codes, I conducted a systematic review to consolidate related codes, thereby generating basic themes that offer a clearer and more organised representation of parental perspectives. After this consolidation process, I simplified the coding framework, reducing the initial 52 codes to 25 more representative basic themes. This integrative approach minimised repetition and facilitated the development of a coherent thematic structure.

For example, multiple parent-caregivers' references to improvements in their children's pronunciation, such as "teacher praise for pronunciation", "specific pronunciation issues improvement," "improvement in pronunciation", and "pronunciation improvement in repetition practice" all pointed to a broader theme of pronunciation enhancement. Consequently, these codes were synthesised under thetheme 'Pronunciation Accuracy Improvement' to comprehensively capture parent-caregivers' feedback on pronunciation development.

In addition, I consolidated the codes based on their function or purpose, grouping those with similar or overlapping functional orientations into broader themes. For instance, the codes ‘Improved participation in English class’ and ‘Overcoming the fear of speaking English’ both reflect an improvement in children's confidence during classroom learning. The former emphasises an increased willingness to raise their hands, actively participate in class activities, and demonstrate greater learning initiative, while the latter focuses on children gradually overcoming psychological barriers in language expression, allowing them to speak English more naturally without fear of mistakes or criticism. Although these two codes differ in detail, they both indicate an enhancement of children's confidence. Therefore, I combined them under the more representative theme ‘Improved Learning Confidence’.

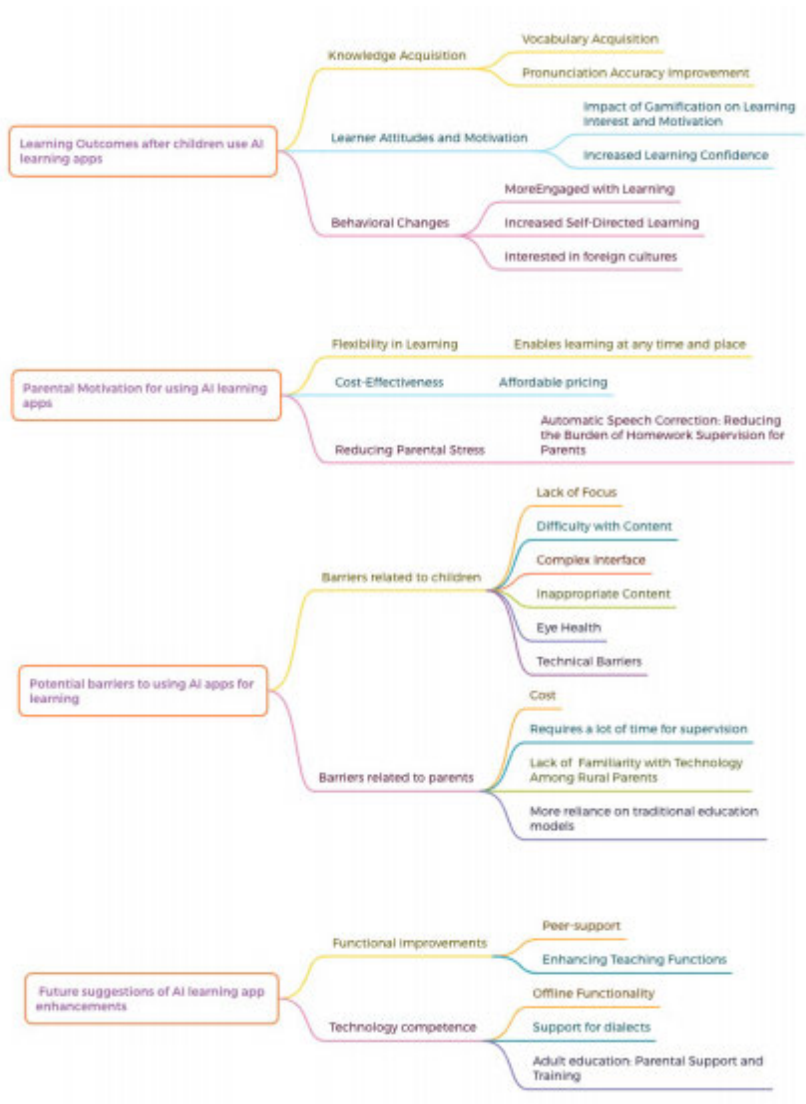
Finally, I ensured the retention of unique and nuanced perspectives that resisted categorisation, such as ‘a lack of familiarity with technology among rural parents’ and the diverse expectations for AI features, including ‘peer support’. This approach enriched the analysis and ensured that the thematic framework reflected a broad spectrum of parent-caregivers’ needs and suggestions.

In the subsequent stage of thematic integration, I systematically refined the 25 basic themes into a more comprehensive and representative set of 10 organizing themes. For example, feedback concerning learning outcomes and knowledge acquisition was synthesised into the theme “Knowledge Acquisition,” while codes related to children’s learning attitudes and motivation were grouped under the theme “Learner Attitudes and Motivation.” This iterative process elevated the codes to a thematic level, revealing prominent patterns within the data and providing a structured framework for further analysis.

Through detailed analysis and systematic refinement, I consolidated the 10 organising themes into four global themes, providing a higher-level perspective on parents' feedback regarding their children’s use of AI-based learning applications for learning English. These four global themes are: ‘Learning Outcomes After Using the Application’, ‘Parental Motivation’, ‘Potential Barriers’, ‘Suggestions for Future Improvements’, which together encompass the main areas of concern expressed in the parent-caregivers’ feedback.

In addition, I employed Thematic Network Analysis (Attride-Stirling, 2001) to systematically reveal the connections and hierarchical structure among the themes within the data. Thematic Network Analysis presents the basic, organising, and global themes in a network-like visualisation, which not only facilitates an understanding of the logical relationships between themes but also enhances the transparency and interpretability of the research findings (Braun & Clarke, 2021). To further clarify the logical relationships and hierarchical structure among the themes, I employed mind maps as a visualisation tool (see Figure 5-6). This approach clearly illustrates the intrinsic connections among themes at various levels, helping researchers and readers better understand the layered structure and interaction patterns within the data.

Figure 5-6. *Thematic Network of Parental Feedback on AI Learning App Usage*



5.2.2 Reflection on my own position in the research process

In conducting this analysis, I became increasingly aware that my cultural background, educational experiences, and familiarity with AI technology might shape my interpretation of the data. As a Chinese international student, my academic training and perspective may have influenced my understanding of feedback from both rural and urban parents. This observation is consistent with the claim made by Palaganas et al. (2017) that a researcher's reflexivity plays a critical role in shaping their interpretation of complex issues, particularly in data analysis. For instance, as a supporter of AI technology, I may be predisposed to highlight positive feedback on AI applications, potentially overlooking concerns or difficulties expressed by parent-caregivers. Furthermore, my educational and technological outlook may differ significantly from that of parent-caregivers in rural areas, indirectly affecting my perception of their specific needs and challenges. For example, while I might prioritise the advanced features of AI technology, rural parent-caregivers may place greater value on essential functions due to unstable internet connections or device limitations. Additionally, the cultural background and communication style of rural families differ from my urban environment, which may lead to biases when interpreting their responses during interviews.

To mitigate the influence of personal perspectives, several methodological strategies were implemented to enhance analytical objectivity. First, systematic reflection and documentation were maintained throughout each stage of analysis, especially during coding, to balance diverse viewpoints in the data. Second, to increase methodological transparency, a third-party review was conducted by my supervisors, who examined the coding framework and critiqued the thematic categorisation to ensure a balanced representation of parent perspectives. Specifically, in integrating themes, equal weight was given to feedback that diverged from my expectations and to the viewpoints of rural parents. For example, during the coding process, rural parent-caregivers' perspectives were retained as separate codes to ensure their feedback was fully and independently represented. Additionally, parent-caregivers' concerns regarding AI functionalities were treated as distinct themes rather than secondary opinions. These reflexive practices enhance the transparency of the analytical process, allowing the final analysis to acknowledge personal limitations while accurately reflecting participants' diverse experiences.

5.2.3 Semi-structured interview findings

To address the research questions, this research employed thematic analysis to analyse the interview data from parent-caregivers (see section 5.2.1) to gain a comprehensive understanding of the data by identifying the key themes related to children's use of an AI-based app to learn a second language. The development of the themes and their supportive quotes is presented in Table 5-7.

Table 5-7. *Summary of Themes, Code, Sub-codes and supporting quotes.*

Themes	Code	Sub-code	Supporting Quotes
a. Learning Outcomes after children use AI-based learning applications	a.1 Knowledge Acquisition	*a.1.1 Vocabulary Acquisition*	(Interviewee 1) *I feel like my kid's English has gotten so much better since she started using the Zebra AI app. Like, now she can actually recognise and read a lot of the words in those English storybooks I boughtfor her. Sometimes, when we're out shopping, I'll randomly ask her stuff like "what's this?" or test her on simple words like "car" or colors, and shejust gets them right away. It's pretty clear she's learned a ton of new words, and her homework is so much better now too. I mean, before, she used to really struggle with reading—she'd maybe getfive or six words out of ten right. But now it's like seven or eight, and even her teacher said she's doing great!* (Interviewee 3) *Ifind it helpful. My child has learned many new words andphrases while using this AI app, and his score has improved in the most recent English test.*

			(Interviewee 7) *Since my kid started using this English learning app, I've really noticed his vocabulary has gotten much better Like, before, he could only say some basic stuffs, you know, like "cat" or "apple." But now, he knows more words, and he can even put together simple sentences all by himself. It's honestly such a big change!* (Interviewee 4) *He has recognised more words.* (Interviewee 2) *To be honest, I think the support these apps provide is limited, and the results are just so-so. I mean, they're basically all about teaching kids vocabulary through memorisation and repetition. But, you know, young kids isn't fully developed yet, so they end up forgetting most of what they learn. Like, if you teach them 10 words, they might only remember 2 or 3 well, with more practice, they might improve a bit, but I don't think you can really rely on an app like this to fully help a child master a second language.* (Interviewee 6) *I am not very familiar with these.*
		a.1.2 Pronunciation Accuracy Improvement	(Interviewee 5) *My child's pronunciation has improved a lot. They used to have to repeat words several times to get them right during follow-along sessions, but now, I've noticed they can get it right much quicker.* (Interviewee 1) *My child came home from school excitedly telling me that the teacher praised his pronunciation.* (Interviewee 3) *My child used to struggle with pronouncing certain words correctly, especially sounds like "th" and "r." He'd always get them mixed up. But lately, I noticed that his pronunciation is so much clearer when he reads aloud, and he barely makes those mistakes anymore.* (Interviewee 4) …and his *pronunciation of words has also improved.*

	a.2 Learner Attitudes and Motivation	*a.2.1 Impact of Gamification on Learning Interest and Motivation*	(Interviewee 1) *Honestly, I think this app really gets my kid interested in learning English. She enjoys following along, reading, and even singing with it. It's such a big change from before when she thought memorizing vocabulary was boring. Now she's way more into it and even wants to learn on her own.* (Interviewee 5) *Our kid's only 5 and, honestly, he just loves playing. Sitting still with a boring book? That's not gonna happen. But these learning apps, they totally grab his attention with all the animations, games, and interactive stuff. It's like learning English turns into a game for him, and he gets so into it—it's actually fun for him!* (Interviewee 7) *Kids love cartoons, so adding that into the learning just makes it way more fun and keeps them interested.* (Interviewee 4) *I totally agree—this kind of app is great for getting kids interested in learning. The games are engaging, and the simple gameplay is perfect for their age. Plus, the cute cartoon characters really grab their attention. Even though they're learning a language, it feels more fun, and it helps them enjoy the whole process.* (Interviewee 2) *If I had to pick the best part, I'd say it's the games. They really help kids practice the words they've learned while playing, so it makes learning way more fun and effective. These modern apps are so much better at keeping kids engaged—they're learning and having fun at the same time, and it really cuts down on their resistance to learning English.*
		a.2.2 Increased Learning Confidence	(Interviewee 8) *Her teacher told me she's way more confident in class now. Before, she never raised her hand to answer questions or speak English. When I asked her why, she'd say she didn't know the answers and was scared of making mistakes or feeling embarrassed. But since she started using this AI app, she's been practicing listening and speaking at home and seems so much braver. The teacher even said she raises her hand more and is way more active in English class. Her pronunciation's been praised too, which she never thought could happen before. It feels like the app really helped her get over her fear of speaking English.* (Interviewee 5) *After using this application, I noticed a improvement in my child's confidence.*

	a.3 Behavioural Changes	*a.3.1 More Engaged with Learning*	(Interviewee 1) *I've noticed a change in her habits. Like, she used to really struggle to sit still. If I asked her to read a book on her own, she'd get distracted and startfidgeting injust afew minutes. But ever since she started using this app, I've seen her focusfor much longer.* (Interviewee 3) *I've noticed a big change in his patiencefor learning. He used to losefocusjust afew minutes into reading or studying and would getfrustrated really quickly. But since he started using this app, his attention span has gotten way better, and he can actually stayfocused on his lessonsfor much longer without getting distracted.*
		a.3.2 Increased Self-Directed Learning	(Interviewee 7) *He's actually started to enjoy English now. Like I said before, he used to hate reading vocabulary, but learning on the phone has made him so much less resistant. Now, when he gets home, he'll even take the initiative to study English and read the vocabulary out loud—he never used to do that before.* (Interviewee 5) *He's way more interested in learning English now. Before, I had to keep reminding him to study, but these days, hejust grabs his phone or tablet and starts learning on his own without me having to say anything.*
		a.3.3 Culture Curiosity	(Interviewee 4) *This app has actually taught my kid a lot about thefestivals and customs in English-speaking countries, like where Thanksgiving and Halloween comefrom and their traditions. She can even explain it all in detail! What's really cool is she's notjust sitting back and taking it in—she's also asking me and her mom all these questions, which shows she's curious about the culture. It's like she's notjust learning the language, but also seeing the world differently. This increased cultural awareness is something I never expected.* (Interviewee 3) *She really enjoys learning now, loves chatting withforeigners, and has even started talking about studying*

			abroad someday.
b. Parental Motivation for using AI-based learning applications	b.1 Flexibility in Learning	*b.1.1 Enables learning at any time and place*	(Interviewee 7) *Using this AI tool, my child can now learn English right at home just by using a smartphone, which is much more convenient. They can study whenever they have some free time, and there are so many learning resources easily accessible.* *(Interviewee 8) Honestly, we don't need to send our kid to English tutoring classes anymore. Now they can learn English anytime, anywhere, right at home.* *(Interviewee 2) It's very convenient ... they can learn at home anytime.*
	b.2 Cost-Effectiveness	*b.2.1 Affordable pricing*	*(Interviewee 7) Before, parents used to send their kids to all kinds of extracurricular classes to learn English, but now we don't have to rush around to different training centres anymore. It really saves a lot of time, effort, and money... plus, it's way more affordable than those traditional tutoring classes.* *(Interviewee 5) Most of these apps are free, so they don't add any extra financial stress on families when it comes to their kids' education.*
	b.3 Reducing Parental Stress	*b.3.1 ASR-based pronunciation feedback: Reducing the stress of Homework Supervision for Parents*	*(Interviewee 1) I think this feature is really practical because, as a parent, what I care about most is whether my kid's pronunciation is correct. At her age, just memorizing words is hard enough. This feature is amazing—it saves us so much time and effort. My husband and I don't have to look up every single word to check her pronunciation anymore. Honestly, sometimes we're not even sure about the correct pronunciation ourselves, and we worry about teaching her the wrong way.* *(Interviewee 8) The teacher often gives reading words as homework and sometimes asks parents to check their kid's pronunciation. Honestly, I don't know much about English and can't really tell if she's saying the words right. In those moments, I just nod along, not feeling too sure. But with this speech recognition feature, it's so much easier! The app automatically picks up on her pronunciation and gives instant feedback. If she makes a mistake, it points it out and shows her the correct way to say it. This way, she learns more accurately, and I don't have to stress about not understanding or*

			teaching her the wrong way. *(Interviewee 5) As parents, we want to help our kids practice, but pronunciation can be tricky because it's so easy to be influenced by accents. Honestly, our own pronunciation isn't perfect either, so we're often hesitant to teach them ourselves—worried we might pass on the wrong way to say things. But this app is great because it can accurately pick out where my kid's pronunciation needs improvement. It really takes the pressure off us and makes sure they're learning the right way.*
c. Potential barriers to using AI apps for learning	c.1 Barriers related to children	*c.1.1 Lack of Focus*	*(Interviewee 1) My child doesn't have much self-discipline, so I have to be there all the time to supervise and make sure they stay on track...Not open other entertainment apps.* *(Interviewee 5) I think kids are still pretty young and not fully developed when it comes to self-discipline and managing themselves, which is something we parents worry about. A lot of the time, when they're using AI apps to learn, they get distracted really easily—one-minute they're playing games, and the next they're watching cartoons. This makes their learning less efficient and pulls them away from their original goals.* *(Interviewee 3) The biggest problem is that he has trouble staying focused and gets distracted really easily. He'll start studying but then end up playing with something else. So, my husband and I always have to sit with him, keeping an eye on him and encouraging him to stay on task. But honestly, we're both busy with work and can't always be there with him. Because of that, his learning isn't as effective. When we're not around, he tends to slack off with the app, doesn't take it seriously, and has a hard time sticking to it.*
		c.1.2 Difficulty with Content	*(Interviewee 7) The main problem is that sometimes the content is just a bit too hard. My child already has a weak foundation in English, so when the material gets too challenging, he really struggles to keep up and has a hard time understanding it.* *(Interviewee 2) Sometimes the content in the app can be quite challenging, especially when it comes to difficult words. My child struggles with pronunciation in those situations, so I usually play the pronunciation multiple times, hoping he can pick it up. However, I've noticed that if he feels he can't pronounce the words correctly, he quickly loses interest and may even*

			become reluctant to continue practicing altogether. *(Interviewee 4) The app's content difficulty could use some adjustment. I mean…some of the material is just too hard for young kids, and they can't remember really long words at all.*
		c.1.3 Complex Interface	*(Interviewee 2) My child is still pretty young, so at first, he didn't really know how to use the app. He would just tap on things randomly and needed us to be there to guide him step by step.* *(Interviewee 5) When my kid first started using the app, he felt pretty overwhelmed. The menus were confusing, and he had no idea where to start. We had to sit down with him, go through it together, and show him how to use it step by step.*
		c.1.4 Inappropriate Content	*(Interviewee 2) Sometimes the cartoons have bad guys and intense scenes that are just too much for little kids. They don't really get why the characters are fighting, and they might end up copying the bad behaviour. If the kids are already a bit rebellious and like to copy things, it could lead to some bad habits.*
		c.1.5 Eye Health	*(Interviewee 6) His mom's always saying he should spend less time on his phone because it's bad for his eyes. The kid next door already has to wear glasses, and he's so young—we really don't want that to happen to him as well.* *(Interviewee 7) Spending too much time staring at a phone screen can also be harmful to the eyes.* *(Interviewee 3) Yeah, and too much screen time can really hurt her eyes. If her vision starts getting worse, I'll probably have to cut down on how much time she spends on the screen.* *(Interviewee 4) We always make sure to limit his screen time at home because we don't want him ending up needing glasses so young.*

		c.1.6 Technical Barriers	*(Interviewee 7) I've noticed that the app seems to require a more advancedphone to run smoothly. Ourfamily has an average financial situation and we're using an older device, and the app keeps freezing or crashing. My kid ends up losing his progress and has to start over, which really frustrates him. It's especially tough when he'sfinally motivated to learn—itjust kills his interest and makes learning way less effective.* *(Interviewee 8) Another problem is the internet connection. In our rural area, it's often slow and unstable. If the app needs to be online all the time, it gets really frustrating. Sometimes, just as my kid gets into learning, the connection drops, and it messes up their progress. It's hardfor them to stayfocused and motivated like that.*
	c.2 Barriers related to parents	*c.2.1 Cost*	*(Interviewee 7) I have to think about cost-effectiveness. As he gets older, he'll need tutoring in other subjects ... Chinese and math, for example. If the app charges afee, we might not gofor it because, honestly, I don't think it's as effective as having a real teacher.* *(Interviewee 1) We also have to think about the cost. My husband and I are planning to enrol her in offline one-on-one tutoring classes. If the app staysfree, we'd use it occasionally, and that'sfine. But if it starts costing more than what we'd payfor tutoring, we probably wouldn't keep using it. I'd be okay with paying up to a third of what we'd spend on offline classes since I see the app more as a supplementary tool to get her interested.* *(Interviewee 6) If it gets more expensive, her parents will probably have to think about whether it's really worth the cost.*
		c.2.2 Requires a lot of timefor supervision	*(Interviewee 1) Ifparents are really busy, they don't have much time to sit with their kid while using the app, especially since they also need to help with other schoolwork and only have so much time in the evening. It'dbe much better if the app helped kids learn on their own without needing us to constantly supervise.* *(Interviewee 2) It really depends on our schedules. If my husband and I get too busy with work, we might not be able to keep supervising our child's learning regularly. In that case, we'dprobably look into offline tutoring classes where the teachers can directly guide and support them.*

		c.2.3 Lack of Familiarity with Technology Among Rural Parents	*(Interviewee 6) I feel like I'm falling behind the times and don't really get how all these new technologies work.* *(Interviewee 8) Well... I'm not familiar with that AI technology.*
		c.2.4 More reliance on traditional education models	*(Interviewee 5) I think these apps are more like extra tools and shouldn't be something we totally depend on. Kids should still focus on learning at school in the end. What they learn in class and the teacher's guidance matter the most. You can't expect an app alone to get them perfect test scores.* *(Interviewee 1) It really depends on the schoolteachers. If they recommend using it at home, we'll keep using it. But if they say it's not a good idea, we'll probably drop it.* *(Interviewee 2) I think this app is just a supplementary tool for learning English. Teachers have more teaching experience and can provide face-to-face instruction. In the classroom, teachers can monitor children's learning progress in real-time and address issues immediately. With teachers' involvement, we as parents feel more at ease.* *I think this app is just an extra tool to help with learning English. Teachers have way more teaching experience and can teach face-to-face. In class, they can keep an eye on kids' progress and fix any problems right away. With teachers involved, we as parents feel a lot more comfortable.*
d. Future suggestions of AI-based learning	d.1 Functional improvements	*d.1.1 Peer-support*	*(Interviewee 5) I hope in the future, this app can have a smart reading companion feature. Well, just like a friend who learns with the kid. It could teach them English, chat with them, and even play games…learning would feel more fun and not like such a task.*

application enhancements		*d.1.2 Enhancing Teaching Functions*	*(Interviewee 1) I hope these AI applications can continue to advance, perhaps even reaching a level where they can replace real teachers. If AI could teach children as effectively as a teacher, that would be amazing.* *(Interviewee 3) I hope that in the future, AI applications can better understand each child's needs. For instance, the app could automatically adjust the difficulty level based on my child's abilities, avoiding content that's too challenging. Enhancing personalised features would be a great addition since every child learns differently. It would be even better if the content could be tailored to my child's preferences.*
		d.1.3 Improving Parent Support Features	*(Interviewee 2) I hope they can develop more intelligent features, primarily to improve children's English proficiency... It would be helpful if these AI apps were designed to be more user-friendly, with features like clear learning plans and reminders to help parents keep track of their children's progress. Even if we're busy, our kids would still be able to stay on track with their learning.* *(Interviewee 3) Moreover, better communication with parents would be helpful, such as providing updates on what the child has learned and their progress, so we can have a clearer understanding of their learning journey.*
		d.1.4 Integration of AI Applications with Traditional Teaching	*(Interviewee 5) I believe it would be ideal if schoolteachers could integrate this type of learning app into their classrooms. Traditional teaching methods can sometimes be a bit monotonous, but teachers' guidance is essential since they have more experience working with children. Combining both approaches would likely lead to better learning outcomes.*

	d.2 Technology competence	*d.2.1 Offline Functionality*	*(Interviewee 8) I think it'd be awesome if the app had an offline mode. That way, even if the internet isn't great, kids could keep learning without any interruptions. For places with unstable connections, this would make it so much easierfor kids to use the app without any worries.*
		d.2.2 Supportfor dialects	*(Interviewee 7) I really hope these AI apps can be simpler and easier to use, and it'dbe great if they supported dialects too. In our community, a lot ofpeople speak in dialects since their Mandarin isn't very standard, which makes using the app a bit tricky. If the app could recognise and understand dialects, it'd be so much better—I mean it'd help parents get more involved in their kids' learning.*
		d.2.3 Adult education: Parental Support and Training	*(Interviewee 3) It'd be really helpful if the app could share more updates with parents, like what the child has learned and how they're doing…we'd have a better idea of their progress and could support them more effectively.* *(Interviewee 8) Besides the network issues I mentioned, I really hopefuture apps can think about getting parents involved, especiallyfor those of us who aren't great with tech. It'dbe helpful to have simple guides or even a bit of training—notjust for the kids, butfor parents too—on how to use the app properly. So we can support our kids better and make the whole learningprocess smoother.*

5.2.3.1 Learning Outcomes after children use AI-based learning applications

5.2.3.1.1 Knowledge Acquisition

This section presents a comprehensive account of parents' and caregivers' feedback on their children's learning outcomes following the use of AI-based learning apps. The feedback not only covers changes in children's knowledge acquisition but also includes shifts in their behavioural habits, improvements in learning attitudes, and an increase in learning motivation.

Specifically, parents and caregivers reported that AI-based learning applications had a significant impact on their children's vocabulary acquisition. Half of the parents noted that their children encountered a broader range of words through these apps, resulting in a substantial increase in their vocabulary. One parent further observed that their child's language development extended beyond word memorisation, evolving from mastery of basic vocabulary to the independent construction of simple sentences (sub-code a.1.1; Interviewee 7). Additionally, parents observed that these gains were not limited to vocabulary improvement but also positively influenced their children's English learning performance. For instance, one parent highlighted that their child's homework completion rate improved markedly, earning praise from the teacher (sub-code a.1.1; Interviewee 1). Another parent mentioned a significant improvement in their child's performance on English tests (sub-code a.1.1; Interviewee 3), suggesting that AI-based learning applications contributed to academic success.

Moreover, approximately half of parents and caregivers noted substantial improvements in their children's English pronunciation. For example, one parent reported that their child no longer needed to repeat words multiple times during reading exercises and was able to pronounce them correctly more quickly (sub-code a.1.2; Interviewee 5). Another parent shared that their child overcame challenges with certain sounds, particularly making notable progress with sounds such as ‘th’ and ‘r’ (sub-code a.1.2; Interviewee 3).

While some parents emphasised the positive impact of AI-based learning applications on learning outcomes, others expressed concerns about their effectiveness in supporting children's L2 learning. One parent highlighted that these applications predominantly rely on memorisation and repetition to teach vocabulary, which may not be suitable for children whose cognitive abilities are still developing. Consequently, children often forget much of what they have learned. Although continuous practice may lead to some improvements, the parent remained doubtful about whether these apps could genuinely facilitate comprehensive mastery of a second language (sub-code a.1.1; Interviewee 2). These observations underscore the limitations of AI-based learning applications, particularly when relying on rote memorisation, as they may struggle to ensure long-term retention or support the full acquisition of a second language (Lee et al., 2015).

Furthermore, a parent from a rural area noted their unfamiliarity with these applications (sub-code a.1.1; Interviewee 6), suggesting that access to or understanding of such tools may present a barrier for some families.

In conclusion, parents' feedback indicates that AI-based learning applications have been notably effective in enhancing children's vocabulary and pronunciation and have had a significant impact on their academic performance and language development. These positive outcomes have been widely acknowledged in both home learning and classroom settings.

5.2.3.1.2 Learner Attitudes and Motivation

In the discussion of learning attitudes, almost two-thirds (63%) of parents and caregivers reported that AI-based learning applications significantly enhanced their children's interest and motivation in learning English. Gamified learning and multimedia content are known to effectively increase learner motivation and engagement, particularly in language learning contexts (Dehghanzadeh et al., 2021). Specifically, the application's use of animations, games, and interactive features effectively captured the children's attention, transforming English learning into an enjoyable and engaging experience. This design significantly stimulated the children's enthusiasm for learning, encouraging them to actively participate in educational activities (sub-code a.2.1; Interviewees 1, 4, 2, 5, 7). Additionally, two parents specifically highlighted that these applications were particularly effective for younger children (sub-code a.2.1; Interviewees 4, 5). The use of cartoon characters and engaging design elements not only enhanced the enjoyment of learning but also was consistent with the cognitive development stages and interest preferences of younger children.

Furthermore, the use of AI-based learning applications significantly enhanced the children's confidence in both learning and using English. One parent reported that their child, who had previously avoided speaking in English class due to a fear of making mistakes, became more confident and began actively participating in class discussions after using the app (sub-code a.2.2; Interviewee 8). Another parent noted that their child showed greater confidence, willingly engaging in English-related activities after using the application (sub-code a.2.2; Interviewee 5). These experiences suggest that the application's continuous practice and positive feedback mechanisms played a crucial role in fostering the children's confidence in their language abilities.

5.2.3.2 Behavioural Changes

Many parents reported positive changes in their children's learning behaviours. The interactive features of AI-based learning tools enhanced children's engagement in learning and fostered greater independence, reducing their reliance on external supervision. Specifically, around a quoter of parents observed a significant improvement in their children's focus and engagement after using AI-based learning applications. For instance, one parent noted that their child, who had previously struggled to remain still and would often become distracted while reading, was now able to sustain focus for longer periods (sub-code a.3.1; Interviewees 1 and 3). Furthermore, AI-based learning applications appeared to enhance children's independent learning abilities. One parent shared that their child, who had once resisted vocabulary memorisation, now independently reviewed and recited words at home (sub-code a.3.2; Interviewee 7). Another parent indicated that, while frequent reminders were once necessary, their child now initiated learning sessions on a phone or tablet without further prompting (sub-code a.3.2; Interviewee 5). These findings suggest that AI-based learning applications not only contribute to language skill development but also play a significant role in developing independent learning habits.

In addition to academic participation, AI-based learning applications have stimulated children's interest in cultural learning. One parent reported that their child had learned about various festivals and customs in English-speaking countries, such as Thanksgiving and Halloween, and was able to explain these traditions in detail. This increased cultural awareness prompted the child to actively ask questions, demonstrating a strong curiosity about cultural understanding (sub-code a.3.3; Interviewee 4). Another parent noted that their child now enjoys engaging with foreigners and has even started discussing plans for studying abroad in the future (sub-code a.3.3; Interviewee 3). This cross-cultural learning not only expanded the child's knowledge but also fostered greater respect and understanding of diverse cultures. Furthermore, this curiosity can serve as motivation to learn related knowledge (Singh & Manjaly, 2022).

5.2.3.3 Parent-caregivers' Motivationfor using AI-basedlearning applications

This theme examines the factors motivating parents to choose AI-based learning applications. The research identifies three key reasons: the flexibility of learning anytime and anywhere, the cost-effectiveness of the applications, and their potential to reduce parent-caregivers' educational stress. These factors were found to be the primary motivations behind parent-caregivers' decisions to encourage their children to use these tools.

5.2.3.3.1 Flexibility in Learning

Around one- third of parents reported that these applications substantially improve the convenience and accessibility of their children's learning. Some parents mentioned that their children can use smartphones to study English at home, making efficient use of their free time while accessing a variety of learning resources (sub-code b.1.1; Interviewees 2 and 7). Additionally, some parents reported that these applications allow their children to avoid traditional English tutoring classes, enabling independent study according to their own schedules, thus reducing the burden on parents (sub-code b.1.1; Interviewee 8). This flexibility not only reduces parents' time and financial pressures but also offers children a more personalised learning experience. This change is especially significant in the context of China's 2021 ‘Double Reduction’ policy, which restricts the expansion of traditional tutoring classes. As a result, many parents and students now turn to alternative learning tools, such as AI-based learning applications, to meet educational needs.

5.2.3.3.2 Cost-Effectiveness

In addition, the cost-effectiveness of AI-based learning applications plays a significant role in parent-caregivers' decision-making process. One parent pointed out that, in comparison to traditional or one-on-one tutoring classes, these applications are more affordable, resulting in savings in both time and financial resources (sub-code b.2.1; Interviewee 7). Another parent added that many of these applications are free, which effectively eases the financial burden of family education while still providing high-quality learning support for children (sub-code b.2.1; Interviewee 5).

5.2.3.3.3 Reducing Parent-caregivers’ Tutoring Stress

Moreover, the speech recognition feature of AI-based learning applications plays a crucial role in reducing the tutoring stress on parent-caregivers. This pressure primarily arises from parent-caregivers' uncertainty regarding the correct pronunciation of words when assisting their children. According to one-third of the interviewed parents, this feature automatically detects and corrects pronunciation errors, offering immediate feedback. For parents with limited confidence in their own English proficiency, this functionality reduces concerns about potential teaching inaccuracies, thereby easing the pressure they experience when guiding their children (sub-code b.3.1; Interviewees 1, 5, 8).

5.2.3.4 Potential barriers to using AI-based applications for learning

The potential barriers to using AI-based learning applications can be broadly categorised into two main groups: child-related barriers and parent-related barriers. These challenges primarily focus on the practical use of the applications, including children's usage and learning habits, as well as the challenges parent-caregivers face in terms of resource investment, technical support, and educational philosophy. These barriers reflect the limitations of AI-based learning tools in practice.

5.2.3.4.1 Barriers related to children

Some parents reported challenges with their children's attention and self-discipline while using AI-based learning applications. During the interviews, parents mentioned that continuous supervision was necessary to ensure their children remained focused on learning tasks (sub-code c.1.1; Interviewee 1). Additionally, some parents indicated that younger children still struggle with self-management and resisting distractions. Interviewees 5 and 3 (sub-code c.1.1) noted that their children had difficulty maintaining focus, frequently becoming distracted while using the AI-based learning applications, switching between playing games and watching cartoons, which led to decreased learning efficiency. These findings are consistent with existing research on the impact of screen use on children's attention. Studies have shown that extended use of screen-based devices, particularly those with rapidly changing content, can shorten children's attention spans (Vedechkina & Borgonovi, 2021).

In addition to concerns regarding attention, parents expressed concerns about the difficulty level of the application's content. One parent noted that some materials were too challenging for children with limited English proficiency, causing them to fall behind and experience frustration (sub-code c.1.2; Interviewee 7). Another parent mentioned that when children struggled with pronunciation or complex vocabulary, they often lost interest and refused to continue practising (sub-code c.1.2; Interviewee 2). Research has shown that excessive difficulty can lead to frustration, thereby diminishing learners' motivation (Cai et al., 2023).

Some parents expressed concerns that the application's interface might confuse children (sub-code c.1.3; Interviewees 2 and 5), with this issue being particularly raised by parents of younger children. One parent noted that their child struggled to navigate the application initially and required step-by-step guidance (sub-code c.1.3; Interviewee 2). The importance of interface simplicity and user experience design in educational technology has been widely discussed. Research indicates that for younger children, a clear and intuitive interface is essential for effective use, as it can enhance their learning engagement and efficiency (Masood & Thigambaram, 2015).

Concerns regarding the content of educational applications have been raised by one parent, particularly with respect to their appropriateness for young children. She mentioned that the “villain” characters or intense scenes in the animations could potentially confuse children and even encourage them to imitate negative behaviours (sub-code c.1.4; Interviewee 2). In addition, excessive screen time is frequently cited by parents as a significant health concern, with many associating extended uses with potential harm to children's eyesight. Half of the parents were worried that extended use of the app could adversely affect their children's eye health, and several have proactively implemented measures to limit screen time to prevent the early onset of vision problems (sub-code c.1.5; Interviewees 3 and 4).

Finally, technical issues, particularly the limitations of device performance and network stability, were frequently mentioned by most rural parents. They reported that the application’s requirement for high-performance devices often led to interruptions in children’s learning due to lag or crashes on older devices, thereby negatively impacting learning progress (sub-code c.1.6; Interviewees 7 and 8). Technical challenges and device requirements are common barriers in educational technology. Studies have consistently shown that device performance and network stability directly influence the effectiveness of online learning (T. Chen et al., 2020).

5.2.3.4.2 Barriers related to parents

Time and cost are significant barriers preventing parents from choosing to use the application. Although many applications are available for free, a proportion of parents expressed concerns regarding the costs associated with premium features or subscriptions. One parent mentioned that, due to the need to pay for additional tutoring in other subjects, they may prioritise traditional tutoring over paid applications (sub-code c.2.1; Interviewee 7). Another parent indicated that they consider the application as a supplementary tool and would only continue its use if the costs were limited to less than one-third of the price of offline tutoring (sub-code c.2.1; Interviewee 1).

Additionally, around a quarter of parents highlighted that supervising their children's use of the application required a significant amount of time. They noted that this could lead to a very tight schedule (sub-code c.2.2; Interviewee 1). If their work becomes busier, they may be more willing to choose offline tutoring, as it offers direct teacher support and guidance (sub-code c.2.2; Interviewee 2).

Some parents also prefer traditional educational methods. Three parents argued that AI-based learning applications serve only as supplementary tools and cannot replace formal schooling. One parent emphasised that classroom teachers can provide real-time guidance and monitor their child's learning progress, thus offering more reliable educational support (sub-code c.2.4; Interviewee 2). Another parent noted that the decision to use the application is contingent on the teacher's recommendation; if the teacher does not endorse it, they are likely to discontinue its use (sub-code c.2.4; Interviewee 1). In general, parents place greater importance on the value of teachers' instructional experience and face-to-face interactions for supporting their children's learning, as compared to learning through apps; this view is consistent with previous research (Aruta et al., 2019).

For most rural parents, unfamiliarity with technology presents a challenge. Several rural parents admitted to being unfamiliar with modern technology, which makes it difficult for them to support their children's use of AI-based learning applications (sub-code c.2.3; Interviewees 6 and 8).

5.2.3.5 Future Suggestions for AI-based Learning Application Enhancements

To optimise AI-based learning applications, parent-caregivers have provided valuable insights that can be categorised into two primary aspects: functional enhancements and technology competence.

Regarding the future optimisation of these tools learning applications, parent-caregivers have expressed a desire for greater interactivity and personalised support in AI-based learning applications. Specifically, they advocate for the integration of advanced teaching functionalities and improved systems for parent-caregiver collaboration. These features aim to foster a more engaging and enjoyable learning experience for children while simultaneously ensuring measurable educational outcomes.

In terms of enhancing technical capabilities, parent-caregivers have highlighted the importance of technical features such as offline functionality, support for regional dialects, and resources tailored for parent training and engagement. These enhancements are particularly critical for reducing technical barriers, improving usability, and broadening access to education, especially in rural or resource-constrained areas (Hillier, 2020). Moreover, these improvements are expected to enable parent-caregivers to play an active role in their children's education, thereby enhancing overall learning efficacy.

5.2.3.5.1 Functional improvements

In the interview, one parent suggested that introducing peer-support features could better support student engagement and enjoyment in AI-based learning applications. He specifically proposed adding an innovative feature called the 'Smart Reading Companion'. This AI virtual assistant would accompany children in learning English, interact with them, and even participate in games, creating a supportive and engaging learning environment (sub-code d.1.1; Interviewee 5). Research indicates that peer interaction and engagement significantly enhance learning outcomes (Hilts et al., 2018), provide emotional support, and help learners develop a sense of belonging during the learning process (Zhao et al., 2021).

Another parent suggested that the application should incorporate more advanced personalisation features, such as automatically adjusting content difficulty based on children's learning abilities and preferences, to better meet the needs of each individual child (sub-code d.1.2; Interviewee 3). Research highlights that integrating personalised learning with AI-driven adaptive systems significantly enhances learning outcomes, particularly for students with diverse abilities. Dynamic content adjustment can effectively address the diverse needs of learners (Kolluru et al., 2018).

Additionally, one parent emphasised the importance of enhancing the application's usability, including user instructions and progress monitoring tools. He proposed that the application integrate features such as clear learning plans, progress tracking, and reminder notifications to enable busy parents to effectively supervise their children's learning (sub-code d.1.3; Interviewee 2). Furthermore, parents suggested improving communication functionalities by providing regular updates on their children's learning content and achievements. These enhancements would offer parents a more comprehensive understanding of their children's progressand and foster greater engagement in the educational process (sub-code d.1.3; Interviewee 3). Empirical studies underscore that parent-caregivers' involvement is a critical determinant of children's learning outcomes, particularly in technology-assistede ducational environments (Gonzalez-DeHass et al., 2022). These findings suggest that enhancing personalised support and parent collaboration features is a critical direction for improving AI-based learning applications.

Some parents have suggested integrating AI-based learning applications with traditional classroom instruction. One parent noted that traditional teaching methods can sometimes appear monotonous, whereas the use of AI tools not only enhances the engagement and dynamism of the content but also complements the teacher's central educational role. This hybrid approach could further improve teaching effectiveness (sub-code d.1.4; Interviewee 5). Blended learning models, which combine AI-based applications with conventional teaching, have been shown to provide students with broader educational opportunities and a richer learning experience (Cunningham, 2014). Research indicates that this model significantly enhances the academic performance and learning efficiency of remote students.

Furthermore, another parent expressed their expectations for the future development of AI technology, predicting that AI will eventually evolve to a level where it can effectively replace human teachers (sub-code d.1.2; Interviewee 1).

5.2.3.5.2 Technology competence

In areas with unstable internet connections, most rural parents (around two-thirds of the participants) expressed a desire for the app to support offline functionality. One parent pointed out that this feature would enable children to continue their learning even without an internet connection (sub-code d.2.1; Interviewee 8). This need is especially crucial in rural or resource-limited areas. Research suggests that learning apps with offline capabilities not only provide continuous learning opportunities for students in remote areas but also help address some of the challenges associated with the unequal distribution of educational resources (Hillier, 2020).

Another suggestion for improvement is to add support for local dialects. One parent mentioned that in communities where Mandarin is not widely spoken, an app that supports dialects could make learning more inclusive and accessible, helping both parents and children engage in the learning process more effectively (sub-code d.2.2; Interviewee 7).

Some parents also highlighted their significant role in their children's learning and recommended that future apps include more resources to support them. These resources could involve simple user guides, regular updates on children's learning progress, and even basic training for parents who are less familiar with technology. With these features, parents would be better able to assist their children more effectively, ensuring a smoother learning experience (sub-code d.2.3; Interviewees 3 and 8).

These suggestions offer valuable insights for the future development of AI-based learning applications, emphasising the need to balance student learning outcomes with the practical needs and challenges of parent-caregivers.

5.2.4. Summary of Qualitative Findings

This research collected qualitative data to address RQ3. The findings indicate that most parents hold a positive attitude towards their children's use of AI-based learning applications for L2 learning. Based on parent-caregivers' feedback, the key learning outcomes observed after children used AI-based learning applications can be categorised into three main aspects: 1) Knowledge Acquisition; 2) Learner Attitudes and Motivation; and 3) Behavioural Changes.

In addition, this study explored parent-caregivers' motivations for choosing these applications and analysed the challenges faced by both children and parents during the usage process, which affected their willingness and ability to continue using them. Parents commonly mentioned time constraints, cost concerns, and unfamiliarity with technology, all of which hindered the effective use of these tools to some extent.

Finally, parent-caregivers shared their expectations for future AI-based learning applications' features, hoping that the applications could offer more personalised learning paths, support offline learning, integrate with school curricula, and include additional parent-caregivers' support tools. These findings provide valuable insights for improving AI-assisted L2 learning applications for children, making them more effective in meeting the needs of both children and parents.

5.3 Integration of Quantitative and Qualitative Findings

To deepen the understanding of the intervention's effects, this section presents a methodological triangulation of the findings by integrating the experimental and questionnaire data with semi-structured interviews. The experimental and questionnaire data provided quantitative evidence on learning gains, app engagement, and hedonic experience, while the interviews offered contextualised explanations for the variability observed in the quantitative patterns. This design enabled the strengths of one method to compensate for the limitations of another, particularly in relation to children's adherence patterns and their motivations for using the application.

Due to the ceiling effect observed in the parent logs, children in the experimental group were categorised into a 'Maximal Adherence' subgroup (completion of all 5 required days per week; $N = 22$) and a 'High Adherence' subgroup (<5 days per week; $N = 21$).

Independent-samples t-tests indicated distinct learning and engagement profiles between the two groups. As presented in Table 5-8, children with maximal adherence achieved substantially greater vocabulary gains (Δreceptive single-word vocabulary breadth) (M = 18.64) than those with variable adherence (M = 9.81), t (35.48) = 6.11, p < .001, with a large effect size (Cohen's d = 1.84). Maximal adherence was also associated with significantly higher hedonic ratings (M = 17.64 vs. M = 14.00), t (21.62) = 4.26, p < .001, Cohen's d = 1.33. Notably, the High Adherence subgroup showed much greater variability in hedonic experience (SD = 3.83) than the Maximal Adherence subgroup (SD = 0.79), suggesting heterogeneous engagement processes.

Table 5-8. *Independent Samples t-test Results for Study Variables by Adherence Group.*

Measure	Adherence Group	*N*	*M*	*SD*	*t(df)*	*p*	Cohen's *d*
ΔReceptive single-word vocabulary breadth	Maximal Adherence	22	18.64	5.71	6.11(35.48)	*p* < .001	1.84
	High Adherence	21	9.81	3.57			
Hedonic rating	Maximal Adherence	22	17.63	.79	4.26(21.62)	*p* < .001	1.33
	High Adherence	21	14.00	3.83			

Note. Welch's independent-samples *t*-tests were used due to unequal variances. *ΔReceptive single-word vocabulary breadth = post-test minus pre-test PPVT-5 scores.*

In summary, the statistical comparisons indicate a clear pattern favouring the ‘Maximal Adherence’ subgroup. The data suggest that consistent usage, defined as completing the full five-day regimen, is closely associated with stronger vocabulary acquisition and a more positive and stable hedonic experience. In contrast, the substantial variability observed within the ‘High Adherence’ group suggests that irregular usage may interrupt the continuity needed for both effective learning and sustained enjoyment.

While the quantitative findings establish an association between adherence and learning outcomes, the interview data provide further insight into the behaviours underlying these patterns. Consistent with the stability observed in the ‘Maximal Adherence’ group, qualitative accounts pointed to a shift towards more autonomous engagement. For example, Interviewee 5 reported that their child began to “initiate learning sessions without further prompting”, indicating that high adherence was often supported by the development of independent learning habits (sub-code a.3.2).

By contrast, accounts from parents in the ‘High Adherence’ (variable) group frequently referred to difficulties with distractibility and a continued need for supervision. As noted by Interviewees 1 and 3, some children struggled to remain focused without ongoing parental monitoring and often switched to non-educational content (sub-code c.1.1). These observations are consistent with the irregular usage patterns identified in the quantitative data.

Finally, the high variability in hedonic experience (SD = 3.83) among variable users is reflected in parental reports of frustration related to task difficulty. Although many parents described the gamified elements as enjoyable, Interviewee 7 noted that excessive difficulty led to “frustration” and, in some cases, refusal to continue practising (sub-code c.1.2). This contrast between children who found the application engaging and those who experienced frustration helps to explain the wide spread of affective ratings within the less adherent group.

5.4 Conclusion

This chapter presented the results of both quantitative and qualitative data analyses to address the three research questions (RQ1, RQ2, and RQ3).

Regarding RQ1, the quantitative results provided robust evidence for the intervention's efficacy, with the experimental group demonstrating significant gains in receptive vocabulary breadth (fully supporting Hypothesis H1a) and working memory. Crucially, these statistical improvements were corroborated by qualitative insights from RQ3, where parents reported distinct gains in children's knowledge acquisition and confidence. However, Hypothesis H1b was only partially supported; regression analysis indicated that receptive vocabulary growth was primarily predicted by group allocation, age, and WM rather than ToM gains, suggesting complex developmental dynamics.

In terms of children's hedonic experience (RQ2), the survey data revealed high levels of satisfaction and exceptional adherence, thereby confirming Hypothesis H2. Additionally, ease of use, interactive and gamified features, and high-quality visual presentation significantly influenced children's hedonic experience. Statistical analyses highlighted that children's hedonic experience was a pivotal predictor of both behavioural engagement and vocabulary acquisition. Qualitative data provided a deeper understanding of this dynamic from RQ3. While parent-caregivers praised the interactive design and ease of use as primary motivators, they also identified several barriers to the long-term use of AI-based learning applications, including technical adaptability, supervision requirements, cost concerns, and the irreplaceability of traditional classroom instruction.

Finally, based on these experiences, parent-caregivers expressed practical expectations for future AI-based learning application improvements, emphasising the need for personalised learning pathways, offline support, integration with school curricula, and parent-caregivers' monitoring tools.

These findings will be compared with and discussed in relation to relevant literature in the next chapter to further explore the role and impact of AI-based learning application (Zebra AI) in children's L2 learning.

Chapter 6. DISCUSSION

This chapter provides a comprehensive analysis and discussion of the research findings and main outcomes of this study. The presentation and interpretation of these results are based on the mixed-methods approach adopted in this dissertation, which combines both qualitative and quantitative methods and adopts a range of data collection strategies to explore answers to three core research questions. Specifically, data were gathered through experiments, questionnaires, and interviews, allowing for an in-depth analysis from multiple perspectives, including children's language learning and parent-caregivers' viewpoints. The research subjects comprised Chinese children aged 5–7 and their parents or primary caregivers. Furthermore, to provide a more comprehensive explanations of the research questions, this chapter compares and integrates these findings with existing related studies, analysing their consistency and unique contributions in relation to the existing literature, to address the core research questions of this study.

RQ1: To what extent does the AI-based app improve children's receptive vocabulary, and how are these outcomes predicted by their WM and ToM?

RQ2: Which functions and design interface influence children's hedonic experience with the AI-based language learning application?

RQ3: What are the parent-caregivers' views regarding the effectiveness of using AI-based learning apps for learning a second language?

6.1 Triangulation

Triangulation is a research strategy that integrates multiple methods, data sources, and theoretical perspectives to investigate the same phenomenon, with the aim of enhancing the credibility, validity, and interpretive power of the research (Denzin, 1970). This approach overcomes the limitations and biases of any single method by cross-verifying results from diverse sources, thereby providing a more comprehensive understanding of the research problem. In the social sciences, triangulation has been widely applied. Denzin (2006) categorises data triangulation into three types: time, space, and individuals, based on the notion that the reliability of data maybe influenced by factors such as the time of collection, the identity of the participants, and the context in which the data are collected (Leech & Onwuegbuzie, 2007).

In this study, I primarily employed methodological, data, and theoretical triangulation to ensure a multidimensional examination of the role of AI-based learning applications in children's L2 learning. Specifically, methodological triangulation was achieved by combining experiments, surveys, and parent interviews: the experiments and questionnaires provided quantitative data on the impact of the AI-based learning application on children's L2 learning, while the parent-caregivers' interviews offered in-depth insights into their observations and reflections on the use of the AI-based learning application, thereby compensating for the potential biases and data gaps in any single method. Data triangulation was implemented by integrating data from experiments, questionnaires, and interviews, using the mutual corroboration among these diverse sources to further enhance the diversity and reliability of the research findings. Finally, theoretical triangulation was conducted by employing multiple theoretical frameworks, including sociocultural theory (Vygotsky, 1978), self-determination theory (Deci & Ryan, 2000), and the technology acceptance model (Davis, 1989), to interpret the phenomenon from different angles and provide a robust theoretical foundation for data analysis and discussion. Through this multi-level triangulation approach, the study achieved systematic validation on methodological, data, and theoretical fronts, thereby significantly enhancing the overall credibility and interpretive power of the research conclusions.

6.2 Research Findings

The following sections, based on the quantitative and qualitative data collected in this research (see Chapter 5), provide an in-depth discussion of the role of AI-based learning applications in children's L2 learning. Specifically, Section 6.3.1 examines the effectiveness of the application in facilitating children's L2 learning; Section 6.3.2 analyses which functions and design factors influence children's hedonic experiences; and

Section 6.3.3 discusses, from the perspective of parents or primary caregivers, their evaluations of their children's use of the application. The overall objective is to assess the outcomes achieved by integrating AI-based learning applications into L2 learning interventions among Chinese children aged 5 to 7.

6.2.1 To what extent does the AI-based app improve children’s receptive vocabulary, and how are these outcomes predicted by their WM and ToM?

The findings in Chapter 5 indicate a positive relationship between the use of AI-based learning applications and children's L2 learning. Quantitative data from the pre-test and post-test results demonstrate that children’s receptive single-word vocabulary breadth significantly improved after the intervention using Zebra AI for English learning. This result is consistent with the findings from participant interviews, where approximately half of the parents reported that their children's vocabulary expanded after using this application. Taken together, these findings suggest that the observed receptive vocabulary gains can be meaningfully interpreted through a convergence of theoretical perspectives from cognitive psychology, SLA, and CALL.

From a cognitive perspective, the findings are consistent with core principles of MLT (Mayer, 2005) and the Input Hypothesis (Krashen, 1985). In particular, Zebra AI’s central design feature of presenting animated narrative contexts in close temporal and spatial alignment with native-speaker auditory input is consistent with the temporal contiguity and modality principles proposed by Mayer (2005). Such alignment is likely to facilitate dual-coding processes and may reduce extraneous cognitive load. This design can be interpreted as providing favourable conditions for associative binding within the episodic buffer, thereby supporting the retention of word–meaning mappings through strengthened links between visual referents and verbal labels. While the Input Hypothesis does not specify detailed cognitive mechanisms, it offers a useful pedagogical lens for understanding how enriched and comprehensible input may support early lexical development in young learners.

Beyond input processing, the potential contribution of the ASR tasks can be interpreted by combining Chapelle's (1998) CALL framework with the Interaction Hypothesis (Long, 1996). According to Chapelle, computer-mediated negotiation is critical because it provides opportunities for corrective feedback, which drives language development. Although the application does not involve human-to-human negotiation, the ASR feedback loop creates a form of technology-mediated interaction that mimics the interactional modification found in natural conversation. When the ASR system rejects an incorrect pronunciation, it provides negative feedback similar to a clarification request or a signal of communication breakdown. This feedback creates opportunities for 'noticing' (Schmidt, 1990), drawing the learner's attention to mismatches between their own interlanguage and the target form. This process encourages learners to engage in self-repair and modify their subsequent attempts, thereby facilitating more precise phonological encoding without explicitly relying on pedagogical instruction. Importantly, these interpretations reflect theoretically reasonable mechanisms grounded in the interactive features of the design, rather than processes directly measured in the present study.

From a broader perspective, these findings are consistent with previous research, which suggests that technology use can enhance children's second-language vocabulary acquisition (Eltalhi et al., 2021). Building on this broader evidence base, more recent studies have specifically highlighted the effectiveness of AI-based language learning applications, which have been shown to have a significant impact on promoting children's second language vocabulary (Vadivel et al., 2023). For example, Liu and Chen (2023) found that elementary school students who used AI-based applications for English learning scored significantly higher in post-tests compared to those who learned using traditional, non-AI technologies. Similarly, Wen et al. (2024) conducted a study on 140 second-grade students in Singapore and observed that after 8 to 10 weeks of continuous use of the AI-driven ARCHe application, these students showed significant improvements in their second-language vocabulary acquisition of Chinese characters. These studies further highlight the potential of AI technology in enhancing children's language learning outcomes, particularly in vocabulary acquisition.

However, not all studies support this perspective. Rachels and Rockinson-Szapkiw (2018) reported that after using Duolingo in the classroom for 12 weeks, there was no significant difference in Spanish proficiency among third- and fourth-grade students. Similarly, James and Mayer (2019) found no significant differences between undergraduate students who used Duolingo at home for seven lessons to learn Italian and those who learned Italian using online slides. One possible explanation is the limited exposure to the intervention: in the first study, students used the application only once per week, and in the second, the intervention lasted just seven lessons, conditions that may have been insufficient to produce measurable effects. In contrast, the present research involved a longer intervention period, spanning 12 weeks, with students using the AI-based learning application five days per week. This extended exposure may explain why the experimental group outperformed the control group. Therefore, the duration of AI-based learning app usage appears to have influenced vocabulary acquisition levels between the two groups.

To deepen the understanding of this 'dosage effect,' the triangulation of data in the present study offers further explanatory power. Within the experimental group, higher vocabulary gains were strictly associated with maximal adherence. Qualitative findings reveal that this consistency was largely driven by children's 'autonomous engagement.' High-adherence learners often initiated sessions independently, reflecting a shift towards autonomous motivation as described in SDT (Ryan & Deci, 2000). Conversely, for children with variable adherence, parents frequently cited a dependence on external prompts and the 'need for supervision.' This reliance on external regulation often failed to sustain continuity when faced with distractions. This implies that whilst the extended duration provided the opportunity for learning, the actualisation of gains was mediated by the nature of the child's motivation, namely the capacity for self-determined engagement versus reliance on external control. The specific design mechanisms within the application that foster this sense of autonomy and competence will be further analysed in the section 6.2.2.

Although the quantitative data in the current study reveals a trend of receptive single-word vocabulary breadth growth among children, its limitation lies in its inability to directly capture how children use newly learned words in real-life communication. In contrast, the qualitative data compensates for this shortcoming by providing deeper insights into the practical application of vocabulary. Findings from parent-caregivers' interviews indicate that children's language development extends beyond mere vocabulary memorisation.

Instead, they progress from mastering basic words to independently constructing simple sentences (Interviewee 7). This observation is consistent with Nation (2001) who suggests that children do not remain at the passive recognition stage of word learning but gradually transition to active usage. This shift reflects the transfer of vocabulary knowledge to higher levels of language proficiency. This implies that the combination of high-dosage input and interactive features not only expanded the mental lexicon but also facilitated the retrieval processes necessary for productive language use.

Despite these positive findings, it is important to clarify the nature of the gains observed in this study. As the research compared the language learning app against a control condition (a non-language app), the observed improvements reflect a 'content effect' defined as the benefit of exposure to targeted language content rather than isolating the specific efficacy of AI-driven features (such as adaptive algorithms or ASR). While the theoretical mechanisms discussed above (such as multimodal input and interactive feedback) are plausible, future dismantling studies are needed to verify the specific causal contributions of individual AI components.

The experimental data not only confirm the positive impact of AI-based learning application on receptive single-word vocabulary breadth but also reflect changes in children’s cognitive abilities. SLA relies not only on the enhancement of vocabulary input but is also constrained by learners’ cognitive abilities, particularly the role of WM in associative binding and retention, as well as the contribution of ToM in language comprehension and social interaction (Doughty & Long, 2008). Therefore, the following section will further discuss the potential impact of the intervention on children’s cognitive resources.

In the cognitive ability assessment, children’s WM was evaluated. The results indicate that the experimental group showed a significant improvement in WM scores from pre-test to post-test, suggesting a potential positive relationship between the AI-based intervention and cognitive development.

While the Zebra AI intervention did not explicitly target WM training, theoretical mechanisms concerning multimodal processing offer a plausible explanation for these gains. Specifically, the application's design is consistent with the Associative Binding function of the Episodic Buffer in Baddeley's model (2000). Zebra AI delivers vocabulary instruction through highly integrated animated narratives, where each target word is embedded within a specific story-driven context. By consistently presenting native-speaker auditory input in synchronous alignment with visual stimuli, the system establishes a dual-channel learning environment. This presentation likely necessitates the continuous cross-modal binding of phonological and visuospatial information to extract meaning. Consequently, repeated engagement in this cognitive binding process over the intervention period may have exercised the children's capacity to hold and integrate multimodal information, thereby indirectly training the specific cognitive resource measured by the WM task.

Beyond passive input, features such as 'listen-and-select' games (e.g., selecting the correct image of an 'apple' from a four-quadrant grid upon hearing the corresponding auditory prompt) require active cognitive manipulation. To succeed, children must maintain the phonological instruction in memory while simultaneously scanning visuo-spatial information. This process engages the Central Executive to manage attention and exercise inhibitory control over irrelevant distractors (Miyake et al., 2000). This is consistent with the Retrieval Practice Effect (Roediger & Karpicke, 2006), transforming the activity from simple exposure into a complex 'storage-plus-processing' task. Such high-intensity online manipulation exercises the attentional resources required for efficient binding, thereby likely contributing to the observed gains in WM scores.

These findings align with existing research suggesting that multimedia learning environments can support the development of memory systems (Alzubi et al., 2018; Johann & Karbach, 2018; Lamrani & Abdelwahed, 2020). For example, the findings of this study are consistent with those of Gu (2024), who reported that integrating technology-driven online tools such as Quizlet into vocabulary instruction can enhance undergraduate students' WM. Since WM provides the necessary workspace for associative binding, the cognitive benefits observed by Gu support the plausibility of the WM gains found in the current study. This consistency across different age groups provides preliminary support for the applicability of AI-based language learning at various developmental stages. However, it is important to note that the two studies differ significantly in terms of their participants. While Gu’s study focused on university students, the present study targeted children aged 5 to 7, who differ in cognitive development and learning strategies.

It is important to note that this study did not isolate specific AI features to test their individual contributions to cognitive gains. Therefore, the interpretation that the app caused WM improvement through associative binding remains hypothetical. Future research should employ dismantling designs to verify whether specific features, such as the synchronicity of audio-visual input or the frequency of gamified retrieval tasks, are the causal drivers of these WM improvements.

Moreover, previous research has indicated that improvements in cognitive capacity typically require prolonged and systematic training (Melby-Lervag & Hulme, 2013). Therefore, future studies could consider extending the duration of intervention to examine whether long-term use of AI-based language learning applications can lead to more substantial and sustainable cognitive gains, thereby providing a more comprehensive understanding of the true potential of AI interventions in supporting children’s cognitive development.

Beyond WM, ToM is also employed as a measure of children's cognitive abilities. The findings of this study indicate that while children's ToM performance improved over time; however, this improvement was not significantly associated with the use of Zebra AI. This suggests that the gains were likely attributable to developmental maturation or test-retest practice effects, rather than a specific benefit derived from the AI intervention.

To explain this lack of differentiation, it is necessary to examine the mechanistic alignment between the application's design features and the cognitive processes required for ToM. Research suggests that ToM development relies heavily on 'social contingency', the non-linear, unpredictable responsiveness of a social partner, and the interpretation of epistemic uncertainty (e.g., inferring hidden intentions or false beliefs) (Wellman, 2014).

In contrast to these requirements, the interaction design of Zebra AI is predominantly deterministic and behaviourist. For instance, in the 'listen-and-select' tasks, the system’s feedback loops are explicit and binary: a correct touch input invariably triggers a pre-programmed visual reward (e.g., animations of three stars), whilst an incorrect input prompts a fixed corrective cue (e.g., animations of one star or try again). This logic follows a strict 'If A, then B' algorithm, lacking the ambiguity found in naturalistic social exchanges. Consequently, the AI agent functions as a responsive interface rather than an intentional social partner. Because the information flow between the learner and the digital agent is transparent and rule-based, the application likely did not necessitate mental state attribution. Children were not required to infer why the character acted or resolve any informational discrepancies (i.e., 'I know something the character does not'). Without scenarios that simulate social friction or require perspective-taking, the specific neural networks associated with social cognition may not have been exercised beyond the baseline level of the control group.

This highlights a critical distinction in the 'feature-outcome' mapping of AI interventions. While algorithmic adaptivity and multimodal synchronicity effectively target information processing efficiency (benefiting WM), they do not yet sufficiently simulate the pragmatic complexities of human interaction required to drive social-cognitive development. The digital agents in the current application function as responsive tools rather than intentional agents, thereby limiting their capacity to scaffold ToM. Therefore, future research aiming to support ToM should move beyond simple gamification and incorporate features that simulate social friction or perspective-taking scenarios, such as interactive narratives where learners must infer a character's incorrect belief to solve a problem.

The regression analysis revealed that the group variable was the most robust predictor of receptive vocabulary growth. This indicates that the AI-based intervention itself, characterised by its high-dosage multimodal input and interactive feedback, was the primary driver of lexical acquisition, outweighing individual differences in baseline cognitive abilities. This confirms the efficacy of the content effect discussed earlier: exposure to the systematic, gamified curriculum provided by Zebra AI offered a significant learning advantage over the control condition.

In terms of cognitive mechanisms, the growth in WM emerged as a positive predictor of receptive vocabulary growth. This finding can be explained by the fact that early second language vocabulary acquisition is essentially a task of arbitrary mapping, which involves creating a stable link between a novel phonological form (the English word) and its visual referent (the object). This process is structurally identical to the cognitive demands of the WM assessment used in this study. Both tasks rely heavily on the Episodic Buffer in Baddeley's (2000) model to execute cross-modal binding, integrating information from the phonological loop and the visuo-spatial sketchpad. This finding corroborates existing literature linking WM capacity to language outcomes (Baddeley, 2003; Linck et al., 2014). Reinforcing this view, Monnier et al. (2022) reported a significant correlation between WM and language comprehension, suggesting that enhanced WM abilities allow children to allocate cognitive resources more efficiently during language tasks, thereby facilitating superior performance on vocabulary assessments.

In contrast, gains in ToM did not significantly predict receptive vocabulary growth.
This finding appears to diverge from established literature, such as Hale and Tager-Flusberg (2003), who conducted a quantitative analysis of the relationship between children's social cognition and language abilities, further demonstrated that children with higher ToM skills tend to exhibit greater advantages in language processing and communication. While these studies posit that interpreting others' intentions is crucial for language comprehension, the current study suggests that this relationship is context-dependent. Specifically, the lack of prediction can be primarily attributed to the nature of the learning task. The target vocabulary in this intervention consisted predominantly of concrete nouns (e.g., animals, fruits), the acquisition of which relies heavily on associative mapping rather than social-pragmatic inference. Unlike abstract verbs or social terms, learning these concrete labels involves direct object-word pairing and does not necessarily require mental state attribution (Bloom, 2000). In addition, Zebra AI provides a transparent, error-free learning environment where the intent is explicitly coded into the audiovisual synchronicity as discussed. Consequently, the children's social-cognitive abilities were effectively bypassed during the vocabulary uptake process.

This finding offers a critical reference for the design of similar AI-based learning tools. It suggests that while deterministic, drill-based applications are highly effective for rote receptive vocabulary expansion, they may fail to leverage the learner's broader social-cognitive resources. Future applications should move beyond simple associations and incorporate interactive narratives requiring intent inference to fully activate the reciprocal relationship between social cognition and language learning.

Furthermore, age was confirmed as a significant positive predictor of vocabulary growth, reflecting the well-established role of developmental maturation in SLA. This finding is consistent with the research of Wells and Wells (1985), which suggests that as children grow older, their vocabulary size and linguistic comprehension naturally improve due to increasing cognitive maturity. The advantage observed in older children may stem from their more advanced general cognitive processing speeds and learning strategies. In an AI-assisted environment like Zebra AI, which requires sustained attention and independent interaction, older children are typically better equipped to extract information from digital stimuli and navigate the gamified interface efficiently. Unlike the findings of Al-Jarf (2021), iPads were found to be more effective in facilitating language learning among younger children compared to older children in grades 1–3 and 4–6. This study suggests that within the 5–7 age range, the cognitive readiness of older children allows them to leverage technological interventions more effectively.

Despite its significance, the predictive power of age may have been constrained by the relatively narrow age range (5–7 years) of the current sample. A broader developmental span might have revealed a clearer, perhaps non-linear, trajectory of age-related gains (Al-Jarf, 2021). Additionally, as noted by Dörnyei (1998), individual differences such as home language environment and learning motivation may overshadow age as children grow. Future research should employ longitudinal designs with broader age cohorts to explore the interaction between age and AI interventions more comprehensively.

6.2.2 Which functions and design interface influence children’s hedonic experience with the AI-based language learning app?

In the process of SLA, children's enjoyment plays a crucial role in stimulating their interest in learning and maintaining their motivation. In other words, when children experience fun and pleasant emotions during learning, they are more likely to remain engaged (Filgona et al., 2020; Isen & Reeve, 2005). Existing research has shown that positive emotional experiences can significantly enhance overall learning outcomes and are especially important in SLA (Dörnyei, 1998; Gardner, 2007; MacIntyre, 2003).

The survey results from this study strongly suggest that Zebra AI successfully integrated high HQ into its design. 88% of participants reported feeling happy during use, while 86% found the user experience exciting. These indicators confirm that the application went beyond mere usability to create a positive affective state. According to Isen and Reeve (2005), such positive affect does not merely make the task pleasant; it broadens cognitive resources and facilitates the flexibility required for language encoding.

From the perspective of application design, the findings reveal a high level of PeoU, with 86% of the participants find Zebra AI very easy to use, with only 5% holding a neutral attitude and none reporting any difficulty in usage. As posited by the TAM (Davis et al., 1989), higher PEoU is associated with more favourable attitudes and stronger intentions to continue use. For young learners, ease of use is not merely a usability advantage but a critical pedagogical factor. Intuitive navigation, minimal operational demands, and clear interaction flows reduce extraneous cognitive load, allowing children to allocate more attentional resources to linguistic input rather than interface management. As highlighted by Faudzi et al. (2024), high usability can significantly reduce the user's cognitive load. In the context of early language learning, this reduction in cognitive burden is particularly important given children's still-developing executive and self-regulatory capacities.

In terms of visual presentation, 93% of the participants (N = 40) provided positive feedback on the quality of images and videos, further demonstrating the significant role of high-quality visual elements in enhancing user experience. In the specific context of Zebra AI, this high approval rating highlights the effectiveness of its rich multimedia features, such as animated narratives and dynamic character interactions.

From a theoretical standpoint, this result supports Mayer's Cognitive Theory of Multimedia Learning (Mayer, 2009), which posits that learners process information more effectively when visual and verbal channels are engaged simultaneously. The high-quality visual elements in Zebra AI do not merely serve as aesthetic decoration; they function as cognitive scaffolds that make abstract linguistic concepts more concrete and comprehensible for young learners. Beyond supporting cognitive processing, such superior visual design increases the attractiveness of the interface and forsters an intuitive and enjoyable interactive experience (Blair-Early & Zender, 2008; Shneiderman, 2004).

From the perspective of specific application functionalities, the data from this study clearly indicates that the game feature has the greatest impact on children's hedonism. Specifically, the in-app game feature not only received a 90% positive rating but also accounted for 28% of the children's favourite functions. This preference is strongly linked to how Zebra AI's specific gamified designs align with SDT (Ryan & Deci, 2000). For instance, the application employs instant AI-based pronunciation scoring and a star reward system immediately following interactive quizzes. Theoretically, these feedback loops fulfill the child's psychological need for competence. By providing real-time, tangible validation of their performance, the app reinforces the child's sense of mastery.

This satisfaction of psychological needs transforms language practice from a repetitive drill into an intrinsically motivated activity. Similar findings have been supported by existing literature. Boyle et al. (2012) pointed out that digital games possess a high level of attractiveness and interactivity, which can stimulate students' enthusiasm for participation and creativity, while research by Li and Chu (2021) demonstrated that gamified elements in educational settings can significantly enhance user engagement and satisfaction, findings that are well consistent with the positive feedback observed in this study.

This quantitative finding is strongly supported by qualitative data. In in-depth interviews with parents, many indicated that game elements significantly increase their children's interest in learning, making them more engaged and proactive during the learning process. Huang et al. (2019) similarly noted that games not only serve as effective educational tools but also enhance students' cognitive engagement through their entertainment value. Moreover, the empirical study by Chapman and Rich (2018) showed that learning environments incorporating gamified design can effectively boost students' learning motivation and outcomes. These studies provide solid theoretical and empirical support for the conclusions of this study, further confirming the important role of the game feature in enhancing children's hedonism and learning motivation.

In addition to the game feature, other application functionalities have also played an important role in attracting children. The data show that the simulated video call feature received a 23% preference rate, making it the second most popular function after games.

This result indicates that simulated video calling provides children with an experience of real interaction with the outside world, making them feel as if they are in an authentic cross-cultural communication environment, thereby stimulating their interest and confidence in second language communication. Research indicates that interaction and the simulation of authentic contexts have a significant positive effect on language acquisition (Y. L. Chen et al., 2022; Huang et al., 2022). This feature helps children obtain more vivid and intuitive language input by simulating real communication scenarios.

In addition, animated content (16%), cartoon characters (12%), and the English song feature (12%) have also played a positive role in enhancing children's hedonism. The animated content, with its vivid, lively, and creative visual presentation, can transform abstract language knowledge into intuitive and concrete information, making the learning process more relaxed and enjoyable. In addition, cartoon characters, through their unique appearances and distinctive personalities, help establish an emotional connection with children, thereby increasing the interest and affinity of the learning context (Dalacosta et al., 2009). Furthermore, music not only assists children in remembering language information more easily through its rhythm and melody but also creates an enjoyable learning atmosphere, further enhancing their emotional experience and learning interest. Multiple studies have shown that music and songs significantly facilitate language learning, improving both phonetic discrimination and language intuition (Chou, 2014; Millington, 2011).

The high levels of enjoyment observed in the study correlated directly with strong behavioral intentions. Regarding usage intention (Q4, Q5), 84% of participants expressed a willingness to use the application again, while 79% indicated a desire for long-term use. These data suggest that most children not only show a high willingness to use the application but also maintain a positive attitude toward continued learning. According to the TAM (Davis et al., 1989), this 'intention to use' serves as the key determinant to actual usage behavior. This finding is consistent with the regression analysis presented earlier, where children’s hedonic experience explained approximately 56.7% of the variance in their behavioural adherence. Consequently, it suggests that for young learners, the intrinsic enjoyment derived from the app is not merely a byproduct but the primary driver of their sustained engagement.

However, to fully understand the nuances of adherence, it is essential to examine elements of the user experience that may create barriers to sustaining engagement. Although the overall feedback was positive, the ASR-based pronunciation feedback (Q9) obtained a relatively weaker emotional response from children. While 72% of participants provided positive feedback, neutral and negative responses accounted for 24%, which was the highest proportion among all functions. This suggests that this feature may not fully meet children's emotional expectations in terms of ease of operation and prompt feedback, thereby impacting the overall pleasantness of the experience.

This specific finding offers a critical perspective for explaining the 'unstable adherence' pattern observed in the qualitative data. While the gamified elements provided motivation, technical limitations in the ASR feedback likely led to moments of frustration. For children with lower tolerance for frustration, these technical barriers disrupt the 'Flow' state (Csikszentmihalyi, 1990), causing a shift from enjoyment to anxiety or boredom. This mirrors the qualitative observation where parents of unstable users cited 'frustration with task difficulty' as a reason for discontinuation. Therefore, while high HQ drives the intention to use, the continuity of actual usage is facilitated by the system's ability to balance challenge with technical responsiveness.

In summary, the study confirms that Zebra AI's high HQ and PEoU effectively minimised cognitive load and sustained engagement. Gamified elements emerged as the primary driver of intrinsic motivation by satisfying children's need for competence (SDT), while multimedia features supported dual-coding processing (MLT). Crucially, this positive affect was the strongest predictor of sustained usage, explaining 56.7% of the variance in adherence (TAM). However, technical limitations in ASR feedback introduced friction that disrupted the 'Flow' state for some learners. This specific barrier provides a critical explanation for the 'unstable adherence' patterns observed, highlighting that long-term efficacy depends on balancing gamified motivation with technical responsiveness.

6.2.3 What are the parents' views regarding the effectiveness of using AI-based learning apps for learning a second language?

The third research question of this study aims to assess parents' perceptions of children's use of AI learning applications for SLA. To explore this issue, the study collected in-depth insights from parents through interviews and analysed the data using thematic analysis, identifying the following four? key themes: 1) Learning outcomes after children use AI learning apps, 2) Parental motivation for using AI learning apps, 3) Potential barriers to using AI apps for learning, 4) Future suggestions of AI learning app enhancements. Each of these themes is discussed in detail in the following sections.

6.2.3.1 Learning Outcomes after children use AI learning apps

As part of this study's exploration of parents' perceptions regarding AI learning applications, a key theme that emerged from the interviews was the perceived learning outcomes associated with children's use of AI learning apps. Understanding these learning outcomes is crucial for evaluating the effectiveness of AI-assisted language learning and its potential to shape children's SLA experiences.

Interview data suggest that parents perceived AI learning applications as having a substantial positive influence on their children's English language development. These perceived benefits primarily manifest in three aspects: (i) Knowledge Acquisition, where children demonstrate improvements in vocabulary and comprehension; (ii) Learner Attitudes and Motivation, encompassing their increased engagement and enthusiasm for learning; and (iii) Behavioural Changes, including active engagement with learning, increased self-directed learning, and interest in culture.

Parental feedback indicates that Knowledge Acquisition is one of the most noticeable learning outcomes brought by AI learning applications. Many parents observed that their children's vocabulary increased after using AI applications, a finding consistent with qualitative research results, as discussed in section 6.2.1. Furthermore, parents noted that the growth in vocabulary was not only reflected at the linguistic level but also had a positive impact on their children's academic performance. This perspective is supported by many previous studies (Fidan & Tuncel, 2019). Wei (2023) highlighted in interviews with 14 students that AI-powered adaptive learning platforms, by providing sustained and tailored instruction, significantly enhance engagement and motivation, ultimately leading to improved academic outcomes. Approximately half of the parents observed noticeable progress in their children's English pronunciation, particularly in phonetic accuracy and fluency. This observation is consistent the findings of Ganapathy and Seetharam (2016), who demonstrated that multimodal teaching strategies enhance pronunciation by integrating auditory and visual inputs. Building on this, Dennis (2024) highlighted that AI-driven pronunciation correction and contextualised exposure to target language input contribute to more precise phonetic articulation and smoother speech production, reinforcing the role of AI in SLA.

Most parents observed noticeable improvements in their children's interest and motivation in learning English, with two thirds reporting that the AI learning application enhanced their child’s enthusiasm for language learning. This increase in motivation can be interpreted through the lens of SDT (Ryan & Deci, 2000), which posits that intrinsic motivation is strengthened when learners experience autonomy, competence, and relatedness. AI learning applications facilitate these psychological needs by offering gamified tasks, interactive elements, and real-time feedback, thereby fostering greater engagement and sustained motivation (Dehghanzadeh et al., 2021). However, it is worth considering whether this increased motivation reflects long-term learning engagement or is driven by the novelty effect of digital learning tools. Future research should explore the sustainability of AI-driven motivation over time.

Additionally, some parents specifically mentioned in the interviews that the application was particularly effective for younger children, as cartoon characters, animations, and interactive tasks greatly captured their attention. This phenomenon can be explained through Cognitive Development Theory (Piaget, 1954), which suggests that children between the ages of 5 and 8 are typically in the preoperational stage of development. At this stage, children learn most effectively through visual, symbolic, and hands-on interactions, rather than through abstract reasoning. Therefore, AI applications integrate learning contexts with animations, making the language learning process more aligned with children's cognitive characteristics (Hsu et al., 2021). However, some studies present different perspectives on AI's role in enhancing children's motivation in language learning. For instance, Griffith and Arnold (2019) found that in the absence of appropriate parental or teacher guidance, some children perceive digital learning applications purely as gaming tools rather than educational tools, leading to decreased focus on language learning objectives.

Furthermore, AI learning applications not only stimulate children's enthusiasm for learning but also enhance their confidence in learning and using English to some extent. According to parental feedback, some children who previously hesitated to express themselves in English in class have become more confident and actively participated in classroom discussions after using AI-based learning applications. This phenomenon can be explained by Self-Efficacy Theory (Bandura, 1977) which proposes that when individuals continuously experience success and positive feedback in a given task, their belief in their ability to accomplish that task is strengthened.

AI applications provide a low-risk learning environment for children through continuous practice and positive feedback mechanisms, such as ASR-based pronunciation feedback and gamified reward systems, allowing them to obtain positive reinforcement through repeated practice (Fu et al., 2020). However, research findings on the effect of AI learning applications in boosting children's confidence are not entirely consistent. For instance, Jeon (2024) noted that not all children can experience positive outcomes from AI feedback systems, some may feel frustrated by frequent ASR-based pronunciation feedback, thereby negatively affecting their learning confidence. Additionally, Ali et al. (2020) emphasise that building confidence depends not only on learners' personal experiences but also on factors such as classroom atmosphere, teacher support, and peer interaction. Therefore, although AI learning applications can enhance children's confidence to some extent, in the absence of emotional support from teachers or parents, AI learning tools may not fully substitute the critical role of interpersonal interaction in the development of confidence (Engwall & Lopes, 2022).

Many parents have reported positive changes in their children's learning behaviours. The interactive features of AI learning tools have enhanced children's engagement in learning and fostered greater autonomy, thereby reducing their reliance on external supervision. Specifically, a quarter of the parents observed a significant improvement in their children's focus and participation after using AI learning applications. This behavioural change is consistent with Self-Regulated Learning Theory (Bandura, 1991), which emphasizes learners' ability to plan, monitor, and evaluate their learning processes independently. AI learning applications create a self-regulated learning environment through features such as immediate feedback, gamified challenges, and adaptive learning, enabling children to depend less on external supervision and gradually develop independent learning skills. Bains et al. (2022) research also supports this theory, demonstrating that technology-based interactive learning environments can effectively enhance students' engagement and self-management abilities. Additionally, Liu (2015) noted in his research on language learning motivation that enhancing learners' autonomy and intrinsic motivation is key to improving learning outcomes.

This study also found that AI learning applications play a positive role in fostering children's self-regulated learning habits. Several parents mentioned that while their children previously needed reminders to study, they are now able to independently schedule their learning time. This finding is consistent SDT (Deci & Ryan, 2008), which posits that when a learning environment satisfies learners' needs for autonomy, competence, and relatedness, their intrinsic motivation is significantly enhanced. AI learning applications facilitate this process through adaptive feedback, personalised learning paths, and progress tracking, allowing children to experience a sense of achievement, thereby increasing their motivation for self-directed learning.

Beyond promoting academic learning, this study also found that AI learning applications stimulate children's interest in cultural learning. Parents reported that their children learned about holidays and customs in English-speaking countries through AI applications, expressed a willingness to communicate with foreigners, and even began discussing future study abroad plans. This finding is consistent with Byram's Intercultural Communicative Competence Model (1997), which emphasises that language learning is inseparable from cultural understanding and that language education fosters intercultural awareness. AI learning applications incorporate cultural storytelling, interactive teaching, and diverse language environments, helping children develop a deeper understanding and appreciation of cultural diversity. Furthermore, Alshenqeeti (2016) highlighted that integrating cultural content into language learning not only enhances linguistic proficiency but also promotes learners' openness and global awareness.

6.2.3.2 Parent-caregivers' Motivation for using AI-based learning applications

The findings of this study indicate that parents' decisions to choose AI learning applications are influenced by multiple factors. Interview analysis revealed three core motivations: the flexibility of learning anytime and anywhere, the cost-effectiveness of the apps, and their potential to reduce parental educational stress. These factors played a crucial role in parents' decision-making processes, making them more inclined to encourage their children to use AI tools for English learning.

This study found that around a third of parents believed that AI learning applications greatly enhanced the convenience and accessibility of children's learning. Some parents mentioned that their children could use smartphones to learn English at home anytime, making efficient use of fragmented time while accessing a wealth of learning resources (Interviewees 2 and 7). This finding is consistent with Park (2011), who argued that mobile learning enables learners to study anytime and anywhere, freeing them from the constraints of traditional classroom environments. Similarly, Liu and Chen (2023) emphasised that the constant availability of AI learning applications allows children to take greater control over their learning pace, enhancing flexibility in the learning process.

In addition to saving time, the cost-effectiveness of AI learning applications is also a key factor influencing parents' choices. The study found that compared to traditional one-on-one English tutoring sessions, AI language learning applications are more affordable while also saving both time and money (Interviewee 7). This finding is consistent with the study by D. Wang et al. (2022), which found that following the implementation of the "Double Reduction" policy, an increasing number of families have turned to digital learning resources as a supplement to compensate for the reduction in off campus tutoring programs. This shift is particularly significant in the context of China's 2021 “Double Reduction” policy. In this context, technology- based solutions such as AI-powered learning apps have gained increasing attention as potentially effective tools to support self-regulated learning at home while aligning with the policy’s goals of reducing burden and improving efficiency. Additionally, Kormos and Wisdom (2021) highlighted that the cost-effectiveness of digital learning tools is particularly crucial for low-income families, as these applications provide a low-cost yet effective approach to language learning. This perspective is consistent with the views expressed by rural parents in the interviews.

This study found that around one-third of parents perceived the speech recognition feature in AI language learning applications as playing a significant role in reducing their tutoring burden. In home language learning environments, parents often take on the role of guiding their children's pronunciation and correcting their speech errors. However, for parents with limited English proficiency, this task can lead to additional anxiety and uncertainty. Interview data from this study revealed that many parents worried about the accuracy of their own pronunciation or their inability to effectively identify and correct their children's mistakes, which in turn impacted their children's language learning outcomes (Interviewees 1, 5, 8). The ASR-based pronunciation feedback in AI-based language learning applications provides an automated support system through real-time pronunciation correction and feedback mechanisms. This feature automatically detects and corrects children's pronunciation errors, allowing them to receive instant and precise pronunciation feedback during independent learning. By reducing the need for direct parental intervention in pronunciation instruction, AI applications enable parents to transition from the role of direct instructors to facilitators and supporters, enhancing their confidence and engagement in their children's language learning process. While existing literature primarily focuses on AI's role in developing learners' language proficiency, its impact on parental stress, family education models, and the shifting role of parents in children's learning remains an area that warrants further exploration. Through parental perspectives and data analysis, this study provides new evidence demonstrating that AI language learning applications not only positively influence children's language development but also serve as potential support tools within home learning environments.

6.2.3.3 Potential barriers to using AI-based applications for learning

Parents have identified a range of issues and challenges that children face when using AI-driven learning applications. These challenges not only affect children's learning outcomes but also have the potential to negatively impact their learning motivation. Specifically, the difficulties encompass both the direct challenges children encounter while using the app and the obstacles parents face in supervising and supporting their children. The following section discusses these issues in detail.

Regarding children's use of the application, some parents reported that their children experienced difficulties in maintaining attention while using the app. This concern is supported by Başar and Elyıldırım (2021), who argue that smartphones, as multifunctional devices, contain various entertainment applications (such as games, videos, and social media) that are highly appealing to children. This multitasking environment makes it difficult for children to focus solely on a learning application. Research has shown that children's attention is easily diverted by other stimuli on the device, particularly in the absence of parental supervision (Girela-Serrano et al., 2024). As a result, the multifunctionality of smartphones itself becomes a potential interference factor affecting the effectiveness of learning tools. Theopilus et al. (2024) further highlight that addiction to the internet, smartphones, or video games can create various developmental barriers that hinder children's development.

Some parents have expressed concerns about the difficulty level of AI-driven learning applications. One parent noted that certain learning materials were too complex for children with weaker English foundations, making it hard for them to keep up and leading to frustration (Interviewee 7). This is consistent with the findings of Furió et al. (2015), who demonstrated that the difficulty level of learning content directly influences learners' motivation. When tasks are excessively challenging, children's interest in learning may decline. Conversely, appropriate difficulty adjustments can enhance their learning experience. These insights suggest that AI learning applications should offer more precise personalised learning pathways, ensuring that children learn within their ZPD (Vygotsky, 1978) to sustain their motivation.

Furthermore, around a quarter of parents expressed concerns about the complexity of AI learning applications' interfaces, particularly those with younger children (Interviewees 2 and 5). One parent mentioned that their child initially felt confused when navigating the application and required step-by-step guidance to use it effectively (Interviewee 2). This finding is consistent with Masood and Thigambaram (2015) research, which highlights that a clear and intuitive interface design is crucial for enhancing young children's engagement and learning efficiency. A complex interface may disrupt the learning flow, creating unnecessary obstacles. Therefore, AI learning applications should adhere to child-friendly design principles, refining interactive logic and simplifying navigation to lower the barriers to learning (G. Wang et al., 2022).

Additionally, one parent raised concerns about the content of AI learning applications, specifically the presence of "antagonist" characters or intense animated scenes, which might confuse children or even lead them to imitate negative behaviours (Interviewee 2). Children's emotional responses to digital learning content are closely linked to Social Learning Theory (Bandura, 1977), which suggests that children learn behaviours by observing and imitating what they see in media. As a result, AI learning applications should place greater emphasis on age-appropriate content, avoiding elements that could encourage undesirable behaviours while ensuring that the educational material aligns with children's cognitive and emotional development needs (G. Wang et al., 2022).

Almost half of parents in this study worried that prolonged use of AI learning applications may negatively affect their children's vision. As a preventive measure, they have implemented restrictions on screen time to mitigate potential vision-related issues (Interviewees 3 and 4). These concerns are well-founded, as existing research suggests that excessive screen exposure can have adverse effects on children's visual health and overall well-being (Canadian Paediatric Society & Digital Health Task Force, 2017; Stiglic & Viner, 2019). For instance, Yang et al. (2020) emphasise that prolonged screen time can lead to eye strain, an increased risk of myopia, and reduced attention span. Moreover, excessive screen exposure has been linked to disruptions in sleep quality, which may indirectly impact children's learning performance.

During the interviews, technological issues, particularly device performance limitations and network stability emerged as one of the most frequently mentioned barriers among rural parents. Many reported that AI learning applications require high-performance devices, making older devices prone to lagging or crashing, which significantly disrupts children's learning experiences (Interviewees 7 and 8). In addition, unstable internet connections further hinder children's ability to engage in continuous learning. These findings align with the study by Sun et al. (2008), which indicates that device performance and network stability are critical factors influencing the effectiveness of online learning. For users in rural areas, technological barriers can lead to frequent learning interruptions and even weaken student's motivation to continue learning (Salemink et al., 2017).

While many AI learning applications offer free versions, some parents expressed concerns about the costs associated with premium features or subscription fees. Some of them also believed that traditional tutoring may provide a better return on investment compared to AI-based learning applications. This phenomenon is consistent with the findings of Li and Lin (2022), who noted that many parents still prefer traditional offline tutoring because face-to-face instruction provides more direct teacher guidance and support.

A small proportion of parents reported that supervising their children's use of AI learning applications requires a significant amount of time, which may lead to increased strain on their daily schedules (Interviewee 1). This phenomenon is consistent with the high parental investment model in Chinese education. Mu and Hu (2023) found that Chinese parents typically bear high educational expectations and responsibilities, devoting substantial time and effort to extracurricular tutoring and learning supervision to ensure their children gain a competitive advantage in an increasingly rigorous educational environment. Additionally, Zheng et al. (2020) highlighted that Chinese parents not only focus on their children's academic performance but also actively participate in extracurricular tutoring, homework supervision, and learning planning, which places significant pressure on their time management.

Some parents still prefer traditional educational methods and view AI learning applications as supplementary tools rather than replacements for formal schooling. Additionally, some parents indicated that their decision to continue using AI learning applications largely depends on teachers' recommendations. If the school or teachers do not endorse these applications, they are more likely to discontinue their use (Interviewee 1). This perspective is consistent with the findings of Aruta et al. (2019), who noted that in Asian cultural contexts, parents tend to place greater trust in teachers' authority, making school endorsement a crucial factor in the technology acceptance process. Furthermore, Chan and Tsi (2024) emphasised that teachers have distinct attributes, including critical thinking and emotional sensitivity, which cannot be fully replaced by AI technology. This suggests that the future promotion of AI learning applications should focus on deep integration with the school education system. Possible strategies include teacher training programs, school-endorsed recommendations, and parental education initiatives to enhance parents' trust and acceptance of AI-assisted learning.

For most rural parents in this study, lack of familiarity with technology was a significant barrier. Some rural parents admitted that they were not well acquainted with modern technology, making it difficult for them to support their children in using AI learning applications at home (Interviewees 6 and 8). This phenomenon is particularly prevalent in rural areas of China, as existing research has indicated that the digital divide in rural households affects children's access to digital educational resources (Wang et al., 2021). Studies have shown that despite the Chinese government's ongoing efforts to promote digitalisation in rural areas, rural families still lag behind urban households in technology usage skills, digital literacy, and internet infrastructure (McQuaide, 2009). This issue is especially pronounced among caregivers, who often lack experience in using smart devices (Zhao & Chen, 2023).

Future research could explore how to provide effective technology training for rural parents, helping them become familiar with AI learning applications and improving their ability to support their children in digital learning environments. Additionally, at the policy level, research should investigate how to promote digital education equity in rural areas, including ways in which the government can invest in improving internet infrastructure, reducing technological costs, and making AI-based educational tools more accessible to rural families.

6.2.3.4 Future suggestions of AI learning app enhancements

This study, through an in-depth analysis of parental feedback, found that when optimising AI learning applications, parents generally focus on two key areas: functional enhancement and technological adaptability. The following discussion will elaborate on these two core areas in detail.

Firstly, parents hope to stimulate their children's interest in learning by adding social interaction and immersive learning experiences. For example, one parent suggested introducing an "intelligent reading partner" feature, where an AI virtual assistant not only accompanies the child in learning English but also interacts with them through games, thereby creating a more supportive and interactive learning atmosphere

(Interviewee 5). Research shows that peer interaction can significantly enhance learning outcomes by providing emotional support and fostering a sense of belonging, which in turn motivates learners (Kiefer et al., 2015; Korpershoek et al., 2020).

Secondly, enhancing personalised learning features is another key focus for parents. Some parents have proposed that AI learning applications should automatically adjust the difficulty of content based on the child’s learning ability and preferences, thus better meeting individual needs (Interviewee 3). Related studies have found that AI-based adaptive learning systems can dynamically adjust learning content to suit learners at different levels, avoiding frustration from overly challenging material while still providing sufficient challenges to maintain interest (Cui et al., 2018). Personalised learning not only supports more efficient knowledge acquisition but also promotes learner autonomy, motivation, and sustained engagement over time.

In addition, parents particularly emphasised the need to improve the usability of these applications, especially in terms of user guidance and learning progress monitoring. One parent suggested that the application should provide clear learning plans, progress tracking, and reminder notifications so that parents can effectively supervise their children’s learning even amid a busy work schedule (Interviewee 2). Meanwhile, some parents expressed a desire to receive regular updates on their child’s learning content and achievements, which would help them comprehensively understand their child’s progress. Empirical research indicates that parental involvement is a critical factor influencing children’s learning outcomes in technology-assisted learning environments (Matban, 2023). Therefore, improving the parent monitoring system and communication features (such as providing progress reports and AI-generated personalised feedback) will not only help parents better support their children’s learning but also enhance their trust and acceptance of AI learning applications.

Furthermore, some parents have suggested that AI learning applications be integrated with traditional classroom teaching to form a blended learning model. One parent pointed out that traditional teaching methods can sometimes be monotonous, whereas the inclusion of AI tools can enhance classroom interactivity and engagement while still retaining the teacher’s central educational role (Interviewee 5). Research has demonstrated that blended synchronous learning models can provide students with broader learning opportunities and a richer learning experience. This approach is particularly suitable for remote learners, as it helps improve academic performance and efficiency while enabling a better integration of AI-assisted learning content with classroom instruction (Wang et al., 2017; Xiao & Jiang, 2023).

Finally, some parents expressed optimism about the future development of AI, with one parent even suggesting that AI might eventually evolve to a level where it could replace human teachers (Interviewee 1). However, existing research indicates that although AI shows tremendous potential in adaptive learning and automated feedback, it is currently unable to substitute for teachers in terms of providing emotional support, managing classrooms, and delivering personalised guidance (Selwyn, 2019). Therefore, future AI language learning tools are more likely to augment rather than replace teachers, complementing traditional educational models to jointly enhance the quality of education.

Research indicates that in areas with unstable internet connectivity, most rural parents, approximately two-thirds prefer AI learning applications to include offline functionality, ensuring that children can continue learning even without internet access (Interviewee 8). This finding is consistent with Munoto et al. (2021), who emphasised the critical role of offline educational technology in ensuring learning accessibility in resource-limited regions. Additionally, some parents suggested that AI learning applications should support local dialects to enhance linguistic inclusivity, enabling both parents and children to engage more seamlessly in the learning process (Interviewee 7). This perspective is supported by Ahia et al. (2024), whose research highlights the importance of multilingual and dialect-adaptive features in promoting educational equity and accessibility. Moreover, parents generally perceive themselves as playing a crucial role in their children's learning process and expect AI learning applications to provide more resources specifically designed to support their involvement and guidance. Based on these findings, future research should further explore the following areas: (1) the development of efficient and stable offline learning models to ensure continuity in low- connectivity environments; (2) the optimisation of AI systems to accommodate various dialects, thereby improving linguistic inclusivity and learning accessibility; and (3) the design of parent-oriented support systems to strengthen their role in AI-assisted learning. Investigating these directions will contribute to enhancing the applicability of AI learning applications and promoting the equity and sustainability of technology-assisted learning in diverse educational settings.

6.3 Conclusion

This chapter presented an in-depth discussion of the key findings of this study. As stated in the introduction to this chapter, this research adopted an embedded mixed-methods approach. In the quantitative study, data were collected from 85 children aged 5 to 7, who participated in two assessment tests and a questionnaire to evaluate the impact of AI learning applications on children's SLA. Additionally, the qualitative study involved semi-structured interviews with the parents of eight children in the experimental group, aiming to explore parents' perceptions of the effectiveness of AI learning applications.

This study employed a quasi-experimental design, with the intervention lasting 12 weeks. For data analysis, quantitative data were processed using statistical analysis, while qualitative data were analysed using thematic analysis. To enhance the credibility and robustness of the findings, this study implemented a triangulation approach, integrating multiple data sources to cross-validate the research findings.

In conclusion, this study provides strong evidence that AI-based learning applications, as exemplified by Zebra AI, can effectively support young children's second language vocabulary acquisition. Importantly, the observed learning gains cannot be attributed to technological novelty alone. Instead, they are explained by psychologically informed design features. Specifically, the combination of visual and auditory input supports cross-modal associative learning, while ASR-based feedback offers opportunities for correction and focused attention, both of which facilitate vocabulary development. In addition to language outcomes, a clear dose effect was identified, indicating that the effectiveness of AI-based interventions depends on sustained exposure, which in turn is more strongly driven by children's self-directed engagement than by parental enforcement.

From a theoretical perspective, the findings contribute to a clearer understanding of cognition–technology interaction in early language learning. WM was found to be a positive predictor of vocabulary growth, suggesting that repeated visual–auditory pairing and active retrieval demands within AI-based applications may indirectly strengthen WM and attentional processes, even without explicit cognitive training. In contrast, no significant effects were observed for ToM. This highlights an important pedagogical limitation, as current AI systems are largely rule-based and lack the uncertainty and intention inference that characterise genuine social interaction. As a result, while AI-based applications can effectively support perceptual and associative aspects of language learning, they can only complement, rather than replace, human-mediated social language experiences.

The study also highlights the central role of enjoyment in early childhood learning. Children’s continued engagement was shaped by hedonic quality, which helped to reduce cognitive load and support feelings of competence, thereby encouraging self-determined participation. However, technological challenges, particularly occasional limitations in ASR sensitivity, remained a key barrier. Such issues can disrupt children’s flow and reduce sustained engagement, emphasising the need for reliable technology alongside sound pedagogical design. Overall, these findings suggest that future AI-based learning tools should move beyond basic gamification and towards more robust, responsive, and socially informed designs that balance educational challenge with technical stability.

Parental views largely supported these conclusions. Most parent-caregivers reported positive attitudes towards AI-based learning applications, recognising their benefits for children’s language development and learning motivation, as well as their potential to reduce parental involvement in direct instruction. At the same time, parents raised concerns about long-term use, including issues related to technical reliability, supervision requirements, cost, and the continued importance of traditional classroom teaching. Their expectations for future development, such as more personalised learning pathways, offline access, closer alignment with school curricula, and improved parental monitoring tools, provide valuable guidance for the further development and refinement of AI-supported language learning applications.

Overall, this study provides empirical support for the application of AI technology in children's SLA and identifies key directions for optimizing AI learning applications in the future. However, further research is needed to explore long-term effects, individual differences, and technological adaptability to ensure that AI-based language learning tools can be more widely and effectively used to support children's language development.

Chapter 7. CONCLUSION

7.1 Introduction

This research systematically evaluated the effectiveness of AI-based language learning applications in supporting L2 learning among Chinese children aged 5 to 7, with a particular focus on their impact on children's English receptive vocabulary and cognitive abilities (WM and ToM), as well as the interface designs and features that influence children's hedonic user experience.

A mixed-methods approach was employed, utilising an embedded research design that integrated quantitative (quasi-experimental design and questionnaires) and qualitative (semi-structured interviews) methodologies.

In the quantitative phase, 85 children participated in pre- and post-tests to assess learning outcomes and completed a questionnaire to report on their user experience with the AI learning application. In the qualitative phase, semi-structured interviews were conducted with 8 parents or caregivers of children from the experimental group to gather in-depth insights into their perceptions of the application's effectiveness and their feedback during the intervention.

The quasi-experimental intervention lasted for 12 weeks and involved the use of Zebra AI, an AI-powered language learning app that combines ASR, adaptive learning and learning analytics. During the intervention, children in the experimental group regularly engaged with the AI application, while the control group used non-educational entertainment apps. Post-intervention, qualitative interviews explored parental views on app design, learning outcomes, and the role of the technology in home learning.

The study aimed to provide a comprehensive evaluation of the effectiveness of AI learning applications in improving children’s receptive vocabulary in a second language, explore their potential influence on cognitive development (WM and ToM), and investigate the experiences and attitudes of both children and their parents throughout the use of the application. In addition, the study drew on the TAM and the UX modle framework to examine the multidimensional mechanisms underlying children's hedonic experiences and to explore parental expectations and acceptance of AI-assisted language learning.

Grounded in a thorough literature review and empirical analysis, the study critically examined the outcomes and underlying mechanisms of the intervention. This chapter summarises and discusses the key findings of the study, structured around the following three core research questions:

RQ1: To what extent does the AI-based app improve children's receptive vocabulary, and how are these outcomes predicted by their WM and ToM?

RQ2: Which functions and design interface influence children's hedonic experience with the AI-based language learning app?

RQ3: What are parent-caregivers' views regarding the effectiveness of using AI-based learning apps for learning a second language?

In addition, this chapter will propose recommendations for optimising AI-based language learning tools based on research data and literature review, further explore the logical connections between the research questions and the findings, and offer suggestions for future research directions.

7.2 Main Research Findings

This study addressed three core research questions (RQ1–RQ3) and systematically investigated the impact of AI-based language learning applications on SLA among Chinese children aged 5 to 7, using a combination of quantitative and qualitative data.

Regarding RQ1, the study confirms that the AI-based learning application had a significant positive impact on children's receptive English vocabulary. Both quantitative and qualitative data substantiate this finding, with the experimental group achieving significantly higher vocabulary scores than the control group following the 12-week intervention. Beyond linguistic gains, the intervention also led to a significant improvement in WM. Although the application did not explicitly train WM, its design indirectly exercised WM and attentional resources by requiring continuous cross-modal binding (visual-auditory association) and active retrieval. In contrast, no significant inter-group difference was observed in ToM development. This suggests that the application's deterministic, rule-based interactions lacked the social contingency and unpredictability necessary to challenge and develop social cognition.

Regarding predictive factors, the intervention itself emerged as the most potent predictor of vocabulary growth. Among the cognitive variables, WM was identified as a significant positive predictor, suggesting that lexical acquisition in this context shares a reliance on cross-modal binding mechanisms with WM. In contrast, ToM did not yield significant predictive power for vocabulary outcomes, implying that the acquisition of concrete nouns in this AI-supported environment may rely more heavily on associative mechanisms than on complex social-intention inference. Additionally, age was confirmed as a significant predictor, with older children in the sample generally demonstrating higher learning gains.

Regarding RQ2, the study found that AI-based language learning applications significantly enhanced children's hedonic experience during the second language learning process. Questionnaire results showed that most children expressed positive emotional responses, such as enjoyment, excitement, and high engagement when using the Zebra AI application. Further analysis identified several key design features of the app, such as ease of use, high-quality visual presentation, and gamified interactive mechanisms as critical in stimulating children's emotional engagement and learning motivation. Among these, the game function was most favoured by both children and parents, serving as a central driver of emotional investment and learning enthusiasm. Additional features such as simulated video calls, animated characters, and English songs also enhanced the sense of fun and immersion, making the language learning process more engaging and approachable. These findings were further supported by qualitative data from parent interviews. Most parents noted that the app's engaging, and interactive design effectively increased their children's interest and initiative in learning. Some parents also highlighted that gamified tasks and virtual characters significantly enhanced their children's anticipation and willingness to participate in English learning activities. Nevertheless, the study also found individual differences in children's experiences, with some showing limited interest or difficulty maintaining attention. This indicates a need for future AI learning tools to better accommodate developmental and cognitive differences among users.

For RQ3, the analysis of parent-caregiver' interview data revealed three main types of positive changes in children after using AI language learning applications. First, in terms of knowledge acquisition, children demonstrated improvements in vocabulary, pronunciation accuracy, and overall language expression, with some showing increased learning engagement both at home and in school. Second, children's learning motivation and attitudes were enhanced, as evidenced by greater interest, confidence, and willingness to participate in English learning. Third, positive behavioural changes were observed, including improved focus, the development of autonomous learning habits, and a growing interest in exploring English-speaking cultures. These findings reflect the multifaceted impact of AI applications on children's cognitive, emotional, and behavioural development.

At the same time, the study identified several barriers to the effective use of AI learning tools. While most parents recognised the educational potential of such applications, practical challenges such as time constraints, financial concerns, concerns about children's eye health, and a lack of technological familiarity reduced the frequency and sustainability of their use. Some parents also reported that certain app interfaces were unintuitive, or that the learning content was too difficult, which negatively affected their children's experience. Others raised concerns about inappropriate animated characters or content that could potentially influence children's behaviour. Additionally, some children exhibited signs of distraction and low self-regulation during use, highlighting the importance of age-appropriate design and better learning environment management.

The interviews also revealed parent-caregivers' expectations for future improvements to AI learning applications. These include the desire for more adaptive personalised learning paths, offline functionality to support learning in areas with limited internet access, better integration with school curricula, and expanded features for parental monitoring and feedback. Furthermore, some parents suggested the incorporation of virtual “peer support” functions, such as AI learning companions that simulate social interactions to boost engagement and foster a sense of belonging. Parents also expressed the need for simplified operational guidance and basic digital training resources to better support their involvement in AI-assisted learning at home.

In summary, this study provides robust evidence that AI-based language learning applications can significantly enhance children’s vocabulary acquisition, learning motivation, and behavioural engagement. At the same time, it identifies ongoing challenges in areas such as social cognition, individual adaptability, and family-based support. These findings not only expand the theoretical landscape of SLA and user experience research in early childhood but also offer practical implications for developers, educators, and policymakers aiming to design and implement more effective AI-assisted language learning tools for young learners.

7.3 Limitations

While this research offers valuable insights into AI-supported language learning and its underlying cognitive processes, it is important to acknowledge certain limitations. These arise primarily from practical experimental constraints and the inherent complexity of the subject matter. To maintain a balanced perspective and inform future studies, the following sections outline the research shortcomings in three key areas: experimental design and intervention fidelity, measurement and operationalisation of constructs, and sample representativeness.

A primary limitation concerns the intervention tool itself. The study relied exclusively on a single proprietary application, the trial version of Zebra AI. While the findings demonstrate promising effects, these outcomes are necessarily specific to the design features and technological capabilities of this particular app. The category of ‘AI-based learning tools’ encompasses a wide variety of systems that differ markedly in pedagogical frameworks, feedback mechanisms, and adaptivity. Therefore, the results reported here cannot be assumed to generalise automatically to other AI applications with different algorithmic models or interaction designs.

Closely related to this is the design of the control group, which restricted the precise interpretation of the intervention mechanism. The choice of Candy Crush Saga, a non-educational game was based on feasibility and aimed at matching the experimental group on screen time and task engagement. However, this design could not effectively isolate the core technological mechanisms of the Zebra AI application (such as AI-driven immediate feedback and voice recognition) from its structured educational content. Consequently, the significant improvement in receptive single-word vocabulary breadth is more likely attributable to the system's high-quality vocabulary input rather than the independent contribution of specific AI-driven mechanisms. Future research could adopt mechanism-separation designs, comparing an 'AI-enabled' condition with a 'traditional content' condition to clarify the relative influence of content versus mechanism.

Furthermore, there is a potential risk of a novelty effect. Although the intervention spanned 12 weeks, the high hedonic ratings and rapid learning gains observed, particularly in the initial phases, might partially reflect participants' enthusiasm for the novel instructional format. Consequently, it is difficult to determine with certainty the extent to which engagement was sustained by the intervention's intrinsic pedagogical value versus the novelty of the experience. Longitudinal research with delayed post-tests is necessary to evaluate the durability of these learning gains and behavioural changes over time.

Several limitations exist regarding the instruments used to measure key constructs. Regarding language assessment, while the PPVT-5 is a robust measure of receptive vocabulary, it does not capture the full spectrum of linguistic competence, such as vocabulary depth (e.g., collocations, polysemy) or sentence-level listening comprehension. Thus, the observed gains should be interpreted strictly as improvements in single-word recognition, and caution is exercised when generalising these findings to broader communicative proficiency.

Significant caveats also apply to the cognitive measures. The image-based Animal Span task assesses associative memory rather than the broader construct of WM, which involves simultaneous storage and active executive processing. While this task was chosen to minimise executive load for young children (aged 5–7) and to capture cross-modal binding, it implies that observed gains reflect improvements in associative binding efficiency rather than a fundamental expansion of general WM capacity. Moreover, the interpretation of these cognitive outcomes is tempered by unexpectedly high baseline scores in the WM task. Factors such as prior exposure to digital games or the recruitment of a relatively advantaged urban sample likely contributed to a ceiling effect, limiting the extent of observable improvement. Future research should employ more complex span paradigms (e.g., backward digit span) or increase task difficulty to accurately capture cognitive development nuances.

Additionally, limitations pertain to the measurement of hedonic experience. The self-developed questionnaire, while grounded in the UX model and TAM, has not undergone standardised reliability and validity testing with a large sample. This issue is particularly salient in child populations, where subjective evaluations vary widely. Furthermore, young participants using Likert scales often avoid extreme response options, potentially reducing data sensitivity. Future research is advised to incorporate objective methods, such as behavioural observation or interaction log analysis, to triangulate these findings.

Finally, unmeasured covariates present a limitation. The study did not statistically control for participants' L1 ability, non-verbal intelligence (IQ), or socioeconomic status (SES). Previous research indicates these factors are predictors of language learning (Schmidt & Blumenthal, 2022; Sparks et al., 2009). The exclusion of these variables introduces the possibility of unobserved heterogeneity. Future studies should assess these characteristics using standardised measures (e.g., Raven's Matrices) to isolate the intervention's specific contribution more precisely.

The final set of limitations concerns the sampling strategy and geographical reach. First, the quantitative data were collected from two primary schools in Nanyang, China. Given China's vast regional diversity in educational resources and parental attitudes toward technology, findings from this urban context may not be fully generalisable to rural areas or less-developed provinces.

Second, the qualitative component relied on a small sample of eight parents for interviews. While offering valuable contextual insights, this limited number restricts representativeness. Moreover, the self-selection nature of the sample suggests that participating parents may have had more positive attitudes towards AI or higher involvement in their children's learning than the general population. This potential bias underscores the need for future studies to recruit larger, more stratified samples using purposive or quota sampling strategies to ensure diverse representation across demographic groups.

Despite these constraints, the consistent measurement across pre- and post-intervention phases ensures that the findings offer reliable, albeit context-specific, insights into how AI technologies support early childhood language development.

7.4 Suggestions for Further Research

Based on the findings and limitations of the current study, several avenues for future research are recommended to further explore and enhance the application of AI-based language learning tools in early childhood SLA. These suggestions focus on research design, measurement tools, intervention content, and sociocultural factors.

7.4.1 Research Design and Scope

Future studies should consider expanding the sample size and geographic coverage to improve the generalisability of findings. While this study was conducted in two primary schools in Nanyang, China, further research could include participants from diverse urban and rural settings across different regions. This would allow for exploration of how factors such as socioeconomic status may affect the AI-based interventions.

In addition, future research should extend the intervention duration beyond the current 12 weeks and adopt more rigorous longitudinal designs to evaluate the long-term sustainability of language learning outcomes. Specifically, incorporating delayed post-tests, such as an assessment three-month post-intervention, is essential to determine whether the observed benefits persist once the initial novelty has subsided. Furthermore, employing a phased exposure design where the intervention is introduced to different groups at staggered intervals would help to isolate specific treatment effects from temporal novelty factors, thereby providing more robust evidence of the intervention's true efficacy.

Researchers may also consider comparative studies that evaluate the effectiveness of AI-based learning tools against traditional instruction or alternative digital platforms. For example, contrasting the effects of Zebra AI with other commercial language apps could help identify which features are most effective for young learners.

7.4.2 Measurement and Analytical Tools

To enhance reliability and validity, future research should refine localised, age-appropriate instruments for assessing children's hedonic experience and engagement. While this study used adapted UX and TAM based questionnaires, integrating objective behavioural data (e.g., clickstreams, response times, facial emotion tracking) could provide a more nuanced understanding of user experience.

Additionally, future studies could go beyond receptive vocabulary and include assessments of expressive language, syntactic development, and oral fluency through storytelling or interactive verbal tasks. Introducing additional variables such as learning motivation, attention control, or self-regulated behaviours would help construct a more holistic view of how children engage with AI-based tools during the learning process.

7.4.3 Intervention Design and Technological Adaptation

Future research should focus on the latest advancements in the functional iterations and technological evolution of AI-based language learning applications, exploring their potential to further facilitate children’s language acquisition. With ongoing developments in AI technologies, such as natural language processing, speech recognition, and adaptive recommendation algorithms AI applications are expected to introduce new forms of interaction and learning mechanisms. For example, future studies could incorporate updated AI features (e.g., personalised emotion recognition, immersive learning experiences) into newly designed intervention programs and examine their effects on children's learning outcomes and user experience.

7.4.4 Parental Perspectives and Sociocultural Factors

Given parents' critical role in mediating children's technology use, future studies should further explore how digital literacy, educational attitudes, and family routines influence the uptake and effectiveness of AI language learning apps. Mixed-methods research combining large-scale surveys with qualitative follow-ups could yield deeper insights.

In conclusion, these future research directions aim to build upon the findings of the present study by expanding understanding of how AI-based tools support SLA in early childhood, improving technological design, and promoting equitable and sustainable learning outcomes. Integrating insights from children's cognitive and linguistic development, UX models, and parent-caregivers' feedback offers a comprehensive roadmap for advancing the field of AI-supported language education.

REFERENCES

Ahia, O., Aremu, A., Abagyan, D., Gonen, H., Adelani, D. I., Abolade, D., Smith, N. A., & Tsvetkov, Y. (2024). Voices Unheard: NLP Resources and Models for Yor\ub\'a Regional Dialects. *arXiv*. https://doi.org/https://doi.org/10.48550/arXiv.2406.19564

Aitchison, J. (2012). Words in the mind: An introduction to the mental lexicon. John Wiley & Sons.

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, *20*, 1-11. https://doi.org/10.1016/j.edurev.2016.11.002

Akgun, F. (2023). Parents' Attitudes and Opinions towards Their Children's Use of Technology. *International Journal of Research in Education and Science*, *9*(3), 597-622. https://doi.org/10.46328/ijres.3157

Al-Jarf, R. (2021). Differential effects of the iPad on first and second language acquisition by Saudi children during the Covid-19 pandemic. eLearning and Software for Education,

Alhussayen, A., Alrashed, W., & Mansor, E. I. (2015). Evaluating the user experience of playful interactive learning interfaces with children. *6th International Conference on Applied Human Factors and Ergonomics (Ahfe 2015) and the Affiliated Conferences, Ahfe 2015*, *3*, 2318-2324. https://doi.org/10.1016/j.promfg.2015.07.378

Ali, Z., Masroor, F., & Khan, T. (2020). Creating positive classroom environment for learners' motivation towards communicative competence in the english language. *Journal of the Research Society of Pakistan*, *57*(1), 317. https://pu.edu.pk/images/journal/history/PDF-FILES/24_57_1_20.pdf

Alkhudiry, R. (2022). The Contribution of Vygotsky's Sociocultural Theory in Mediating L2 Knowledge Co-Construction. *Theory and Practice in Language Studies*, *12*(10), 2117-2123. https://doi.org/10.17507/tpls.1210.19

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A.-M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*(2), 85-106. https://doi.org/10.1016/j.jecp.2003.10.002

Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International journal of teaching and education*, *III*(3), 21-34. https://doi.org/10.52950/TE.2015.3.3.002

Alshenqeeti, H. (2016). Social and Cultural Focus in EFL Learning. Does Understanding Target Language Culture Increase Interaction in the EFL Classroom? *International Journal of Linguistics*, *8*(6). https://doi.org/10.5296/ijl.v8i6.10265

Alzubi, T., Fernández, R., Flores, J., Duran, M., & Cotos, J. M. (2018). Improving the Working Memory During Early Childhood Education Through the Use of an

Interactive Gesture Game-Based Learning Approach. *IEEE Access*, *6*, 53998-54009. https://doi.org/10.1109/Access.2018.2870575

Amaral, L. A., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, *23*(1), 4-24. https://doi.org/10.1017/S0958344010000261

Arifani, Y., Asmana, C. H., & Perwita, T. (2021). Mobile-Assisted Language Learning Based Using Scientific Approach to Improve Students Vocabulary. *English Teaching, Literature, and Applied Linguistics*, *4*(1), 43-51. https://doi.org/10.30587/jetlal.v4i1.2390

Arifiati, N., Nurkhayati, E., Nurdiawati, E., Pamungkas, G., Adha, S., Purwanto, A., Julyanto, O., & Azizi, E. (2020). University students online learning system during Covid-19 pandemic: Advantages, constraints and solutions. *Systematic reviews in pharmacy*, *11*(7). https://doi.org/10.31838/srp.2020.7.81

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental psychology*, *35*(5), 1311-1320. https://doi.org/10.1037//0012-1649.35.5.1311

Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative research*, *1*(3), 385-405. https://doi.org/10.1177/146879410100100307

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315094922

Azungah, T. (2018). Qualitative research: deductive and inductive approaches to data analysis. *Qualitative research journal*, *18*(4), 383-400. https://doi.org/10.1108/Qrj-D-18-00035

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

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*(11), 417-423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory and language: an overview. *Communication Disorders*, *36*(3), 189-208. https://doi.org/10.1016/s0021-9924(03)00019-4

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bai, Y., Tejedor-García, C., Hubers, F., Cucchiarini, C., & Strik, H. (2021). *An asr-based tutor for learning to read: How to optimize feedback to first graders* Speech and Computer: 23rd International Conference,

Bains, M., Kaliski, D. Z., & Goei, K. A. (2022). Effect of self-regulated learning and technology-enhanced activities on anatomy learning, engagement, and course outcomes in a problem-based learning program. *Advances in Physiology Education*, *46*(2), 219-227. https://doi.org/10.1152/advan.00039.2021

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Gersten, R., Haymond, K., Kieffer, M. J., Linan-Thompson, S., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school (NCEE 2014-4012)* (What Works Clearinghouse, Issue. I. o. E. S. National Center for Education Evaluation and Regional Assistance (NCEE), U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications_reviews.aspx

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. https://doi.org/10.1037/0033-295x.84.2.191

Bandura, A. (1991). Social Cognitive Theory of Self-Regulation. *Organizational behavior and human decision processes*, *50*(2), 248-287. https://doi.org/10.1016/0749-5978(91)90022-L

Banić, L., & Orehovački, T. (2024). A comparison of parenting strategies in a Digital Environment: A systematic literature review. *Multimodal technologies and interaction*, *8*(4), 32. https://doi.org/10.3390/mti8040032

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language learning*, *57*(1), 35-56. https://doi.org/10.1111/j.1467-9922.2007.00398.x

Barendregt, W., Bekker, M. M., Bouwhuis, D. G., & Baauw, E. (2006). Identifying usability and fun problems in a computer game during first use and after some practice. *Human-Computer Studies*, *64*(9), 830-846. https://doi.org/10.1016/j.ijhcs.2006.03.004

Başar, T., & Elyıldırım, E. (2021). The Role of Multimedia in Concept Learning from the Parents' Perspective. *Learning and Teaching in Digital Age*, *7*(1), 16-29. https://doi.org/10.53850/joltida.945975

Bax, S. (2003). CALL—past, present and future. *System*, *31*(1), 13-28. https://doi.org/10.1016/S0346-251X(02)00071-4

Beitchman, J., & Brownlie, E. (2005). Language development and its impact on children's psychosocial and emotional development. *Encyclopedia on early childhood development*, 1-7. https://www.child-encyclopedia.com/language-development-and-literacy/according-experts/language-development-and-its-impact-childrens

Berguno, G., & Bowler, D. M. (2004). Communicative interactions, knowledge of a second language, and theory of mind in young children. *Genetic Psychology*, *165*(3), 293-309. https://doi.org/10.3200/GNTP.165.3.293-309

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and cognition*, *13*(4), 525-531. https://doi.org/10.1017/S1366728909990423

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of educational psychology*, *98*(1), 44. https://doi.org/10.1037/0022-0663.98.1.44

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language learning*, *56*, 9-49. https://doi.org/10.1111/j.1467-9922.2006.00353.x

Birdsong, D., & Vanhove, J. (2016). Age of second language acquisition: Critical periods and social concerns. In E. M. Nicoladis, S. (Ed.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 163–181). American Psychological Association. https://doi.org/10.1037/14939-010

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). MIT Press.

Blair-Early, A., & Zender, M. (2008). User interface design principles for interaction design. *Design Issues*, *24*(3), 85-107. https://doi.org/10.1162/desi.2008.24.3.85

Blake, R. (2016). Technology and the Four Skills. *Language Learning & Technology*, *20*(2), 129-142. http://llt.msu.edu/issues/june2016/blake.pdf

Bloom, P. (2000). How children learn the meanings of words. MIT Press.

Bloom, P. (2002). How children learn the meanings of words. MIT press.

Borgers, N. d. L., E.; Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin of Sociological Methodology/Bulletin de méthodologie sociologique*, *66*(1), 60-75. https://doi.org/10.1177/075910630006600106

Bourgonjon, J., Valcke, M., Soetaert, R., & Schellens, T. (2010). Students' perceptions about the use of video games in the classroom. *Computers & Education*, *54*(4), 1145-1156. https://doi.org/10.1016/j.compedu.2009.10.022

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in human behavior*, *28*(3), 771-780. https://doi.org/10.1016/j.chb.2011.11.020

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57-71). American Psychological Association. https://doi.org/10.1037/13620-004

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications. https://digital.casalini.it/9781526417305

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer handbook of robotics* (pp. 1935-1971). Springer. https://doi.org/10.1007/978-3-319-32552-1_71

Brito, R., & Dias, P. (2020). "Which apps are good for my children?": How the parents of young children select apps. *International Journal of Child-Computer Interaction*, *26*, 100188. https://doi.org/10.1016/j.ijcci.2020.100188

Broekman, F. L., Piotrowski, J. T., Beentjes, H. W. J., & Valkenburg, P. M. (2016). A parental perspective on apps for young children. *Computers in human behavior*, *63*, 142-151. https://doi.org/10.1016/j.chb.2016.05.017

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (Vol. 1, pp. 4-64). Blackwell. https://www.blackwellpublishing.co.uk/content/BPL_Images/Content_store/Sample_chapter/0631205918/Archibald.pdf

Bucholtz, M. (2000). The politics of transcription. *Journal of pragmatics*, *32*(10), 1439-1465. https://doi.org/10.1016/S0378-2166(99)00094-6

Bull, S., & Kay, J. (2010). Open learner models. In *Advances in intelligent tutoring systems* (pp. 301-322). Springer. https://doi.org/10.1007/978-3-642-14363-2_15

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., Ormerod, C. M., Fabiyi, D. G., Ivan, R., & Walsh, C. (2024). The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges. *arXiv preprint arXiv:2406.18900*. https://doi.org/10.48550/arXiv.2406.18900

Bunz, M., & Janciute, L. (2018). *Artificial intelligence and the Internet of Things: UK policy opportunities and challenges*. University of Westminster Press. https://doi.org/10.16997/book25

Cachia, M., & Millward, L. (2011). The telephone medium and semi-structured interviews: a complementary fit. *Qualitative Research in Organizations and Management: An International Journal*, *6*(3), 265-277. https://doi.org/10.1108/17465641111188420

Cai, S., Liu, E., Shen, Y., Liu, C., Li, S., & Shen, Y. (2023). Probability learning in mathematics using augmented reality: impact on students' learning gains and attitudes. In *Cross Reality (XR) and Immersive Learning Environments (ILEs) in Education* (pp. 22-35). Routledge. https://doi.org/10.1080/10494820.2019.1696839

Calvert, S. L., Richards, M. N., & Kent, C. C. (2014). Personalized interactive characters for toddlers' learning of seriation from a video presentation. *Journal of Applied Developmental Psychology*, *35*(3), 148-155. https://doi.org/10.1016/j.appdev.2014.03.004

Canadian Paediatric Society, & Digital Health Task Force. (2017). Screen time and young children: Promoting health and development in a digital world. *Paediatrics & Child Health*, *22*(8), 461. https://doi.org/10.1093/pch/pxx123

Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of child language*, *23*(3), 645-673. https://doi.org/10.1017/S0305000900008989

Carey, S., & Bartlett, E. (1978). *Acquiring a single new word* https://eric.ed.gov/?id=ED198703

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, *132*(3), 354. https://doi.org/10.1037/0033-2909.132.3.354

Cerezo, R., Calderón, V., & Romero, C. (2019). A holographic mobile-based application for practicing pronunciation of basic English vocabulary for Spanish speaking children. *International Journal of Human-Computer Studies*, *124*, 13-25. https://doi.org/10.1016/j.ijhcs.2018.11.009

Chan, C. K. Y., & Tsi, L. H. (2024). Will generative AI replace teachers in higher education? A study of teacher and student perceptions. *Studies in Educational Evaluation*, *83*, 101395. https://doi.org/10.1016/j.stueduc.2024.101395

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, *8*(4), 293-332. https://doi.org/10.1207/s1532690xci0804_2

Chapelle, C. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology*, *2*(1), 22-34. https://doi.org/10.64152/10125/25030

Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

Chapman, J. R., & Rich, P. J. (2018). Does educational gamification improve students' motivation? If so, which game elements work best? *Journal of Education for Business*, *93*(7), 315-322. https://doi.org/10.1080/08832323.2018.1490687

Chaudron, S., DI, G. R., & Gemo, M. (2018). Young children (0-8) and digital technology-A qualitative study across Europe. https://doi.org/10.2760/294383

Chen, C. H., Koong, C. S., & Liao, C. (2022). Influences of Integrating Dynamic Assessment into a Speech Recognition Learning Design to Support Students' English Speaking Skills, Learning Anxiety and Cognitive Load. *Educational Technology & Society*, *25*(1), 1-14. https://www.jstor.org/stable/48647026

Chen, L. J., Chen, P. P., & Lin, Z. J. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chen, T., Peng, L., Yin, X., Rong, J., Yang, J., & Cong, G. (2020). Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic. *Healthcare*, *8*(3), 200. https://doi.org/10.3390/healthcare8030200

Chen, Y., Mayall, H. J., York, C. S., & Smith, T. J. (2019). Parental perception and English learners' mobile-assisted language learning: An ethnographic case study from a technology-based funds of knowledge approach. *Learning, Culture and Social Interaction*, *22*, 100325. https://doi.org/10.1016/j.lcsi.2019.100325

Chen, Y. L., Hsu, C. C., Lin, C. Y., & Hsu, H. H. (2022). Robot-Assisted Language Learning: Integrating Artificial Intelligence and Virtual Reality into English Tour Guide Practice. *Education Sciences*, *12*(7), 437. https://doi.org/10.3390/educsci12070437

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Developmental psychology*, *32*(5), 867. https://doi.org/10.1037/0012-1649.32.5.867

Chiasson, S., & Gutwin, C. (2005). *Design principles for children's technology* (HCI-TR-2005-02). (interfaces, Issue. https://chorus.scs.carleton.ca/wp-content/papercite-data/pdf/chiasson2005design-tr.pdf

China, M. o. E. o. t. P. s. R. o. (2019). *Education Modernization 2035*. http://www.moe.gov.cn/jyb_xwfb/moe_2082/201902/t20190223_370857.html

China, S. C. o. t. P. s. R. o. (2017). *New Generation Artificial Intelligence Development Plan*. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

Chomsky, N. (1986). Knowledge of language: Its nature, origin, and use. Praeger.

Chou, M. H. (2014). Assessing English vocabulary and enhancing young English as a Foreign Language (EFL) learners' motivation through games, songs, and stories. *Education 3-13*, *42*(3), 284-297. https://doi.org/10.1080/03004279.2012.680899

Christensen, L. B., Johnson, B., Turner, L. A., & Christensen, L. B. (2011). Research methods, design, and analysis. https://www.pearsonhighered.com/assets/preface/0/2/0/5/0205944566.pdf

Chrysochoou, E., Bablekou, Z., Masoura, E., & Tsigilis, N. (2013). Working memory and vocabulary development in Greek preschool and primary school children. *European Journal of Developmental Psychology*, *10*(4), 417-432. https://doi.org/10.1080/17405629.2012.686656

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The modern language journal*, *80*(2), 183-198. https://doi.org/10.1111/j.1540-4781.1996.tb01159.x

Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge University Press.

Clark, E. V. (2016). *First Language Acquisition* (3rd ed.). Cambridge University Press.

Clark, E. V. (2017). Lexical acquisition and the structure of the mental lexicon. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.96

Clark, E. V., & Casillas, M. (2015). First language acquisition. In A. Hippisley & G. Stump (Eds.), *The Routledge Handbook of Linguistics* (pp. 311–326). Routledge. https://doi.org/10.4324/9781315718453-21

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, *3*(3), 149-210. https://doi.org/10.1007/BF01320076

Clarke, V. B., V. (2016). Thematic analysis. *The journal of positive psychology*, *12*(3), 297-298. https://doi.org/10.1080/17439760.2016.1262613

Cohen, F., Schunke, J., Vogel, E., & Anders, Y. (2020). Longitudinal Effects of the Family Support Program Chancenreich on Parental Involvement and the Language Skills of Preschool Children. *Front Psychol*, *11*, 1282. https://doi.org/10.3389/fpsyg.2020.01282

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, *51*(1), 42-100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & cognition*, *34*(8), 1754-1768. https://doi.org/10.3758/BF03195936

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage publications.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage publications.

Cui, W., Xue, Z., & Thai, K.-P. (2018). Performance comparison of an AI-based adaptive learning system in China. 2018 Chinese automation congress (CAC),

Cunningham, U. (2014). Teaching the Disembodied: Othering and Activity Systems in a Blended Synchronous Learning Situation. *International Review of Research in Open and Distributed Learning*, *15*(6), 33-51. https://doi.org/10.19173/irrodl.v15i6.1793

Dahl, A., & Vulchanova, M. D. (2014). Naturalistic acquisition in an early language classroom. *Front Psychol*, *5*, 329. https://doi.org/10.3389/fpsyg.2014.00329

Dalacosta, K., Kamariotaki-Paparrigopoulou, M., Palyvos, J. A., & Spyrellis, N. (2009). Multimedia application with animated cartoons for teaching science in elementary education. *Computers & Education*, *52*(4), 741-748. https://doi.org/10.1016/j.compedu.2008.11.018

Davis, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results [Doctoral dissertation, Massachusetts Institute of Technology].

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer-Technology - a Comparison of 2 Theoretical-Models. *Management science*, *35*(8), 982-1003. https://doi.org/10.1287/mnsc.35.8.982

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

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional science*, *38*(2), 105-134. https://doi.org/10.1007/s11251-009-9110-0

de Saussure, F. (1959). *Course in general linguistics* (W. Baskin, Trans.). Philosophical Library. (1916)

de Villiers, J. (2007). The Interface of Language and Theory of Mind. *Lingua*, *117*(11), 1858-1878. https://doi.org/10.1016/j.lingua.2006.11.006

De Villiers, J. G., & de Villiers, P. A. (2014). The role of language in theory of mind development. *Topics in Language Disorders*, *34*(4), 313-328. https://doi.org/10.1097/TLD.0000000000000037

Deci, E. L., & Ryan, R. M. (2008). Self-Determination Theory: A Macrotheory of Human Motivation, Development, and Health. *Canadian Psychology-Psychologie Canadienne*, *49*(3), 182-185. https://doi.org/10.1037/a0012801

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Dede, C. (2014). *The role of digital technologies in deeper learning* (Students at the Center: Deeper Learning Research Series, Issue. https://files.eric.ed.gov/fulltext/ED561254.pdf

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: a systematic review. *Computer Assisted Language Learning*, *34*(7), 934-957. https://doi.org/10.1080/09588221.2019.1648298

Dennis, N. K. (2024). Using AI-Powered Speech Recognition Technology to Improve English Pronunciation and Speaking Skills. *IAFOR Journal of Education*, *12*(2), 107-126. https://doi.org/10.22492/ije.12.2.05

Denzin, N. (2006). *Sociological methods: a sourcebook* (5th ed.). Aldine Transaction.

Denzin, N. K. (1970). The research act: A theoretical introduction to sociological methods. Routledge.

Department for Education. (2019). Realising the potential of technology in education: A strategy for education providers and the technology industry. Department for Education London, UK

Diefenbach, S., Kolb, N., & Hassenzahl, M. (2014). The'hedonic'in human-computer interaction: history, contributions, and future research directions Proceedings of the 2014 conference on Designing interactive systems,

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language teaching*, *31*(3), 117-135. https://doi.org/10.1017/S026144480001315X

Doughty, C. J., & Long, M. H. (2008). *The handbook of second language acquisition*. John Wiley & Sons.

Druin, A., Bederson, B. B., Hourcade, J. P., Sherman, L., Revelle, G., Platner, M., & Weng, S. (2001). Designing a digital library for young children. Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries,

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and education. Artificial intelligence*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Dunn, D. M. (2019). *Peabody picture vocabulary test* (5th ed.). Pearson Assessments.

Dunn, L. M. (1959). Peabody picture vocabulary test [Measurement instrument]. American Guidance Service.

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. https://doi.org/10.64152/10125/25032

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second language acquisition*, *18*(1), 91-126. https://doi.org/10.1017/S0272263100014698

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The modern language journal*, *92*(2), 232-249. https://doi.org/10.1111/j.1540-4781.2008.00716.x

Ellis, R. (2015). *Understanding second language acquisition 2nd edition*. Oxford university press. https://doi.org/10.1016/j.system.2017.07.013

Eltalhi, S., Kutrani, H., Imsallim, R., & Elrfadi, M. (2021). The usability of BenKids mobile learning app in vocabulary teaching for preschool. *iJIM*, *15*(24), 5-18. https://doi.org/10.3991/ijim.v15i24.22237

Engel de Abreu, P. M. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory*, *19*(5), 529-537. https://doi.org/10.1080/09658211.2011.590504

Engwall, O., & Lopes, J. (2022). Interaction and collaboration in robot-assisted language learning for adults. *Computer Assisted Language Learning*, *35*(5-6), 1273-1309. https://doi.org/10.1080/09588221.2020.1799821

Faqih, K. M., & Jaradat, M.-I. R. M. (2015). Assessing the moderating effect of gender differences and individualism-collectivism at individual-level on the adoption of mobile commerce technology: TAM3 perspective. *Journal of Retailing and Consumer Services*, *22*, 37-52. https://doi.org/10.1016/j.jretconser.2014.09.006

Faudzi, M. A., Cob, Z. C., Ghazali, M., Omar, R., & Sharudin, S. A. (2024). User interface design in mobile learning applications: Developing and evaluating a questionnaire for measuring learners' extraneous cognitive load. *Heliyon*, *10*(18), e37494. https://doi.org/10.1016/j.heliyon.2024.e37494

Felix, U. (2008). The unreasonable effectivness of CALL: What have we learned in two decades of research? *ReCALL*, *20*(2), 141-161. https://doi.org/10.1017/S0958344008000323

Fidan, M., & Tuncel, M. (2019). Integrating augmented reality into problem based learning: The effects on learning achievement and attitude in physics education. *Computers & Education*, *142*, 103635. https://doi.org/10.1016/j.compedu.2019.103635

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE.

Filgona, J., Sakiyo, J., Gwany, D., & Okoronka, A. (2020). Motivation in learning. *Asian Journal of Education and social studies*, *10*(4), 16-37. https://doi.org/10.9734/ajess/2020/v10i430273

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Addison-Wesley.

Fitria, T. N. (2021). Artificial intelligence (AI) in education: Using AI tools for teaching and learning process. Prosiding Seminar Nasional & Call for Paper STIE AAS,

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of memory and language*, *41*(1), 78-104. https://doi.org/10.1006/jmla.1999.2638

Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, *27*(4), 935-951. https://doi.org/10.1348/026151008X397189

Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. https://doi.org/10.64152/10125/44709

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology. The broaden-and-build theory of positive emotions. *Am Psychol*, *56*(3), 218-226. https://doi.org/10.1037/0003-066X.56.3.218

Fu, S., Gu, H., & Yang, B. (2020). The affordances of AI-enabled automatic scoring applications on learners’ continuous learning intention: An empirical study in China. *British Journal of Educational Technology*, *51*(5), 1674-1692. https://doi.org/10.1111/bjet.12995

Furió, D., Juan, M. C., Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: a comparative study. *Journal of Computer Assisted Learning*, *31*(3), 189-201. https://doi.org/10.1111/jcal.12071

Ganapathy, M., & Seetharam, S. A. (2016). The effects of using multimodal approaches in meaning-making of 21st century literacy texts among ESL students in a private school in Malaysia. *Advances in Language and Literary Studies*, *7*(2), 143-155. https://doi.org/10.7575/aiac.alls.v.7n.2p.143

Gardner, R. C. (2007). Motivation and second language acquisition. *Porta Linguarum*, *8*. https://doi.org/10.30827/Digibug.31616

Gass, S. M., Behney, J., & Plonsky, L. (2020). *Second language acquisition: An introductory course*. Routledge. https://doi.org/10.4324/9781315181752

Gass, S. M., & Selinker, L. (2008). *Second language acquisition* (3rd ed.).

Gathercole, S. E. (1998). The development of memory. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(1), 3-27. https://doi.org/10.1017/S0021963097001753

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in cognitive sciences*, *3*(11), 410-419. https://doi.org/10.1016/S1364-6613(99)01388-1

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513-543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental psychology*, *29*(4), 770. https://doi.org/10.1037/0012-1649.29.4.770

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of memory and language*, *28*(2), 200-213. https://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S. E., Hitch, G. J., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental psychology*, *33*(6), 966. https://doi.org/10.1037/0012-1649.33.6.966

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Dev Psychol*, *40*(2), 177-190. https://doi.org/10.1037/0012-1649.40.2.177

Gillham, B. (2001). *Research interview*. Bloomsbury Publishing.

Girela-Serrano, B. M., Spiers, A. D. V., Ruotong, L., Gangadia, S., Toledano, M. B., & Di Simplicio, M. (2024). Impact of mobile phones and wireless devices use on children and adolescents' mental health: a systematic review. *Eur Child Adolesc Psychiatry*, *33*(6), 1621-1651. https://doi.org/10.1007/s00787-022-02012-8

Given, L. M. (2008). The Sage encyclopedia of qualitative research methods. Sage Publications.

Gjelaj, M., Buza, K., Shatri, K., & Zabeli, N. (2020). Digital Technologies in Early Childhood: Attitudes and Practices of Parents and Teachers in Kosovo. *International Journal of Instruction*, *13*(1), 165-184. https://doi.org/10.29333/iji.2020.13111a

Godwin-Jones, R. (2011). Mobile Apps for Language Learning. *Language Learning & Technology*, *15*(2), 2-11. https://doi.org/10.64152/10125/44244

Gonzalez-DeHass, A. R., Willems, P. P., Powers, J. R., & Musgrove, A. T. (2022). Parental involvement in supporting students' digital learning. *Educational Psychologist*, *57*(4), 281-294. https://doi.org/10.1080/00461520.2022.2129647

Goodall, J. (2018). Leading for parental engagement: Working towards partnership. In (Vol. 38, pp. 143-146): Taylor & Francis.

Goossens, N. A., Camp, G., Verkoeijen, P. P., Tabbers, H. K., & Zwaan, R. A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition*, *3*(3), 177-182. https://doi.org/10.1016/j.jarmac.2014.05.003

Goswami, U. (2019). *Cognitive development and cognitive neuroscience: The learning brain*. Routledge. https://doi.org/10.4324/9781315684734

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124-132. https://doi.org/10.1007/s40593-015-0086-4

Griffith, S. F., & Arnold, D. H. (2019). Home learning in the new mobile age: parent-child interactions during joint play with educational apps in the US. *Journal of Children and Media*, *13*(1), 1-19. https://doi.org/10.1080/17482798.2018.1489866

Gu, J. (2024). Digital Tools in Language Learning: Optimizing Memory and Attention for College Students. *International Journal of Human-Computer Interaction*, 1-11. https://doi.org/10.1080/10447318.2024.2400384

Guillemin, M., Gillam, L., Rosenthal, D., & Bolitho, A. (2012). Human research ethics committees: Examining their roles and practices. *Journal of Empirical Research on Human Research Ethics*, *7*(3), 38-49. https://doi.org/10.1525/jer.2012.7.3.38

Gündüz, N. (2005). Computer assisted language learning. *Journal of language and linguistic studies*, *1*(2), 193-214. https://dergipark.org.tr/en/download/article-file/104667

Guo, S. C., & Lan, Y. J. (2023). Virtual world-supported contextualized multimodal EFL learning at a library. *Language Learning & Technology*, *27*(2), 176-198. https://doi.org/10.64152/10125/73508

Halcomb, E. J., & Davidson, P. M. (2006). Is verbatim transcription of interview data always necessary? *Appl Nurs Res*, *19*(1), 38-42. https://doi.org/10.1016/j.apnr.2005.06.001

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Dev Sci*, *6*(3), 346-359. https://doi.org/10.1111/1467-7687.00289

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work?—A literature review of empirical studies on gamification* 2014 47th Hawaii International Conference on System Sciences,

Hampel, R., & Stickler, U. (2005). New skills for new classrooms: Training tutors to teach languages online. *Computer Assisted Language Learning*, *18*(4), 311-326. https://doi.org/10.1080/09588220500335455

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of educational psychology*, *90*(3), 414. https://doi.org/10.1037/0022-0663.90.3.414

Harris, A., & Goodall, J. (2008). Do parents know they matter? Engaging all parents in learning. *Educational research*, *50*(3), 277-289. https://doi.org/10.1080/00131880802309424

Hassenzahl, M. (2003). The thing and I: understanding the relationship between user and product. In M. A. Blythe, Overbeeke, K., Monk, A. F., Wright, P. C. (Ed.), *Funology* (pp. 31–42). Kluwer Academic Publishers.

Hassenzahl, M. (2010). *Experience design: Technology for all the right reasons* (Vol. 8). Morgan & Claypool Publishers.

Hassenzahl, M. (2018). The thing and I: understanding the relationship between user and product. *Funology 2: from usability to enjoyment*, 301-313. https://doi.org/10.1007/978-3-319-68213-6_19

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. *Mensch & Computer 2003: Interaktion in Bewegung*, 187-196. https://doi.org/10.1007/978-3-322-80058-9_19

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174138

Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in human behavior*, *44*, 81-95. https://doi.org/10.1016/j.chb.2014.11.009

Heift, T. (2022). Intelligent computer assisted language learning. In *Research Questions in Language Education and Applied Linguistics: A Reference Guide* (pp. 655-658). Springer. https://doi.org/10.1007/978-3-030-79143-8_115

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. Routledge. https://doi.org/10.4324/9780203012215

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

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second language acquisition*, *21*(2), 303-317. https://doi.org/10.1017/S0272263199002089

Hillier, M. (2020). Bridging the digital divide with off-line e-learning. In *Expanding Horizons in Open and Distance Learning* (pp. 110-121). Routledge.

Hilts, A., Part, R., & Bernacki, M. L. (2018). The roles of social influences on student competence, relatedness, achievement, and retention in STEM. *Science Education*, *102*(4), 744-770. https://doi.org/10.1002/sce.21449

Hirsh-Pasek, K., & Golinkoff, R. M. (2006). *Action meets word: How children learn verbs*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.001.0001

Hoff, E. (2006). How social contexts support and shape language development. *Developmental review*, *26*(1), 55-88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E., Rumiche, R., Burridge, A., Ribot, K. M., & Welsh, S. N. (2014). Expressive Vocabulary Development in Children from Bilingual and Monolingual Homes: A Longitudinal Study from Two to Four Years. *Early Child Res Q*, *29*(4), 433-444. https://doi.org/10.1016/j.ecresq.2014.04.012

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law* (9287192375). https://discovery.ucl.ac.uk/id/eprint/10158376

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

Hsu, T.-C., Abelson, H., Lao, N., Tseng, Y.-H., & Lin, Y.-T. (2021). Behavioral-pattern exploration and development of an instructional tool for young children to learn AI. *Computers and Education: Artificial Intelligence*, *2*, 100012. https://doi.org/10.1016/j.caeai.2021.100012

Huang, B. Y., Hew, K. F., & Lo, C. K. (2019). Investigating the effects of gamification-enhanced flipped learning on undergraduate students' behavioral and cognitive engagement. *Interactive learning environments*, *27*(8), 1106-1126. https://doi.org/10.1080/10494820.2018.1495653

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237-257. https://doi.org/10.1111/jcal.12610

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology*, *96*(2), 150-166. https://doi.org/10.1016/j.jecp.2006.09.002

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.011

Indriati, I. (2016). Parents' involvement in supporting their children learn English. *Journal on English as a Foreign Language*, *6*(2), 145-170. https://doi.org/10.23971/jefl.v6i2.433

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and emotion*, *29*(4), 297-325. https://doi.org/10.1007/s11031-006-9019-8

James, K. K., & Mayer, R. E. (2019). Learning a second language by playing a game. *Applied Cognitive Psychology*, *33*(4), 669-674. https://doi.org/10.1002/acp.3492

Javor, R. (2016). Bilingualism, theory of mind and perspective-taking: The effect of early bilingual exposure. *Psychology and Behavioral Sciences*, *5*(6), 143-148. https://doi.org/10.11648/j.pbs.20160506.13

Jeon, J. (2022). Exploring a self-directed interactive app for informal EFL learning: a self-determination theory perspective. *Education and Information Technologies*, *27*(4), 5767-5787. https://doi.org/10.1007/s10639-021-10839-y

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, *37*(1-2), 1-26. https://doi.org/10.1080/09588221.2021.2021241

Ji, X. R., Li, G., & Gunderson, L. (2022). Validation of the PPVT–5 for Chinese-English bilingual learners: Application of cross-classified mixed effects model. *Research Methods in Applied Linguistics*, *1*(2), 100013. https://doi.org/10.1016/j.rmal.2022.100013

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47-77. https://doi.org/10.1093/applin/21.1.47

Jiemian. (2020, November 24). *2020 中国 K12 在线教育移动应用下载分析报告 [2020 China K12 online education mobile app download analysis report]*. Jiemian News. https://www.jiemian.com/article/5307505.html

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, *19*(3), 357-377. https://doi.org/10.1093/applin/19.3.357

Johann, V. E., & Karbach, J. (2018). Validation of new online game-based executive function tasks for children. *J Exp Child Psychol*, *176*, 150-161. https://doi.org/10.1016/j.jecp.2018.07.009

Kabali, H. K., Irigoyen, M. M., Nunez-Davis, R., Budacki, J. G., Mohanty, S. H., Leister, K. P., & Bonner, R. L., Jr. (2015). Exposure and Use of Mobile Media Devices by Young Children. *Pediatrics*, *136*(6), 1044-1050. https://doi.org/10.1542/peds.2015-2151

Kalaycı, G., & Ergül, H. (2020). Teachers' perceptions of the role of parental involvement in teaching English to young learners. *Journal of language and linguistic studies*, *16*(3), 1167-1176. https://doi.org/10.17263/jlls.803608

Kallio, H., Pietila, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *J Adv Nurs*, *72*(12), 2954-2965. https://doi.org/10.1111/jan.13031

Karisma, I. G. P. Y. A., & Gui, A. (2023). Understanding e-learning system acceptance: An empirical analysis of key factors among elementary school students using TAM model. *Jurnal TAM (Technology Acceptance Model)*, *14*(2), 213–220. https://doi.org/10.56327/jurnaltam.v14i2.1599

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary school children. *Journal of Applied Research in Memory and Cognition*, *3*(3), 198-206. https://doi.org/10.1016/j.jarmac.2014.07.008

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychological science*, *21*(2), 260-267. https://doi.org/10.1177/0956797609357327

Khlaif, Z. N., Itmazi, J., Farid, S., Shaqour, A. Z., & Kouraïchi, B. (2019). Exploring children experience with educational mobile technology. *Research in Learning Technology*, *27*, 2242. https://doi.org/10.25304/rlt.v27.2242

Kickmeier-Rust, M. D., & Albert, D. (2010). Micro-adaptivity: Protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted Learning*, *26*(2), 95-105. https://doi.org/10.1111/j.1365-2729.2009.00332.x

Kiefer, S. M., Alley, K. M., & Ellerbrock, C. R. (2015). Teacher and peer support for young adolescents' motivation, engagement, and school belonging. *Rmle Online*, *38*(8), 1-18. https://doi.org/10.1080/19404476.2015.11641184

Knott, E., Rao, A. H., Summers, K., & Teeger, C. (2022). Interviews in the social sciences. *Nature Reviews Methods Primers*, *2*(1), 73. https://doi.org/10.1038/s43586-022-00150-6

Kolluru, V., Mungara, S., & Chintakunta, A. N. (2018). Adaptive learning systems: Harnessing AI for customized educational experiences. *International Journal of Computational Science and Information Technology (IJCSITY)*, *6*(1), 2. https://doi.org/10.5121/ijcsity.2018.6302

Kong, S. C., & Li, K. M. (2009). Collaboration between school and parents to foster information literacy: Learning in the information society. *Computers & Education*, *52*(2), 275-282. https://doi.org/10.1016/j.compedu.2008.08.004

Kormos, E., & Wisdom, K. (2021). Rural schools and the digital divide: Technology in the learning experience. *Theory & Practice in Rural Education*, *11*(1), 25-39. https://doi.org/10.3776/tpre.2021.v11n1p25-39

Korpershoek, H., Canrinus, E. T., Fokkens-Bruinsma, M., & de Boer, H. (2020). The relationships between school belonging and students' motivational, social-emotional, behavioural, and academic outcomes in secondary education: a meta-analytic review. *Research papers in education*, *35*(6), 641-680. https://doi.org/10.1080/02671522.2019.1615116

Krashen, S. (1982). Principles and practice in second language acquisition. Pergamon Press.

Krashen, S. D. (1985). The input hypothesis: Issues and implications. Longman.

Krashen, S. D., & Terrell, T. (1983). The natural approach: Language acquisition in the classroom. Pergamon Press.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149-174. https://doi.org/10.1006/jmla.1994.1008

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831-843. https://doi.org/10.1038/nrn1533

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, *20*(3), 271-289. https://doi.org/10.1017/S0958344008000335

Lamrani, R., & Abdelwahed, E. (2020). Game-based learning and Gamification to improve skills in early years education. *Computer Science and Information Systems*, *17*(1), 339-356. https://doi.org/10.2298/Csis190511043l

Lan, Y. (2020). Immersion, interaction and experience-oriented learning: Bringing VR into FL learning. *Language Learning & Technology*, *24*(1), 1-15. https://doi.org/10.64152/10125/44704

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language learning*, *54*(3), 399-436. https://doi.org/10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1-26. https://doi.org/10.1093/applin/22.1.1

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, *6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lee, J., Jang, J., & Plonsky, L. (2015). The Effectiveness of Second Language Pronunciation Instruction: A Meta-Analysis. *Applied Linguistics*, *36*(3), 345-366. https://doi.org/10.1093/applin/amu040

Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis tools: A call for data analysis triangulation. *School psychology quarterly*, *22*(4), 557-584. https://doi.org/10.1037/1045-3830.22.4.557

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59-67. https://doi.org/10.1080/21548331.1967.11707799

Lesaux, N. K., Lipka, O., & Siegel, L. S. (2006). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing*, *19*(1), 99-131. https://doi.org/10.1007/s11145-005-4713-6

Levelt, W. J. (1989). Speaking: From intention to articulation. MIT press.

Levy, M. (2006). Effective Use of CALL Technologies: Finding the Right Balance. In R. P. Donaldson & M. A. Haggstrom (Eds.), *Changing language education through CALL* (Vol. 1, pp. 1-18). Routledge. <Go to ISI>://WOS:000273642800002

Li, G., & Lin, Z. (2022). Chinese-Canadian parents supporting dual online-offline learning and early biliteracy during COVID-19 school closures:“Two schools, two languages”. In *Handbook of Research on Family Literacy Practices and Home-School Connections* (pp. 1-17). IGI Global.

Li, J., Antonenko, P. D., & Wang, J. (2019). Trends and issues in multimedia learning research in 1996–2016: A bibliometric analysis. *Educational Research Review*, *28*, 100282. https://doi.org/10.1016/j.edurev.2019.100282

Li, R. (2020). Using artificial intelligence in learning English as a foreign language: An examination of IELTS Liulishuo as an online platform. *Journal of Higher Education Research*, *1*(2). https://doi.org/10.32629/jher.v1i2.178

Li, X., & Chu, S. K. W. (2021). Exploring the effects of gamification pedagogy on children’s reading: A mixed-method study on academic performance, reading-related mentality and behaviors, and sustainability. *British Journal of Educational Technology*, *52*(1), 160-178. https://doi.org/10.1111/bjet.13057

Li, X., Zhang, X., Zhao, Y., Zhang, L., & Shang, J. (2023). Exploring the role of learning through play in promoting multimodal learning among children: a pilot study in Chinese first-tier cities. *Front Psychol*, *14*, 1103311. https://doi.org/10.3389/fpsyg.2023.1103311

Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *32*(8), 878-919. https://doi.org/10.1080/09588221.2018.1541359

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: a meta-analysis. *Psychon Bull Rev*, *21*(4), 861-883. https://doi.org/10.3758/s13423-013-0565-2

Liu, H.-j. (2015). Learner autonomy: The role of motivation in foreign language learning. *Journal of Language Teaching and Research*, *6*(6), 1165. https://doi.org/10.17507/jltr.0606.02

Liu, P. L., & Chen, C. J. (2023). Using an AI-Based Object Detection Translation Application for English Vocabulary Learning. *Educational Technology & Society*, *26*(3), 5-20. https://doi.org/10.30191/Ets.202307_26(3).0002

Livingstone, S., & Helsper, E. J. (2008). Parental Mediation of Children's Internet Use. *Journal of broadcasting & electronic media*, *52*(4), 581-599. https://doi.org/10.1080/08838150802437396

Livingstone, S. M., & Blum-Ross, A. (2020). Parenting for a digital future: How hopes and fears about technology shape children's lives. Oxford University Press.

Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language teaching*, *51*(3), 285-329. https://doi.org/10.1017/S0261444818000125

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press. https://doi.org/10.1016/B978-012589042-7/50015-3

Looije, R., Neerincx, M. A., & Lange, V. d. (2008). Children's responses and opinion on three bots that motivate, educate and play. https://doi.org/10.14198/JoPha.2008.2.2.03

Luckin, R. (2010). *Re-designing learning contexts: Technology-rich, learner-centred ecologies*. Routledge. https://doi.org/10.4324/9780203854754

Ma, Z., Wang, J., Li, J., & Jia, Y. (2021). The association between obesity and problematic smartphone use among school-age children and adolescents: a cross-sectional study in Shanghai. *BMC public health*, *21*(1), 2067. https://doi.org/10.1186/s12889-021-12124-6

MacIntyre, P. D. (2003). Motivation, anxiety and emotion in second language acquisition. In *Individual differences and instructed language learning* (pp. 45-68). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.05mac

Mackey, A. (2013). Input, interaction and corrective feedback in L2 learning. Oxford University Press.

Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of Reading Comprehension for Struggling Readers: The Case of Spanish-speaking Language Minority Learners. *J Educ Psychol*, *102*(3), 701-711. https://doi.org/10.1037/a0019135

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive science*, *14*(1), 57-77. https://doi.org/10.1207/s15516709cog1401_4

Marton-Williams, J. (1986). Questionnaire design. In R. M. Worcester & J. Downham (Eds.), *Consumer market research handbook* (3rd ed., Vol. 3, pp. 30). McGraw-Hill.

Masood, M., & Thigambaram, M. (2015). The Usability of Mobile Applications for Pre-schoolers. *7th World Conference on Educational Sciences*, *197*, 1818-1826. https://doi.org/10.1016/j.sbspro.2015.07.241

Matban, N. (2023). Exploring Parental Attitudes Towards Technology-Assisted English Language Learning (TALL) For Young Children in Turkiye University of Warsaw].

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics*, *27*(3), 403-422. https://doi.org/10.1017/S0142716406060334

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1-19. https://doi.org/10.1207/s15326985ep3201_1

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85-139). Elsevier. https://doi.org/10.1016/s0079-7421(02)80005-6

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, *41*(1), 31-48. https://doi.org/10.1017/CBO9780511816819.004

Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403-423. https://doi.org/10.1111/jcal.12197

Mayer, R. E. (2020). *Multimedia Learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). https://doi.org/10.1017/9781108894333

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. https://doi.org/10.1207/S15326985EP3801_6

Mazouz, A., & Maatallaoui, Y. (2025). Automatic Speech Recognition in Language Learning Applications for Children: Challenges, Solutions, and Future Directions. https://doi.org/10.13140/RG.2.2.19527.41124

McKnight, L., & Fitton, D. (2010). Touch-screen technology for children: giving the right instructions and getting the right responses. Proceedings of the 9th international conference on interaction design and children,

McQuaide, S. (2009). Making Education Equitable in Rural China through Distance Learning. *International Review of Research in Open and Distributed Learning*, *10*(1). https://doi.org/10.19173/irrodl.v10i1.590

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Dev Psychol*, *49*(2), 270-291. https://doi.org/10.1037/a0028228

Miao, F., & Holmes, W. (2021). *AI and education: A guidance for policymakers*. Unesco Publishing.

Mikelic Preradovic, N., Lešin, G., & Šagud, M. (2016). Investigating Parents' Attitudes towards Digital Technology Use in Early Childhood: A Case Study from Croatia. *Informatics in education*, *15*(1), 127-146. https://doi.org/10.15388/infedu.2016.07

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81. https://doi.org/10.1037/h0043158

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Dev*, *78*(2), 622-646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Millington, N. T. (2011). Using songs effectively to teach English to young learners. *Language education in Asia*, *2*(1), 134-141. https://doi.org/10.5746/LEiA/11/V2/I1/A11/Millington

Ministry of Education, C., Sports, Science and Technology – Japan (MEXT). (2020). *About MEXT: Organization*. Ministry of Education, Culture, Sports, Science and Technology – Japan Retrieved from https://www.mext.go.jp/en/about/organization/index.htm

Mispa, K., Mansor, E. I., & Kamaruddin, A. (2019). Evaluating children's user experience (UX) towards mobile application: The fantasy land prototype. Proceedings of the 5th International ACM In-Cooperation HCI and UX Conference,

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Monnier, C., Boiché, J., Armandon, P., Baudoin, S., & Bellocchi, S. (2022). Is bilingualism associated with better working memory capacity? A meta-analysis. *International Journal of Bilingual Education and Bilingualism*, *25*(6), 2229-2255. https://doi.org/10.1080/13670050.2021.1908220

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118-141. https://doi.org/10.64152/10125/44357

Moore, A. (2004). Hedonism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2004 ed.): Stanford University.

Moorhouse, B. L., & Beaumont, A. M. (2020). Involving parents in their children's school-based English language writing using digital learning. *relc Journal*, *51*(2), 259-267. https://doi.org/10.1177/0033688219859937

Moorthy, K., Yee, T. T., T'ing, L. C., & Kumaran, V. V. (2019). Habit and hedonic motivation are the strongest influences in mobile learning behaviours among higher education students in Malaysia. *Australasian Journal of Educational Technology*, *35*(4), 174-191. https://doi.org/10.14742/ajet.4432

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational psychology review*, *19*, 309-326. https://doi.org/10.1007/s10648-007-9047-2

Morgan, G., & Kegl, J. (2006). Nicaraguan sign language and theory of mind: The issue of critical periods and abilities. *Journal of Child Psychology and Psychiatry*, *47*(8), 811-819. https://doi.org/10.1111/j.1469-7610.2006.01621.x

Mu, Z., & Hu, S. (2023). Unequal childhoods in China: Parental education and children's time use. *J Community Psychol*, *51*(2), 695-723. https://doi.org/10.1002/jcop.22710

Munoto, M., Sumbawati, M. S., & Sari, S. F. M. (2021). The Use of Mobile Technology in Learning With Online and Offline Systems. *International Journal of Information and Communication Technology Education*, *17*(2), 54-67. https://doi.org/10.4018/Ijicte.2021040104

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17-38. https://doi.org/10.1080/09588221.2010.520675

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes*, *61*(1), 107-134. https://doi.org/10.3138/cmlr.61.1.107

Nation, I. (2001). *Learning vocabulary in another language* (Vol. 10). Cambridge university press Cambridge.

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed., Vol. 10). Cambridge University Press. https://doi.org/10.1017/9781009093873

Neumann, M. M. (2015). Young children and screen time: Creating a mindful approach to digital technology. *Australian educational computing*, *30*(2). http://journal.acce.edu.au/index.php/AEC/article/view/67

Neumann, M. M. (2018). Using tablets and apps to enhance emergent literacy skills in young children. *Early childhood research quarterly*, *42*, 239-246. https://doi.org/10.1016/j.ecresq.2017.10.006

Nicholson, S. (2015). A recipe for meaningful gamification. In T. Reiners & L. C. Wood (Eds.), *Gamification in education and business* (pp. 1-20). Springer. https://doi.org/10.1007/978-3-319-10208-5_1

Nikken, P., & Jansz, J. (2014). Developing scales to measure parental mediation of young children's internet use. *Learning Media and Technology*, *39*(2), 250-266. https://doi.org/10.1080/17439884.2013.782038

Nojavanasghari, B., Baltrušaitis, T., Hughes, C. E., & Morency, L.-P. (2016). Emoreact: a multimodal approach and dataset for recognizing emotional responses in children. 8th acm international conference on multimodal interaction,

Noroozi, O., Pijeira-Díaz, H. J., Sobocinski, M., Dindar, M., Järvelä, S., & Kirschner, P. A. (2020). Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies*, *25*, 5499-5547. https://doi.org/10.1007/s10639-020-10229-w

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411. https://doi.org/10.1037/0278-7393.28.3.411

Oliver, R., & Azkarai, A. (2017). Review of child second language acquisition (SLA): Examining theories and research. *Annual Review of Applied Linguistics*, *37*, 62-76. https://doi.org/10.1017/S0267190517000058

Page, M., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737-3753. https://doi.org/10.1098/rstb.2009.0173

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford university press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *45*(3), 255. https://doi.org/10.1037/h0084295

Palaganas, E. C., Sanchez, M. C., Molintas, M. V. P., & Caricativo, R. D. (2017). Reflexivity in qualitative research: A journey of learning. *The Qualitative Report*, *22*(2), 426–438. https://doi.org/10.46743/2160-3715/2017.2552

Palaiologou, I., Arnott, L., & Gray, C. (2019). Parental perspectives of children's use of Internet Connected Toys. In R. Brito & P. Dias (Eds.), *Crianças, famílias e tecnologias: Que desafios? Que caminhos?* (pp. 93–109). Centro Interdisciplinar de Estudos Educacionais. http://hdl.handle.net/10400.21/11659

Palvia, S., Aeron, P., Gupta, P., Mahapatra, D., Parida, R., Rosner, R., & Sindhi, S. (2018). Online education: Worldwide status, challenges, trends, and implications. In (Vol. 21, pp. 233-241): Taylor & Francis.

Papadakis, S. (2021). Tools for evaluating educational apps for young children: a systematic review of the literature. *Interactive Technology and Smart Education*, *18*(1), 18-49.

Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children: what educators and parents need to know. *International Journal of Mobile Learning and Organisation*, *11*(3), 256-277. https://doi.org/10.1504/IJMLO.2017.085338

Papadakis, S., Zaranis, N., & Kalogiannakis, M. (2019). Parental involvement and attitudes towards young Greek children's mobile usage. *International Journal of Child-Computer Interaction*, *22*, 100144. https://doi.org/10.1016/j.ijcci.2019.100144

Paradis, J. (2011). Individual differences in child English second language acquisition Comparing child-internal and child-external factors. *Linguistic approaches to bilingualism*, *1*(3), 213-237. https://doi.org/10.1075/lab.1.3.01par

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (Vol. 55, pp. 174-200). Cambridge University Press. https://doi.org/10.1017/CBO9781139524643.013

Park, Y. (2011). A Pedagogical Framework for Mobile Learning: Categorizing Educational Applications of Mobile Technologies into Four Types. *International Review of Research in Open and Distributed Learning*, *12*(2), 78-102. https://doi.org/10.19173/irrodl.v12i2.791

Pasquarella, A., Chen, X., Lam, K., Luo, Y. C., & Ramirez, G. (2011). Cross-language transfer of morphological awareness in Chinese–English bilinguals. *Journal of Research in Reading*, *34*(1), 23-42. https://doi.org/10.1111/j.1467-9817.2010.01484.x

Pediatrics, A. A. o. (2016). Media and young minds. *Pediatrics*, *138*(5), e20162591. https://doi.org/10.1542/peds.2016-2591

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review*, *18*(4), 315-341. https://doi.org/10.1007/s10648-006-9029-9

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child development*, *76*(2), 502-517. https://doi.org/10.1111/j.1467-8624.2005.00859.x

Piaget, J. (1954). The construction of reality in the child. Routledge.

Piaget, J. (1964). Cognitive development in children: Piaget. *Journal of research in science teaching*, *2*(3), 176-186. https://doi.org/10.1002/tea.3660020306

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children* (Vol. 8). International universities press New York.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in neural information processing systems* (Vol. 28). Curran Associates, Inc. https://papers.nips.cc/paper/5654-deep-knowledge-tracing

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive science*, *34*(4), 602-619. https://doi.org/10.1111/j.1551-6709.2010.01102.x

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of educational psychology*, *90*(1), 25–36. https://doi.org/10.1037/0022-0663.90.1.25

Plowman, L. (2015). Researching young children's everyday uses of technology in the family home. *Interacting with Computers*, *27*(1), 36-46. https://doi.org/10.1093/iwc/iwu031

Plowman, L., McPake, J., & Stephen, C. (2010). The technologisation of childhood? Young children and technology in the home. *Children & Society*, *24*(1), 63-74. https://doi.org/10.1111/j.1099-0860.2008.00180.x

Plowman, L., Stevenson, O., McPake, J., Stephen, C., & Adey, C. (2011). Parents, pre-schoolers and learning with technology at home: some implications for policy. *Journal of Computer Assisted Learning*, *27*(4), 361-371. https://doi.org/10.1111/j.1365-2729.2011.00432.x

Poarch, G. J., & Van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 535-551. https://doi.org/10.1016/j.jecp.2012.06.013

Powney, J., & Watts, M. (1984). Reporting interviews: a code of good practice. *Research intelligence*, *17*, 164-168.

Proudfoot, K. (2023). Inductive/deductive hybrid thematic analysis in mixed methods research. *Journal of mixed methods research*, *17*(3), 308-326. https://doi.org/10.1177/15586898221126816

Qassrawi, R. M., ElMashharawi, A., Itmeizeh, M., & Tamimi, M. H. (2024). AI-powered applications for improving EFL students' speaking proficiency in higher education. *assessment*, *8*, 9. https://doi.org/10.30564/fls.v6i5.6966

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

Quine, W. V. O. (1960). *Word and object*. MIT Press.

Rachels, J. R., & Rockinson-Szapkiw, A. J. (2018). The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer*

Assisted Language Learning, *31*(1-2), 72-89. https://doi.org/10.1080/09588221.2017.1382536

Rasekh Eslami, Z., & Zohoor, S. (2023). Second language (L2) pragmatics and computer assisted language learning (CALL). *Technology Assisted Language Education*, *1*(3), 1-17. https://doi.org/10.22126/tale.2023.2788

Read, J. C. (2008). Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. *Cognition, Technology & Work*, *10*(2), 119-128. https://doi.org/10.1007/s10111-007-0069-9

Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. *Proceedings of the 2006 Conference on Interaction Design and Children*,

Read, J. C., MacFarlane, S., & Casey, C. (2002). *Endurability, engagement and expectations: Measuring children's fun* Interaction design and children, https://www.researchgate.net/publication/228870976

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review*, *8*(4), 364-382. https://doi.org/10.1207/s15327957pspr0804_3

Reeve, J. (2006). Teachers as facilitators: What autonomy-supportive teachers do and why their students benefit. *The elementary school journal*, *106*(3), 225-236. https://doi.org/10.1086/501484

Reinders, H. (Ed.). (2012). *Digital games in language learning and teaching*. Palgrave Macmillan. https://doi.org/10.1057/9781137005267.

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. In A. Ferligoj & A. Mrvar (Eds.), *Developments in applied statistics* (Vol. 19, pp. 159-177). http://mrvar.fdv.uni-lj.si/pub/mz/mz19/reja.pdf

Riconscente, M. M. (2013). Results from a controlled study of the iPad fractions game Motion Math. *Games and Culture*, *8*(4), 186-214. https://doi.org/10.1177/1555412013496894

Rieber, L. P. (1991). Animation, incidental learning, and continuing motivation. *Journal of educational psychology*, *83*(3), 318. https://doi.org/10.1037/0022-0663.83.3.318

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current directions in psychological science*, *16*(4), 183-186. https://doi.org/10.1111/j.1467-8721.2007.00500.x

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., Grau, V., Lagos, F., López, X., & López, V. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education*, *40*(1), 71-94. https://doi.org/10.1016/S0360-1315(02)00099-4

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child development*, *85*(3), 956-970. https://doi.org/10.1111/cdev.12166

Rubin, H. J., & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data* (3rd ed.). sage.

Russell, S. J. N., Peter. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press. https://doi.org/10.1521/978.14625/28806

Safwat, R. F., & Sheikhany, A. R. (2014). Effect of parent interaction on language development in children. *The Egyptian Journal of Otolaryngology*, *30*, 255-263. https://doi.org/10.4103/1012-5574.138488

Saldaña, J. (2011). *Fundamentals of qualitative research*. Oxford University Press.

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE Publications. http://digital.casalini.it/9781529755992

Salemink, K., Strijker, D., & Bosworth, G. (2017). Rural development in the digital age: A systematic literature review on unequal ICT availability, adoption, and use in rural areas. *Journal of Rural Studies*, *54*, 360-371. https://doi.org/10.1016/j.jrurstud.2015.09.001

Saraswati, N. A. A., S.; Purwati, O. (2021). Autonomous MobileAssisted Language Learning for

Young Learners Using Duolingo. . Celtic: A Journal of Culture, English

Language Teaching, Literature and

Linguistics, *8*(2), 235–246. https://doi.org/10.22219/celtic.v8i2.16959

Saxton, M. (2017). Child language: Acquisition and development. SAGE Publications.

Schmidt, C. (2004). The analysis of semi-structured interviews. In U. Flick, E. von Kardorff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 253-258). SAGE Publications.

Schmidt, K., & Blumenthal, Y. (2022). Early immersion education: L2 vocabulary acquisition and the role of non-verbal intelligence. *Cognition and Second Language Acquisition: Studies on pre-school, primary school and secondary school children*, 37-58. https://doi.org/10.24053/9783823391944

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied Linguistics*, *11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*. Cambridge university press.

Selwyn, N. (2019). Should robots replace teachers?: AI and the future of education. Polity.

Sergi, K., Gatewood Jr, R., Elder, A., & Xu, J. (2017). Parental perspectives on children's use of portable digital devices. *Behaviour & Information Technology*, *36*(11), 1148-1161. https://doi.org/10.1080/0144929X.2017.1360941

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

Settles, B., & Meeder, B. (2016). A trainable spaced repetition model for language learning. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),

Shih, B.-Y., Shih, C.-H., Li, C.-C., Chen, T.-H., Chen, Y.-H., & Chen, C.-Y. (2011). Elementary school student's acceptance of Lego NXT: The technology acceptance model, a preliminary investigation. *International Journal of the Physical Sciences*, *6*(22), 5054-5063. https://www.academicjournals.org/app/webroot/article/article1380626307_Shih%20et%20al.pdf

Shneiderman, B. (2004). Designing for fun: how can we design user interfaces to be more fun? *interactions*, *11*(5), 48-50. https://doi.org/10.1145/1015530.1015552

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, *36*(3), 517-554. https://doi.org/10.1080/09588221.2021.1933540

Siegal, M. (1999). Language and thought: The fundamental significance of conversational awareness for cognitive development. *Developmental science*, *2*(1), 1-14. https://doi.org/10.1111/1467-7687.00048

Singh, A., & Manjaly, J. A. (2022). Using curiosity to improve learning outcomes in schools. *SAGE Open*, *12*(1), 21582440211069392. https://doi.org/10.1177/21582440211069392

Singleton, D. M. (1999). *Exploring the second language mental lexicon*. Cambridge University Press.

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In *New methods in cognitive psychology* (pp. 4-31). Routledge.

Skinner, B. F. (1957). *Verbal behavior*. Appleton-century-crofts.

Smart Nation Office. (2022). *Smart Nation: A National Initiative*. https://www.smartnation.gov.sg/

Smiderle, R., Rigo, S. J., Marques, L. B., Peçanha de Miranda Coelho, J. A., & Jaques, P. A. (2020). The impact of gamification on students' learning, engagement and behavior based on their personality traits. *Smart Learning Environments*, *7*(1), 3. https://doi.org/10.1186/s40561-019-0098-x

Smith, M., & Mccurrach, D. (2021). An examination of the relative benefits & limitations of CALL. *Kwansei Gakuin University Journal of International Studies*, *10*(1), 83-95. https://purehost.bath.ac.uk/ws/portalfiles/portal/218484034/Smith_McCurrach_2021_An_Examination_of_the_Relative_Benefits_Limitations_of_CALL.pdf

Spada, N., & Lightbown, P. M. (2019). Second language acquisition. In *An introduction to applied linguistics* (pp. 111-127). Routledge.

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term relationships among early first language skills, second language aptitude, second language affect, and later second language proficiency. *Applied Psycholinguistics*, *30*(4), 725-755. https://doi.org/10.1017/S0142716409990099

Stiglic, N., & Viner, R. M. (2019). Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. *BMJ open*, *9*(1), e023191. https://doi.org/10.1136/bmjopen-2018-023191

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Sullivan, J. R., Winter, S. M., Sass, D. A., & Svenkerud, N. (2014). Assessing growth in young children: A comparison of raw, age-equivalent, and standard scores using the Peabody Picture Vocabulary Test. *Journal of Research in Childhood Education*, *28*(2), 277-291. https://doi.org/10.1080/02568543.2014.883453

Sun, H., Dhingra, B., Zaheer, M., Mazaitis, K., Salakhutdinov, R., & Cohen, W. (2018). Open domain question answering using early fusion of knowledge bases and text. Proceedings of the 2018 conference on empirical methods in natural language processing,

Sun, H., & Yin, B. (2022). Vocabulary development in early language education. In M. Schwartz (Ed.), *Handbook of early language education* (pp. 57-82). Springer. https://doi.org/10.1007/978-3-030-91662-6_3

Sun, J. C.-Y., & Hsieh, P.-H. (2018). Application of a gamified interactive response system to enhance the intrinsic and extrinsic motivation, student engagement, and attention of English learners. *Educational Technology & Society*, *21*(3), 104-116. https://www.jstor.org/stable/26458511

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, *50*(4), 1183-1202. https://doi.org/10.1016/j.compedu.2006.11.007

Susanto, A. (2017). The teaching of vocabulary: A perspective. *Jurnal Kata*, *1*(2), 183. https://doi.org/10.22216/jk.v1i2.2136

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (Vol. 15, pp. 235-253). Newbury House.

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence*, *21*(1), 83-108. https://doi.org/10.1016/0160-2896(95)90040-3

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/Likert scale. *International Journal of Academic Research in Management (IJARM)*, *8*(1), 1-10. https://doi.org/https://hal.science/hal-02557308/

Tai, T.-Y., & Chen, H. H.-J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive learning environments*, *31*(3), 1485-1502. https://doi.org/10.1080/10494820.2020.1841801

Teng, M. F. (2023). The effectiveness of multimedia input on vocabulary learning and retention. *Innovation in language learning and teaching*, *17*(3), 738-754. https://doi.org/10.1080/17501229.2022.2131791

Theopilus, Y., Al Mahmud, A., Davis, H., & Octavia, J. R. (2024). Preventive Interventions for Internet Addiction in Young Children: Systematic Review. *JMIR mental health*, *11*(1), e56896. https://doi.org/10.2196/56896

Tomasello, M. (2003). Constructing a language: A usage-based theory of child language acquisition. Harvard University Press.

Topping, K., Dekhinet, R., & Zeedyk, S. (2013). Parent–infant interaction and children's language development. *Educational Psychology*, *33*(4), 391-426. https://doi.org/10.1080/01443410.2012.744159

Torrington Eaton, C., & Ratner, N. B. (2016). An exploration of the role of executive functions in preschoolers' phonological development. *Clinical linguistics & phonetics*, *30*(9), 679-695. https://doi.org/10.1080/02699206.2016.1179344

Tsuei, M., & Hsu, Y.-Y. (2019). Parents' acceptance of participation in the integration of technology into children's instruction. *The Asia-Pacific Education Researcher*, *28*(5), 457-467. https://doi.org/10.1007/s40299-019-00447-3

Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: the complexities of cybercompanionship. *Connection Science*, *18*(4), 347-361. https://doi.org/10.1080/09540090600868912

U.S. Department of Education. (2023). *Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations*. Washington, DC: Office of Educational Technology Retrieved from https://tech.ed.gov/ai-report/

Uchikoshi, Y. (2006). Early reading in bilingual kindergartners: Can educational television help? *Scientific Studies of Reading*, *10*(1), 89-120. https://doi.org/10.1207/s1532799xssr1001_5

Unsworth, S. (2016). Early child L2 acquisition: Age or input effects? Neither, or both? *Journal of child language*, *43*(3), 608-634. https://doi.org/10.1017/S030500091500080X

Vaala, S., & Takeuchi, L. (2012). *Co-reading with children on iPads: Parents' perceptions and practices* (The Joan Ganz Cooney Center, Issue. T. J. G. C. C. a. S. Workshop. https://joanganzcooneycenter.org/wp-content/uploads/2012/11/jgcc_ereader_parentsurvey_quickreport.pdf

Vadivel, B., Shaban, A. A., Ahmed, Z. A., & Saravanan, B. (2023). Unlocking English proficiency: Assessing the influence of AI-powered language learning apps on young learners' language acquisition. *International Journal of English Language, Education and Literature Studies*, *2*(6), 55-62. https://doi.org/10.22161/ijeel.2.6.7

Vaiopoulou, J., Papadakis, S., Sifaki, E., Stamovlasis, D., & Kalogiannakis, M. (2021). Parents' perceptions of educational apps use for kindergarten children: Development and validation of a new instrument (PEAU-p) and exploration of parents' profiles. *Behavioral Sciences*, *11*(6), 82. https://doi.org/10.3390/bs11060082

Valagkouti, I. A., Troussas, C., Krouska, A., Feidakis, M., & Sgouropoulou, C. (2022). Emotion recognition in human–robot interaction using the NAO robot. *Computers*, *11*(5), 72. https://doi.org/10.3390/computers11050072

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS quarterly*, 695-704. https://doi.org/10.2307/25148660

Van Dijk, E. M., Lingnau, A., & Kockelkorn, H. (2012). Measuring enjoyment of an interactive museum experience. Proceedings of the 14th ACM international conference on Multimodal interaction,

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in human behavior*, *27*(1), 118-130. https://doi.org/10.1016/j.chb.2010.07.038

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221. https://doi.org/10.1080/00461520.2011.611369

Vatalaro, A., Culp, A. M., Hahs-Vaughn, D. L., & Barnes, A. C. (2018). A quasi-experiment examining expressive and receptive vocabulary knowledge of preschool head start children using mobile media apps. *Early Childhood Education Journal*, *46*, 451-466. https://doi.org/10.1007/s10643-017-0877-3

Vedechkina, M., & Borgonovi, F. (2021). A review of evidence on the role of digital technology in shaping attention and cognitive control in children. *Frontiers in Psychology*, *12*, 611155. https://doi.org/10.3389/fpsyg.2021.611155

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, *39*(2), 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178. https://doi.org/10.2307/41410412

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, *141*, 65-82. https://doi.org/10.1016/j.jecp.2015.06.015

Verhallen, M. J., & Bus, A. G. (2010). Low-income immigrant pupils learning vocabulary through digital picture storybooks. *Journal of educational psychology*, *102*(1), 54. https://doi.org/10.1037/a0017133

Verhallen, M. J., Bus, A. G., & de Jong, M. T. (2006). The promise of multimedia stories for kindergarten children at risk. *Journal of educational psychology*, *98*(2), 410–419. https://doi.org/10.1037/0022-0663.98.2.410

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children’s acquisition and generalization of science concepts. *Child development*, *83*(4), 1137-1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. Vol. 86). Harvard university press.

Walliman, N. (2021). *Research methods: The basics* (3rd ed.). Routledge.

Wang, D., Chen, X.-Y., Ma, Z., Liu, X., & Fan, F. (2022). Has the “Double Reduction” policy relieved stress? A follow-up study on Chinese adolescents. *Child and Adolescent Psychiatry and Mental Health*, *16*(1), 91. https://doi.org/10.1186/s13034-022-00530-6

Wang, G., Zhao, J., Van Kleek, M., & Shadbolt, N. (2022). Informing age-appropriate ai: Examining principles and practices of ai for children. CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems,

Wang, J., Tigelaar, D. E., & Admiraal, W. (2021). Rural teachers' sharing of digital educational resources: From motivation to behavior. *Computers & Education*, *161*, 104055.

Wang, P.-Y., Vaughn, B. K., & Liu, M. (2011). The impact of animation interactivity on novices' learning of introductory statistics. *Computers & Education*, *56*(1), 300-311. https://doi.org/10.1016/j.compedu.2010.07.011

Wang, Q., Quek, C. L., & Hu, X. (2017). Designing and improving a blended synchronous learning environment: An educational design research. *The International Review of Research in Open and Distributed Learning*, *18*(3). https://doi.org/10.19173/irrodl.v18i3.3034

Wang, Z., & Han, F. (2021). Developing English language learners' oral production with a digital game-based mobile application. *Plos one*, *16*(1), e0232671. https://doi.org/10.1371/journal.pone.0232671

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language teaching*, *31*(2), 57-71. https://doi.org/10.1017/S0261444800012970

Warschauer, M., Turbee, L., & Roberts, B. (1996). Computer learning networks and student empowerment. *System*, *24*(1), 1-14. https://doi.org/10.1016/0346-251X(95)00049-P

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of educational psychology*, *103*(2), 455. https://doi.org/10.1037/a0023067

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Wei, L. (2023). Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, *14*, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*(6), 728-755. https://doi.org/10.1080/17405629.2018.1435413

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychol Sci*, *17*(12), 1075-1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: longitudinal perspectives. *Child Dev*, *82*(3), 780-792. https://doi.org/10.1111/j.1467-8624.2011.01583.x

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Dev*, *75*(2), 523-541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wells, G. (1985). *Language development in the pre-school years* (Vol. 2). Cambridge University Press.

Wen, Y., Chiu, M., Guo, X., & Wang, Z. (2024). AI-powered vocabulary learning for lower primary school students. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13537

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development*, *1*(2), 197-234. https://doi.org/10.1207/s15473341lld0102_4

Willams, C. (2007). Research methods. *ournal of Business & Economics Research*, *5*(3), 65-72. https://doi.org/10.19030/jber.v5i3.2532

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). The measurement of executive function at age 5: psychometric properties and relationship to academic achievement. *Psychol Assess*, *24*(1), 226-239. https://doi.org/10.17615/jy5b-4a68

Woo, J. H., & Choi, H. (2021). Systematic review for AI-based language learning tools. *arXiv preprint arXiv:2111.04455*. https://doi.org/10.48550/arXiv.2111.04455

Wu, J. G., & Miller, L. (2020). Improving English Learners' Speaking through Mobile-assisted Peer Feedback. *relc Journal*, *51*(1), 168-178. https://doi.org/10.1177/0033688219895335

Xiao, J., & Jiang, Z. J. (2023). An M-Learning Model in the Context of the Blended Synchronous Learning Environment: A Pilot Study. *International Journal of Mobile and Blended Learning*, *15*(2), 1-16. https://doi.org/10.4018/Ijmbl.318243

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, *28*(11), 13943-13967. https://doi.org/10.1007/s10639-023-11742-4

Yan, S. (2017). Designing Icons on User Interfaces for 4-6 year Old Children. ACHI 2017: The Tenth International Conference on Advances in Computer-Human Interaction,

Yang, L., & Zhao, S. (2024). AI-induced emotions in L2 education: Exploring EFL students' perceived emotions and regulation strategies. *Computers in human behavior*, *159*, 108337. https://doi.org/10.1016/j.chb.2024.108337

Yang, N., Shi, J., Lu, J., & Huang, Y. (2021). Language Development in Early Childhood: Quality of Teacher-Child Interaction and Children's Receptive Vocabulary Competency. *Front Psychol*, *12*, 649680. https://doi.org/10.3389/fpsyg.2021.649680

Yarosh, S., Radu, I., Hunter, S., & Rosenbaum, E. (2011). Examining values: an analysis of nine years of IDC research. Proceedings of the 10th international conference on interaction design and children, Yim, D., Yoon, S., & Lee, J. (2016). Receptive Vocabulary Assessment in Korean-English Bilingual Children: Cross-linguistic Investigations. *Communication Sciences and Disorders-Csd*, *21*(1), 131-138. https://doi.org/10.12963/csd.14186

Zakaria, W., Omar, S. K., Aziz, A. I., & Said, A. (2022). Parents' attitudes towards digital technology use in early childhood. *Social Sciences*, *12*(10), 2531-2548. https://doi.org/10.6007/IJARBSS/v12-i10/15204

Zaman, B., Vanden Abeele, V., & De Grooff, D. (2013). Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction*, *1*(2), 61-70. https://doi.org/10.1016/j.ijcci.2012.12.001

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International journal of educational technology in higher education*, *16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhang, Q. (2021). A literature review of foreign studies on the impact of CALL on second language acquisition from 2015. *English Language Teaching*, *14*(6), 76-83. https://doi.org/10.5539/elt.v14n6p76

Zhao, C. K., & Chen, B. (2023). ICT in education can improve students' achievements in rural China: The role of parents, educators and authorities. *Journal of Policy Modeling*, *45*(2), 320-344. https://doi.org/10.1016/j.jpolmod.2023.02.007

Zhao, W. B., Liu, X. Y., Shah, S., Baah, I., Patel, A., & Wise, N. (2021). Peer Support in Smart Learning and Education. 2021 Ieee Smartworld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People, and Smart City Innovations (Smartworld/Scalcom/Uic/Atc/Iop/Sci 2021), 598-605. https://doi.org/10.1109/Swc50871.2021.00089

Zheng, X. D., Wang, C. C., Shen, Z., & Fang, X. M. (2020). Associations of private tutoring with Chinese students' academic achievement, emotional well-being, and parent-child relationship. *Children and Youth Services Review*, *112*, 104934. https://doi.org/10.1016/j.childyouth.2020.104934

Zhou, Z. (2021). A systematic literature review on the use of mobile-assisted language learning (MALL) for enhancing speaking skills in Chinese EFL context. *International Journal of Frontiers in Sociology*, *3*(15), 12-24. https://doi.org/10.25236/IJFS.2021.031502

Zou, B., Guan, X., Shao, Y. H., & Chen, P. (2023). Supporting Speaking Practice by Social Network-Based Interaction in Artificial Intelligence (AI)-Assisted Language Learning. *Sustainability*, *15*(4), 2872. https://doi.org/10.3390/su15042872

APPENDICES

Appendix 1. PIS for Parents/Caretakers and Children – English

University of Glasgow | School of Education

Participation Information Sheet for Parents/Caretakers

Title of Project: Role of embedded artificial intelligence learning apps in children's second language acquisition

Researcher: Miss Jiachen Wang, email: [redacted]**@student.gla.ac.uk**

Supervisor: Dr. Joanna Wincenciak, email: **Joanna.Wincenciak@glasgow.ac.uk;** Dr. Christopher Hand, email: Christopher.Hand@glasgow.ac.uk

Your child/dependent has been being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish your child to take part.

Thank you for reading this.

1. What is the purpose of the study?
We are interested in finding out about the impact of the embedded-AI language learning app (https://banmaapp.com) developed by China Education Technology Company on children's language learning abilities and social competence.

Through quasi-experimental research with children and interviews with parents, this will help children to better learn second language learning through technology and help to optimise the functionality of the same type of app.

2. Why has my child been chosen?

Your child/dependent has been chosen to take part in this study as they are 5-7 years old and their second language is English. Their involvement in the study will prove to be valuable to the results obtained at the end of the research. Before your child/dependent takes part in our study you will be asked to sign a consent form. Their participation is voluntary and you can withdraw them from this study at any time, even if you have signed the consent form. You can also withdraw their data from the study if you do not feel comfortable having certain information shared. This data will not appear in the final report.

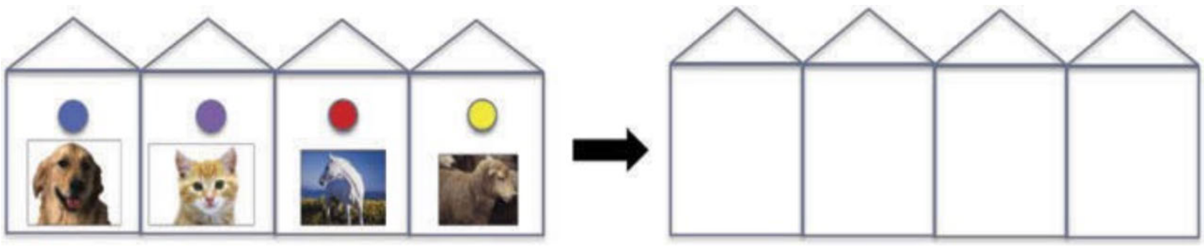
3. Do I have to take part?

It is fully up to you to decide whether or not to allow your child/dependent to take part. If you decide to take part, they are still free to withdraw at any time before the completion of the study and without giving a reason. Your decision to withdraw will not affect you or your child/dependent in any way. However, it will not be possible to withdraw their data after it has been de-identified (approximately 2 weeks after their participation), as de-identified data cannot be traced to them.

4. What will happen to me if I take part?

Your child will be invited to complete three small tests : a vocabulary test, theory of mind test and a working memory test. These tests are designed to better understand the changes in your child's language skills and social skills after using the AI app for language. All three tests are completed in the classroom with the assistance of a researcher. The vocabulary test will assess the impact of using the AI app on children's second language skills through vocabulary. Children will be invited to answer words from pictures made from coloured cards. The theory of mind test is where researchers will show

children a total of 12 animals and 4 colours via PowerPoint and ask them to recall the name or colour of the animal by constantly changing the animal and colour (as shown). This task will take no more than 5-10 minutes to complete.



Finally the child will also be invited to complete a working memory test which will be completed with the assistance of a researcher through a box and doll. At the beginning of the test the tester will tell the child being tested a simple story and will ask the child to accurately predict where the character in the story will be looking for an object that has been moved from one place to another by another person without his/her knowledge. Each test has three questions and this task takes no more than 5-10 minutes to complete.

With your permission, your child will also be asked to fill in a questionnaire which will be used to assess the child's experience and feelings when using the app. The questionnaire contains three sections: emotions, functionality and interface design. A child-friendly "smile-o-meter" scale is also used to assess children's emotions when interacting with different parts of the app. The questions consist of 7 questions and the questionnaire will be printed on paper and will take no more than 5-10 minutes to complete.

In order to fully understand the study population, we will also ask you to fill in a short Demographic questionnair about your child and family online, which will take no more than 5-10 minutes to complete.

If you are interested, you will also be invited to take part in an interview where we will explore your experiences of the programme, or the impact it might have on your child. The interview will take approximately 30-45 minutes. The interview will be conducted over zoom or face to face and offered at times that fit your schedule. With your permission, we will record the interview in order to prepare the transcript. It is up to you in you want to agree to any of the activities related to this research.

5. Will my taking part in this study be kept confidential?

The data gathering process is de-identified, which means that your child/dependent will be given an ID (a combination of numbers and letters) in the process of data collection. All the data will be collected using this ID and not their real name. All information which is collected about your child/dependent during the course of the research will be kept strictly confidential. Please note, however, that due to the small to size of sample and study being conducted in a specific location confidentiality may be impossible to guarantee.

6. What will happen to the results of the research study?

The data gathered will be used to support future academic research and funding applications and may be used for publications in scientific journals. The processed data will be kept in an electronic form only for the duration of up to 10 years, after which it will be destroyed. Your child/dependent, and any of the other participants, will not be identifiable in any of above mentioned ways the results of the data might be used. On your request, the summary of the results of this project will be shared with you. Data collected in may also be used for future projects that focus on any topic and may be unrelated to this study. This new data may be made available to the general public via an open database.

7. Who has reviewed this study?

This study has been considered and approved by the College Research Ethics Committee.

8. Who can I contact for further Information?

If you have any questions about this study, you can ask me, Miss Jiachen Wang (XXXXXXXX@student.gla.ac.uk) or my supervisor, Dr. Joanna Wincenciak (**Joanna.Wincenciak@glasgow.ac.uk**). Dr. Christopher Hand (Christopher.Hand@glasgow.ac.uk)

or the Ethics officer for the School of Education (paul.lynch@glasgow.ac.uk).

Essential statement on confidentiality as required by University Ethics Committee:
Please note that confidentiality will be maintained as far as it possible, unless during data collection we learn anything which make us worried that someone might be in danger of harm, we might have to inform relevant agencies of this.

Thank you for reading this.

End ________________

University of Glasgow | School of Education

1. Child Information Sheet

What is a study?

A research study is what you do when you want to learn about something or find out something new.

Why is this study being done?

We are doing this study to understand how children learn English.

What will happen if I take part?

If you are able, you will be asked to write your name on a form.

University of Glasgow | School of Education

This form is to say that you understand the study and what will happen. You will be given your own copy of the form to keep, as well as this leaflet.

If you take part in this study, the researcher may ask you to play three games, and also ask you to fill in a form and simply tick the emoji that you think best matches you.

Do I have to say yes?

No – not at all. It’s up to you! Just say if you don’t want to join in. Nobody will mind. If you change your mind, that’s ok as well, it will not change the way my friends and me look at you.

What shall I do now?

Now you know about the study you need to think about if you want to take part in the study.

Will the study upset or help me?

At the end of the study, you will receive fun stickers and we hope that the information we get from this study will help boys and girls enjoy their time at school more.

Thank you very much for taking time to read this. Please ask any questions if you need to

Appendix 2. Theory of Mind - 6 Task Stories

1. Diverse Desires

Children see a toy figure of a girl with black hair and yellow skin, and a sheet of paper with an ice cream and an egg drawn on it. The Own desire question “Here’s Xiaohong. It’s snack time, so, Xiaohong wants a snack to eat. Here are two different snacks: an ice cream and an egg. Which snack would you like best? Would you like ice cream or egg best?” If the child chooses the ice cream: “Well, that’s a good choice, but Xiaohong really likes egg. She doesn’t like ice cream. What she likes best are eggs.” (Or, if the child chooses the egg, he or she is told Xiaohong likes ice cream.) Then the child is asked the target question: “So, now it’s time to eat. Xiaohong can only choose one snack, just one. Which snack will Xiaohong choose? An ice cream or an egg?” To be scored as egg, or to pass this task, the child must answer the target question opposite from his or her answer to the own-desire question. This task was adapted from those used by Wellman and Woolley (1990) and Repac Holi and Gopnik (1997).

2. Diverse Beliefs

Children see a toy figure of a girl and a sheet of paper with a bed and a wardrobe drawn on it. “Here’s Lanlan. Lanlan wants to find her socks. Her socks might be hiding under the bed, or it might be hiding in the wardrobe. Where do you think the socks are? Under the bed or in the wardrobe?” This is the own-belief question. If the child chooses the bed: “Well, that’s a good idea, but Lanlan thinks her socks are in the wardrobe.” (Or, if the child chooses the wardrobe, he or she is told Lanlan thinks her socks are under the bed.) Then the child is asked the target question: “So where will Lanlan look for her socks? Under the bed or in the wardrobe?” To be correct the child must answer the target question opposite from his or her answer to the own belief question. This task was adapted from those used by Wellman and Bartsch (1989) and Wellman et al. (1996).

3. Knowledge Access

Children see a plastic box with a drawer containing a small plastic toy ball inside the closed drawer. “Here’s a drawer. What do you think is inside the drawer?” (The child can give any answer he or she likes or indicate that he or she does not know). Next, the drawer is opened and the child is shown the content of the drawer: “Let’s see it’s really a ball inside!” Close the drawer: “Okay, what is in the drawer?” Then a toy figure of a girl is produced: “Lili has never ever seen inside this drawer. Now here comes Lili. So, does Lili know what is in the drawer? (the target question) “Did Lili see inside this drawer?” (the memory question). To be correct the child must answer the target question “no” and answer the memory control question “no.” This task was adapted from those used by Pratt and Bryant (1990) and Pillow (1989).

4. Contents False Belief

Children see a toy figure of a boy and a sheet of paper with a backpack and a closet drawn on it. “Here’s Xinxin. Xinxin wants to find his mittens. His mittens might be in his backpack, or they might be in the closet. Really, Xinxin’s mittens are in his backpack. But Xinxin thinks his mittens are in the closet.” “So, where will Xinxin look for his mittens? In his backpack or in the closet?” (The target question) “Where are Xinxin’s mittens really? In his backpack or in the closet?” (The reality question). To be correct the child must answer the target question “closet” and answer the reality question “backpack.” This task was derived from one used by Wellman and Bartsch (1989) and Siegal and Beattie (1991).

5. Hidden Emotion

Initially, children see a sheet of paper with two faces drawn on it, a happy and a sad face to check that the child knows these emotional expressions. Then that paper is put aside, and the task begins with the child being shown a cardboard cutout figure of a boy drawn from the back so that the boy’s facial expression cannot be seen. “This story is about a boy. I’m going to ask you about how the boy really feels inside and how he looks on his face. He might really feel one way inside but look a different way on his face. Or he might really feel the same way inside as he looks on his face. I want you to tell me how he really feels inside and how he looks on his face.” “This story was about Xiaoming. Xiaoming's uncle was coming back from a business trip abroad. Xiaoming hoped his uncle would bring him a toy gun. If he received the toy gun, Xiaoming would be very happy. If his uncle didn't give him the toy gun, Xiaoming would be sad. But because Xiaoming hadn't seen his uncle in a long time, he didn't want his uncle to think he was sad, so Xiaoming tried to hide how he felt.”

"How would Xiaoming feel if he received the toy gun from his uncle?" (Feel very happy.) "In the story, did Xiaoming want his uncle to see him unhappy?" (No.) Pointing to two emotion pictures: "So, if Xiaoming received a book instead of a toy gun from his uncle, what would Xiaoming truly feel inside? Would he feel happy or sad?" (The target-feel question) "When his uncle didn't give him the toy gun but instead gave him a book, how did Xiaoming try to appear on his face when he received it? Did he look happy or sad?"(The target-look question). To be correct the child’s answer to the target-feel question must be more negative than his or her answer to the target-look question (i.e., sad for target feel and happy for target-look). This task was adapt from one used by Wellman, H. M., et al. (2006).

6. Sarcasm

“The little bear and the little fox are going on a picnic. It's the little bear's idea. He says it's going to be a lovely sunny day. But when they take out the food, dark storm clouds appear. It starts raining, and all the food gets wet. The little fox says, Today is a good day for a picnic."

Children were first asked a preliminary question drawn from Happé (1994): "Is what the little fox said true?" followed by a test question: "Why did the little fox say, 'Today is a good day for a picnic'?" Additionally, the study introduced a comprehension control question: "Was the little fox happy about the rain?" The task included a control question about genuine emotions and a similar "why" test question.

Appendix 3. Demographic Questionnaire - Chinese

What’s your child gender?			
What’s your child date of birth?			
Are you a monolingual or bilingual family？			
Monolingual	Bilingual	Multi-lingual	
Do you have any of the smart devices that can download app (e.g. ipad, smart phone) ?			
Yes	No		
Who is responsible for the main care of the children in your home？			
What is the level of education of the person primarily caring for the child？			
below junior high school	high school	bachelor's degree	bachelor's degree or higher
What is the annual income of your family？			
less than 50000rmb	5000rmb-100000rmb	More than 100000rmb	
What's your employment status ？			
Unemployed	Part-time	Full-time	Self-employed
Does your child have a diagnosis of any of the conditions below?			
ADHD	Autism	Dyslexia	others
Do you have any concerns about your child development?			
Yes (If yes, please provide more details)	No		

Appendix 4. Weekly learning log sheet - Chinese

Zebra AI 应用使用情况 12 周家长记录表

请使用本记录表记录您孩子每天是否完成了 Zebra AI 学习任务。若当天任务完成， 请在对应日期下的方框 内打勾。每周结束时，请标注该周任务是否全部完成，并填写您的备注或观察意见。

周次	周一	周二	周三	周四	周五	周六	周天	是否完成本周任务	父母签字/备注
第一周	☐	☐	☐	☐	☐	☐	☐	☐ 是 ☐ 否	
第二周	☐	☐	☐	☐	☐	☐	☐	☐ 是 ☐ 否	
第三周	☐	☐	☐	☐	☐	☐	☐	☐ 是 ☐ 否	
第四周	☐	☐	☐	☐	☐	☐	☐	☐ 是 ☐ 否	
第五周	☐	☐	☐	☐	☐	☐	☐	☐ 是 ☐ 否	
第六周	☐	☐	☐	☐	☐	☐	☐	☐ 是 ☐ 否	

第七周	□	□	□	□	□	□	□	□ 是 □ 否	
第八周	□	□	□	□	□	□	□	□ 是 □ 否	
第九周	□	□	□	□	□	□	□	□ 是 □ 否	
第十周	□	□	□	□	□	□	□	□ 是 □ 否	
第十一周	□	□	□	□	□	□	□	□ 是 □ 否	
第十二周	□	□	□	□	□	□	□	□ 是 □ 否	

Appendix 5. Interview Questions - English

Semi-structured interviews

Interview questions

1． What do you think about using AI apps as a tool in children's second language learning?

2． Do you think it is helpful for children's second language learning? Please give an example.

3. Do you think that using apps for learning has made a difference to your child? If so, what was it?

4. What do you think is the difference between learning a language using an AI-based learning app and using books?

5． Do you think it has reasonable features? In your opinion, what is the best feature? Why?

6． Did your child have any difficulties in using the app?

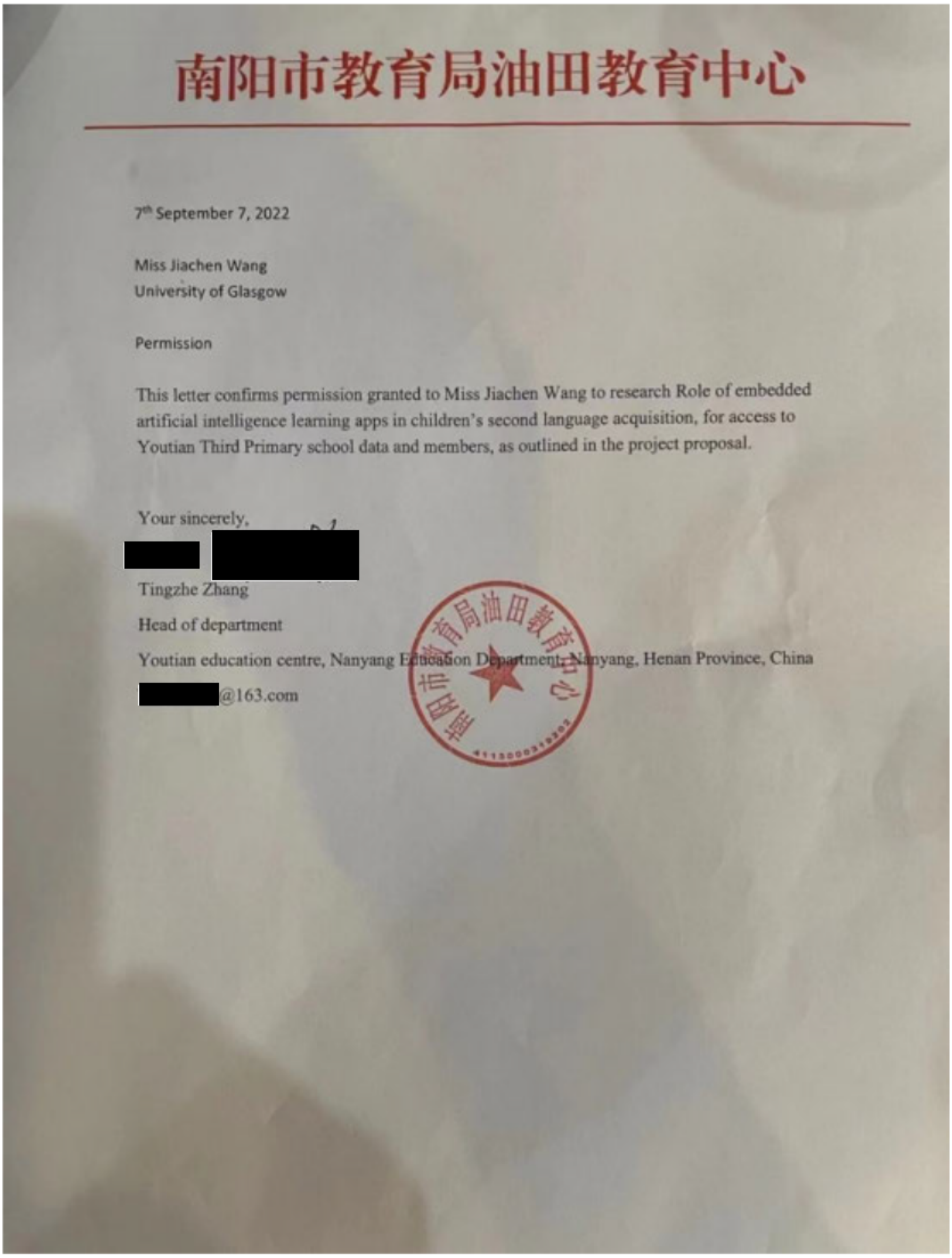
7． What do you think are the pros/cons of the app?

8． What do you think are some factors that might prevent children from using the AI app for second language learning?

9． Do you have any suggestions on how to overcome these problems?

10. What other features would you like to see implemented in AI apps in the future？

Appendix 6. The Approval to Conduct the Study by the Department of Education



南阳市教育局油田教育中心

7th September 7, 2022

Miss Jiachen Wang
University of Glasgow

Permission

This letter confirms permission granted to Miss Jiachen Wang to research Role of embedded artificial intelligence learning apps in children's second language acquisition, for access to Youtian Third Primary school data and members, as outlined in the project proposal.

Your sincerely,

[redacted]

Tingzhe Zhang
Head of department
Youtian education centre, Nanyang Education Department, Nanyang, Henan Province, China
[redacted]@163.com

Appendix 7: Ethical Approval from the University Ethical Committee

University of Glasgow

College of Social Sciences

02 February 2023

Dear Jiachen Wang

College of Social Sciences Research Ethics Committee

Project Title; Role of embedded artificial intelligence learning apps in children's second language acquisition
Application No: 400220113

The College Research Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. It is happy therefore to approve the project, subject to the following conditions:

- Start date of ethical approval: 02/02/2023
- Project end date: 01/02/2025
- Any outstanding permissions needed from third parties in order to recruit research participants or to access facilities or venues for research purposes must be obtained in writing and submitted to the CoSS Research Ethics Administrator before research commences: **socsci-ethics@glasgow.ac.uk**
- The research should be carried out only on the sites, and/or with the groups and using the methods defined in the application.
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (**https://www.gla.ac.uk/media/media_490311_en.pdf**)
- Any proposed changes in the protocol should be submitted for reassessment as an amendment to the original application. The **Request for Amendments to an Approved Application** form should be used:
- Please note reviewers comments. Hope all goes well.

  https://www.gla.ac.uk/colleges/socialsciences/students/ethics/forms/staffandpostgraduateresearchstudents/

Yours sincerely,

Dr Susan A. Batchelor
College Ethics Lead
Susan A. Batchelor, Senior Lecturer
College of Social Sciences Ethics Lead
University of Glasgow
School of Social and Political Sciences &
Scottish Centre for Crime and Justice Research
Ivy Lodge, 63 Gibson Street, Glasgow G12 8LR.
0044+141-330-6167 **socsci-ethics-lead@glasgow.ac.uk**